The Fifteenth International Conference on Learning Analytics & Knowledge

March 3-7, 2025,
Dublin, Ireland

Organized by: SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

acm In-Cooperation sigweb SIGCHI

LAK25

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

ACM ISBN: 979-8-4007-0701-8

LAK25 Program Chairs' Welcome

We are very pleased to welcome you to the Fifteenth International Conference on Learning Analytics and Knowledge (LAK25), organized by the Society for Learning Analytics Research (SoLAR). This year's conference is held in Dublin, Ireland, between March 3rd and 7th, 2025.

The theme for the 15th annual LAK conference is ***Expanding the Horizons of Learning Analytics***. After this many years of research and practice, the learning analytics field has established its own identity, traditions, and community. Pursuing our initial objective of making use of data to better understand and improve learning processes, we have studied and impacted numerous aspects of both formal and informal education. However, as the field enters its teenage years, it faces swift and significant shifts in technological, theoretical, and pedagogical contexts that have a direct effect on our work. For instance, artificial intelligence offers the yet-to-be-proven promise of facilitating and democratizing data analysis, while also posing significant ethical challenges. Critical theories prompt us to examine the values and unintended consequences of our contributions. Novel educational models demand innovative methods for studying learning processes and measuring and assessing learning outcomes. In response, the community is actively reassessing and reshaping what it means to "do learning analytics" within these evolving environments. This process of reinvention often involves stepping out of our comfort zone, established during the field's formative years, to explore new theories, learning processes, data sources, communication modalities, analytical methods, delivery mechanisms, structures for ownership and adoption, and even reconsidering who leads and implements the analytics process. This year's conference aims to highlight and celebrate the trailblazing works that are expanding the horizons of the learning analytics field.

Two excellent keynote talks and a keynote panel present compelling examples of expanding the horizons of learning analytics, but also raise important questions regarding the effects of such an expansion on the learning analytics field itself. Inge Molenaar is the Director of the National Education Lab AI (NOLAI) and a Professor of Education and Artificial Intelligence at the Behavioural Science Institute, Radboud University, Netherlands. Inge's keynote explores the role of artificial intelligence (AI) in education, highlighting the dual role of AI - as both a tool and an actor - and emphasizing the potential for hybrid human-AI collaboration. Gautam Biswas is a Cornelius Vanderbilt Professor of Engineering at Vanderbilt University, whose research focuses on developing intelligent, open-ended learning environments for STEM and computer science education. In his keynote, Gautam presents the work that he and his research team have done on the design, development, and deployment of a multimodal, theoretically grounded learning analytics framework, to analyze and interpret students' collaborative behaviors in STEM environments. The last day of the conference starts with an interactive keynote panel, facilitated by four outstanding learning analytics researchers, namely Rebecca Ferguson (The Open University), Kirsty Kitto (University of Technology Sydney), and Catherine Manly (Fairleigh Dickinson University). Titled "Learning analytics 2035: Pushing the boundaries and meeting the challenges", this interactive panel invites the learning analytics community to ponder on and discuss some of the grand challenges that have been identified during LAK25, explore ways for addressing them and consider what might happen if these are ignored. The conference features two additional panels. One is focused on opportunities, challenges, and risks of adopting learning analytics in higher education settings, whereas the other re-examines the connections and "boundaries" between learning analytics and closely related research fields that rely on educational data (e.g., educational data mining, learning at scale, and quantitative ethnography), in the light of increasing focus on (Generative) AI across all these fields.

This year's conference theme encouraged researchers and practitioners to consider distinct ways of extending the horizons of learning analytics such as proposing novel methods and approaches for data collection, analyses, and communication of analytics results, as well as bringing learning analytics to novel or underexplored learning settings and learning processes, and dealing with ethical issues that novel technologies and learning contexts introduce. This encouragement might partially explain a very large number of high-quality submissions we have received this year, breaking all previous records, and we are extremely grateful for all those who decided to submit the results of their latest research efforts to LAK25. The research track had 337 submissions (232 full paper submissions and 105 short paper submissions). This represents an increase of about 7% in the total number of submissions compared to last year. These papers came from research institutions of 28 countries (11 in Europe, 9 in Asia, 2 in Middle East, 2 in South America, 2 in North America, and 1 in Oceania). Maintaining the high quality of the conference, the program committee for the research track consisted of 280 researchers from the field of learning analytics, educational data mining, learning sciences, educational technology, and related disciplines. Of these, 78 were senior members, all recognized leaders in the field and highly involved in service to the learning analytics community. Overall, from the 337 research submissions, the program committee worked very hard to select 101 papers (70 full research papers and 31 short research papers) that are included in the proceedings of the 15th Learning Analytics and Knowledge Conference. The acceptance rate for both full and short research tracks is 30%.

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 flag any inaccuracies, omissions, or errors in the reviews. This is followed by the meta-review phase during which, for each submission, a senior program committee member, having carefully reviewed the initial reviews and the authors' rebuttal (if submitted), provides a summary meta-review and final recommendation to the program chairs. 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 the learning analytics community. 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. 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 LAK25 participants and other readers of these proceedings will find value in the broad range of contributions to the field of learning analytics contained within. The rapid development and adoption of AI-based technologies, especially generative AI, as well as technological developments more broadly are opening many new opportunities for learning analytics research and practice, but also introducing novel challenges that call for novel methodological approaches and theoretical frameworks. Likewise, further work and novel approaches are needed to assure responsible use and analytics of learning-related data, meet the needs and expectations of diverse stakeholders, as well as ensure ethical conduct in learning analytics research and practice and fair and just treatment of all learners. We hope that the scholarly exchanges at this conference, including the paper presentations, keynotes, panels, and both formal and informal discussions among the participants will contribute to addressing the aforementioned and related challenges and bring us closer to the ultimate objective of understanding and advancing learning and the environments in which it occurs.

Alejandra Martínez Monés
University of Valladolid, Spain

Caitlin Mills
University of Minnesota, USA

Jelena Jovanovic
University of Belgrade, Serbia

Xavier Ochoa
New York University, USA

LAK25 Conference Organization

Program Chairs
Alejandra Martínez Monés, University of Valladolid, Spain
Caitlin Mills, University of Minnesota, USA
Jelena Jovanovic, University of Belgrade, Serbia
Xavier Ochoa, New York University, USA

Practitioner Track Chairs
Tom Olney, Open University UK, United Kingdom
Anouschka van Leeuwen, Utrecht University, Netherlands

Workshop Chairs
Abhinava Barthakur, University of South Australia, Australia
Yizhou Fan, Peking University, China
María Jesús Rodríguez-Triana, University de Valladolid, Spain

Posters and Demonstrations Chairs
Vanessa Echeverria, Monash University, Australia
Sonsoles López-Pernas, University of Eastern Finland, Finland

Doctoral Consortium Chairs
Christopher Brooks, University of Michigan, USA
Nia Nixon, University of California Irvine, United States
Negin Mirriahi, University of South Australia, Australia
Luis Prieto, Universidad de Valladolid, Spain

Proceedings Chairs
Andrew Zamecnik, University of South Australia, Australia
Vishal Kuvar, University of Minnesota, USA
Aaron Wong, University of Minnesota, USA

Website Manager
Yancy Vance Paredes, Arizona State University, USA

Organizing Committee
Geraldine Gray, TU Dublin, Ireland
Bart Rienties, Open University, UK
Grace Lynch, Society for Learning Analytics Research, Australia
Nicole Hoover, Society for Learning Analytics Research, USA

LAK25 Program Committee

Senior Program Committee

Carlos Alario-Hoyos, Universidad Carlos III de Madrid, Spain
Vincent Aleven, "Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA", Turkmenistan
Juan I. Asensio-Pérez, "Collaborative and Intelligent Systems Group, Universidad de Valladolid", Spain
Roger Azevedo, University of Central Florida, United States
Ryan Baker, University of Pennsylvania, United States
Stephanie Baribeau, University of Saskatchewan, Canada
Gautam Biswas, Vanderbilt University, United States
Nigel Bosch, University of Illinois Urbana-Champaign, United States
Miguel L. Bote-Lorenzo, Universidad de Valladolid, Spain
Anthony F. Botelho, University of Florida, United States
Christopher Brooks, University of Michigan, United States
Simon Buckingham Shum, "University of Technology, Sydney", Australia
Guanliang Chen, Monash University, Australia
Irene-Angelica Chounta, University of Duisburg-Essen, Germany
Linda Corrin, Deakin University, Australia
Mutlu Cukurova, University College London, United Kingdom
Rafael D. Araújo, Universidade Federal de Uberlandia, Brazil
Shane Dawson, University of South Australia, Australia
Stavros Demetriadis, Aristotle University of Thessaloniki, Greece
Yannis Dimitriadis, University of Valladolid, Spain
Vania Dimitrova, University of Leeds, United Kingdom
Blazenka Divjak, "University of Zagreb, Faculty of Organization and Informatics", Croatia
Hendrik Drachsler, The Open University, Netherlands
Brendan Eagan, University of Wisconsin-Madison, United States
Yizhou Fan, Peking University, United Kingdom
Stephen Fancsali, "Carnegie Learning, Inc.", United States
Shihui Feng, The University of Hong Kong, Hong Kong
Rebecca Ferguson, The Open University, United Kingdom
Brendan Flanagan, Kyoto University, Japan
Dragan Gasevic, Monash University, Australia
Michail Giannakos, Norwegian University of Science and Technology, Norway
Geraldine Gray, Technological University Dublin, Ireland
Marek Hatala, Simon Fraser University, Canada
Neil Heffernan, Worcester Polytechnic Institute, United States
Davinia Hernandez-Leo, "Universitat Pompeu Fabra, Barcelona", Spain
Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile
Xiao Hu, The University of Hong Kong, Hong Kong
Stephen Hutt, University of Denver, United States
Ioana Jivet, FernUniversität in Hagen, Germany
Srecko Joksimovic, "Education Future, University of South Australia", Australia
Mohammad Khalil, University of Bergen, Norway
Hassan Khosravi, The University of Queensland, Australia
Kirsty Kitto, "University of Technology, Sydney", Australia
Simon Knight, UTS, Australia
Vitomir Kovanovic, The University of South Australia, Australia
Tanja Käser, EPFL, Switzerland
Nancy Law, The University of Hong Kong, Hong Kong
Tobias Ley, University for Continuing Education Krems, Austria
Sonsoles López-Pernas, University of Eastern Finland, Finland

Roberto Martinez-Maldonado, Monash University, Australia
Agathe Merceron, Berliner Hochschule für Technik - Berlin State University of Applied Sciences, Germany
Negin Mirriahi, University of South Australia, Australia
Tanja Mitrovic, "Intelligent Computer Tutoring Group, University of Canterbury, Christchurch", New Zealand
Pedro J. Muñoz-Merino, Universidad Carlos III de Madrid, Spain
Luc Paquette, University of Illinois at Urbana-Champaign, United States
Zach Pardos, "University of California, Berkeley", United States
Oleksandra Poquet, Technical University of Munich, Germany
Mar Pérez-Sanagustín, Université Paul Sabatier Toulouse III, France
Mladen Rakovic, Monash University, Australia
Bart Rienties, The Open University, United Kingdom
Ido Roll, Technion - Israel Institute of Technology, Israel
Olga C. Santos, UNED, Spain
Mohammed Saqr, University of Eastern Finland, Finland
Maren Scheffel, Ruhr University Bochum, Germany
David Shaffer, University of Wisconsin-Madison, United States
Marcus Specht, Delft University of Technology, Netherlands
Daniel Spikol, University of Copenhagen, Denmark
Zachari Swiecki, Monash University, United States
Yi-Shan Tsai, Monash University, Australia
Katrien Verbert, Katholieke Universiteit Leuven, Belgium
Olga Viberg, KTH Royal Institute of Technology, Sweden
Barbara Wasson, University of Bergen, Norway
Phil Winne, Simon Fraser University, Canada
Alyssa Wise, Vanderbilt University, Canada

Program Committee

Seth Adjei, Northern Kentucky University, United States
Samantha Ahern, University College London, United Kingdom
Sarah Alcock, The Open University, United Kingdom
Giora Alexandron, Weizmann Institute of Science, Israel
Riordan Alfredo, Monash University, Australia
Naif Aljohani, University of Southampton, United Kingdom
Laura Allen, University of Minnesota, United States
Ainhoa Alvarez, Universidad del País Vasco (UPV/EHU), Spain
Ishari Amarasinghe, Universitat Pompeu Fabra, Spain
Kimberly Arnold, University of Wisconsin-Madison, United States
Aneesha Bakharia, The University of Queensland, Australia
Abhinava Barthakur, University of South Australia, Australia
Alan Berg, University of Amsterdam, Netherlands
Anabel Bethencourt Aguilar, Universidad de La Laguna, Spain
Christopher Bogart, Carnegie Mellon University, United States
Maurice Boothe, New York University, United States
Conrad Borchers, Carnegie Mellon University, United States
François Bouchet, Sorbonne Université - LIP6, France
Alex Bowers, Columbia University, United States
Julien Broisin, "IRIT, Université Toulouse III - Paul Sabatier, Toulouse, France", France
Okan Bulut, University of Alberta, Canada
Manuel Caeiro Rodríguez, Universidad de Vigo, Spain
Oscar Canovas, University of Murcia, Spain
Dan Carpenter, North Carolina State University, United States
Cristian Cechinel, contato@cristiancechinel.pro.br, Brazil
Teresa Cerratto-Pargman, Stockholm University, Sweden
Mohamed Amine Chatti, University of Duisburg-Essen, Germany

Lujie Chen, "University of Maryland, Baltimore County", United States
Zhongzhou Chen, University of Central Florida Physics Department, United States
Li Chen, "Faculty of Information Science and Electrical Engineering, Kyushu University", Japan
Marco Chiarandini, University of Southern Denmark, Denmark
Katherine Chiluiza, Escuela Superior Politécnica del Litoral / Information Technology Center, Ecuador
Elizabeth Cloude, Tampere University, Finland
Clayton Cohn, Vanderbilt University, United States
Miguel Ángel Conde, University of León, Spain
Alexandra Cristea, Durham University, United Kingdom
Cynthia D'Angelo, University of Illinois at Urbana-Champaign, United States
Daryn Dever, University of Florida, United States
Daniele Di Mitri, DIPF | Leibniz Institute for Research and Information in Education, Germany
Juan Manuel Dodero, Universidad de Cádiz, Spain
Jan Dolonen, "Department of Education, University of Oslo", Norway
Dariane Drake, UW-Madison, United States
Martin Ebner, Graz University of Technology, Austria
Vanessa Echeverria, Monash University, Australia
Matthias Ehlenz, RWTH Aachen University, Germany
Gloria Milena Fernandez-Nieto, "University of Technology, Sydney", Australia
Angela Fessl, "Know-Center, Graz", Austria
Xuewang Geng, "Graduate School of Human-Environment Studies, Kyushu University", Japan
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
Xiaoqing Gu, East China Normal University, China
Julio Guerra, Universidad Austral de Chile, Chile
Ashish Gurung, Carnegie Mellon University, United States
Eduardo Gómez-Sánchez, Universidad de Valladolid, Spain
Ella Haig, "School of Computing, University of Portsmouth", United Kingdom
Ángel Hernández-García, Universidad Politécnica de Madrid, Spain
Martin Hlosta, The Swiss Distance University of Applied Sciences, Switzerland
Adrian Holzer, University of Neuchâtel, Switzerland
Andrea Horbach, "CATALPA, FernUniversität Hagen", Germany
Izumi Horikoshi, Kyoto University, Japan
Sarah Howard, University of Wollongong, Australia
Bowen Hui, The University of British Columbia, Canada
Jui-Long Hung, "Boise State University, USA", United States
Paul Hur, Free University of Berlin, Germany
Santiago Iglesias, Universidad Politécnica de Madrid, Spain
Yeonji Jung, University of Memphis, United States
Rogers Kaliisa, University of Oslo, Norway
Panayiota Kendeou, University of Minnesota, United States
Michael Kickmeier-Rust, Graz University of Technology, Austria
Yoon Jeon Kim, University of Wisconsin-Madison, United States
Kenneth Koedinger, Carnegie Mellon University, United States
Elizabeth Koh, "National Institute of Education, Nanyang Technological University, Singapore", Singapore
Malgorzata Korolkiewicz, UniSA, Australia
Milos Kravcik, DFKI GmbH, Germany
Huan Kuang, Florida State Univeristy, United States
Jakub Kužílek, Humboldt-Universität zu Berlin, Germany
Sébastien Lallé, Sorbonne University, France
Andrew Lan, University of Massachusetts Amherst, United States
Charles Lang, Columbia University, United States
Mikel Larrañaga, Universidad del País Vasco (UPV/EHU), Spain

Eitel Lauria, Marist College, United States
Elise Lavoué, "iaelyon, Université Jean Moulin Lyon 3, LIRIS", France
Alwyn Vwen Yen Lee, Nanyang Technological University, Singapore
Arun Balajiee Lekshmi Narayanan, University of Pittsburgh, United States
Armanda Lewis, New York University, United States
Huiyong Li, "Research Institute for Information Technology, Kyushu University, Japan", Japan
Gen Li, Kyushu University, Japan
Lisa Lim, University of Technology Sydney, Australia
Lyn Lim, Technical University of Munich, Germany
Jionghao Lin, Carnegie Mellon University, United States
Yiwen Lin, "University of California, Irvine", United States
Andreas Lingnau, Deutsche Hochschule für angewandte Wissenschaften (DHAW), Germany
Martín Llamas-Nistal, Universidad de Vigo, Spain
Vanda Luengo, Sorbonne Université - LIP6, France
Collin Lynch, North Carolina State University, United States
Joni Lämsä, University of Oulu, Finland
Catherine Manly, Fairleigh Dickinson University, United States
Mirko Marras, University of Cagliari, Italy
Rebecca Marrone, University of South Australia, Australia
Uwe Matzat, Eindhoven University of Technology, Netherlands
Konstantinos Michos, EPFL, Switzerland
Kamila Misiejuk, FernUniversität Hagen, Germany
Pedro Manuel Moreno-Marcos, Universidad Carlos III de Madrid, Spain
Jorge Munoz-Gama, Pontificia Universidad Católica de Chile, Chile
Hanni Muukkonen, University of Oulu, Finland
Tomohiro Nagashima, Saarland University, Germany
André Nascimento, UFRPE, Brazil
Andy Nguyen, University of Oulu, Finland
Ha Nguyen, University of North Carolina at Chapel Hill, United States
Omid Noroozi, Wageningen University, Netherlands
Alejandro Ortega-Arranz, Universidad de Valladolid, Spain
Korinn Ostrow, Edmentum, United States
Zacharoula Papamitsiou, SINTEF Digital, Norway
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
Gerti Pishtari, University for Continuing Education Krems (Danube University), Austria
Taciana Pontual Falcão, Universidade Federal Rural de Pernambuco, Brazil
Stanislav Pozdniakov, The University of Queensland, Australia
Carlos Gerardo Prieto Alvarez, The University of Sydney, Australia
Ying Que, The University of Hong Kong, China
Joshua Quick, Indiana University Bloomington, United States
Sabina Rako, University Computing Centre University of Zagreb, Croatia
Anand Ramakrishnan, Worcester Polytechnic Institute, United States
Cheng Ren, "University of California, Berkeley", United States
María Jesús Rodríguez-Triana, Universidad de Valladolid, Spain
Daniela Rotelli, University of Pisa, Italy
José A. Ruipérez Valiente, University of Murcia, Spain
René Röpke, RWTH Aachen University, Germany
Merike Saar, Tallinn University, Estonia
Shazia Sadiq, The University of Queensland, Australia
John Saint, Regent's University London, United Kingdom
Fatemeh Salehian Kia, University of British Columbia, Canada
Patricia Santos, Universitat Pompeu Fabra, Spain

Marcel Schmitz, Zuyd Hogeschool, Netherlands
Clara Schumacher, Humboldt Universität zu Berlin, Germany
Avi Segal, Ben Gurion University, Israel
Niels Seidel, FernUniversität in Hagen, Germany
Kshitij Sharma, Norwegian University of Science and Technology, Norway
Bruce Sherin, Northwestern University, United States
Yang Shi, Utah 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
Namrata Srivastava, Vanderbilt University, United States
Benyamin T. Tabarsi, North Carolina State University, United States
Benyamin Tabarsi, North Carolina State University, United States
Yuta Taniguchi, Kyushu University, Japan
Michelle Taub, University of Central Florida, United States
Mohammadreza Tavakoli, TIB, Germany
Craig Thompson, The University of British Columbia, Canada
Paraskevi Topali, "NOLAI, National Education Lab AI, Behavioural Science Institute, Radboud University", Netherlands
Ashwin Tudur Sadashiva, ISIS Vanderbilt, United States
Cristina Villa-Torrano, Universidad de Valladolid, Spain
Nicola Wilkin, University of Birmingham, United Kingdom
Kimberly Williamson, Cornell University, United States
Aaron Wong, University of Minnesota, United States
Simon Woodhead, Eedi, United Kingdom
Jiun-Yu Wu, Southern Methodist 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
Chengjiu Yin, Kyushu University, Japan
Zhang Yupei, Northwestern Polytechnical University, China
Kateryna Zabolotna, University of Oulu, Finland
Jianwei Zhang, University at Albany, United States
Linxuan Zhao, Monash University, Australia
Xinran Zhu, University of Illinois Urbana Champaign, United States
Gaoxia Zhu, "National Institute of Education (NIE), Nanyang Technological University", Singapore

Additional Reviewers

Abdessalam Ouaazki
Abubakir Siedahmed
Alrike Claassen
Amanda Barany
Anailys Hernández Julián
Antonio Balderas
April Clarke
Christopher Steadman
Christos Katsanos
Derek Maki
Dimitrios Tzimas
Emanuel Queiroga

Hongming Li
Jaclyn Ocumpaugh
Jason Bernard
Javier Pastor-Galindo
Jie Cao
José Miguel Mota Macias
Juan Antonio Caballero-Hernández
Kirk Vanacore
Luis Alberto Pinos Ullauri
Manon Berney
Manuel Jesus Gomez Moratilla
Mariano Albaladejo-González
Meng Cao
Morgan Lee
Nabil Al Nahin Ch
Natalia Bartłomiejczyk
Osmel Bordiés López
Paraskevas Lagakis
Qinyi Liu
Ruth Cobos
Sami Baral
Seiyon Lee
Shan Zhang
Shintaro Sato
Shreya Bhandari
Shruti Jain
Sofia Strukova
Vinitra Swamy
Vladimir Macko
Víctor Alonso-Prieto
Wissam Saleh
Yerin Kwak

Practitioner Reports

Demos

Workshops

Lessons Learnt on How to Combine Top-Down Facilities for Learning Analytics with Bottom-Up Initiatives

Marije Goudriaan
Utrecht University
m.goudriaan1@uu.nl

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

Ünal Aksu
Utrecht University
u.aksu@uu.nl

ABSTRACT: We describe a University's general approach to initiating LA projects, namely a combination of a top-down and bottom-up approach, and evaluate this approach based on the results from two LA projects. Project 1 is about providing study delay predictions via a study advisor dashboard. Project 2 focuses on enabling students to self-monitor academic writing skills through a student dashboard. Smooth coordination between pedagogy, privacy, and technology resulted in the successful realization of these projects. However, the initial adoption of the dashboards by end-users was limited. We discuss potential causes, solutions, and general recommendations for institutions that are working on the adoption of LA.

Keywords: Learning Analytics, Dashboard, Adoption, Data Literacy, Stakeholder Management

1 INTRODUCTION AND AIM OF THE PAPER

Learning Analytics (LA) is a complex endeavor requiring input from multiple stakeholders. A particular challenge lies in coordinating the support necessary at an institutional level in terms of policy and funds needed for the initiation and upscaling of LA projects at the staff and student level (Broos et al., 2020). At our University, we have chosen a combination of a top-down and bottom-up approach (Perez-Sanagustin et al., 2022). Top-down, there is institutional support and University policy for the areas in which LA can be applied to improve the quality of education (macro-layer). Bottom-up, a central LA team is available to support LA project submissions (micro-layer) following a roadmap for initiation and evaluation of each pilot (Van Leeuwen et al., 2024). In this paper, we detail our experiences with this way of working by describing two LA projects. A coordinated effort between pedagogy, privacy, and technology led to the successful realization of these projects. However, there was limited end-user adoption of the developed dashboards in these projects. We describe the projects and their initial evaluation (sections 2 and 3), and end with a general reflection (section 4) on the combination of top-down and bottom-up approach and our recommendations for practice.

2 PROJECT 1: STUDY DELAY PREDICTIONS

Project 1 was requested by study advisors at our University. Study advisors' practice in Higher Education includes regularly monitoring students' progress and offering support regarding imminent or present study delay (Sharkin, 2004). In co-design with the study advisors, the central LA team developed a dashboard that provides imminent study delay predictions based on data from earlier cohorts, see Figure 1. This allowed study advisors to identify students in the current cohort that might be at risk and help prevent the negative consequences of study delay (Baars et al., 2022). A group of 9 study advisors worked with the dashboard for several months and were interviewed afterwards.

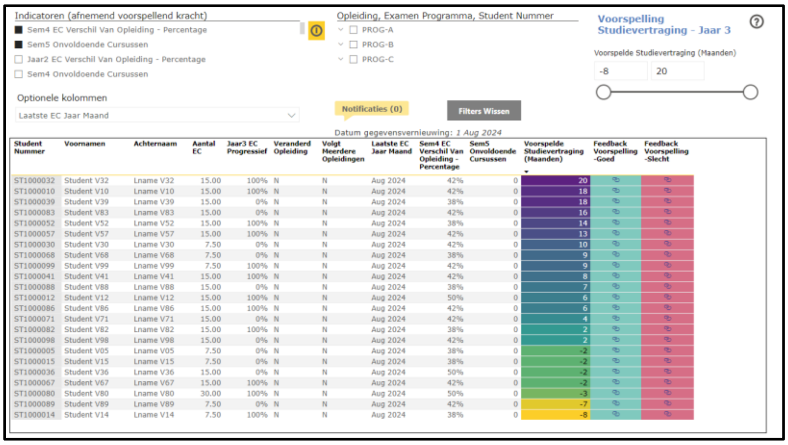


Figure 1: Screenshot of the study advisor dashboard (synthetic data; in Dutch)

The pilot showed that the study advisors found the design of the dashboard sufficient in terms of usability, but they judged its usefulness as low, which resulted in discontinuation of using the dashboard in their practice. Evaluation in terms of interviews showed that there were three core problems underlying this finding. 1) Variation in data literacy skills: the central LA team provided a training and introduction to the prediction modelling that underlies the dashboard. However, some study advisors did not trust or grasp the predictions made by the dashboard and continued to rely on obtained ECs as the only indicator for study delay. 2) Lack of time and means to participate in the project: the study advisors were free to choose whether to participate, and had to do so in parallel to their usual workflow. This resulted in a high reported workload and no full immersion in investigating the potential of the dashboard. 3) Limited new insights: the two study advisors that engaged most with the dashboard indicated the dashboard led to the identification of only a few students that they were not yet aware of. They thought this number was not enough to continue the effort of transitioning to a new way of working.

3 PROJECT 2: SELF-MONITORING ACADEMIC WRITING SKILLS

Project 2 was a request from teachers as part of an education innovation program. Student-facing dashboards aimed at supporting skill development have the potential to increase student skills as well as increase students' persistence in their study program (Grann & Bushway, 2014). In co-design with the teachers, the central LA team created a dashboard visualizing students' academic writing skills overarching several courses, see Figure 2.

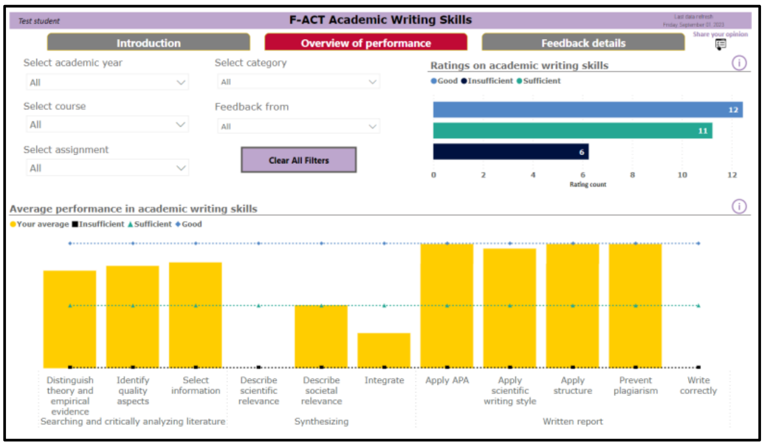


Figure 2: Screenshot of the self-monitoring academic writing skills dashboard (synthetic data)

The dashboard was used in an undergraduate course in which around 350 students were enrolled and was introduced in week 1. It could be accessed through the LMS, along with a video tutorial. The dashboard's usability and usefulness were assessed with a student questionnaire and a focus group with students and teachers. Unfortunately, only 12 students (3.5% of the students) consulted the dashboard regularly. Based on the (limited) information we received, we identified the following reasons for the limited uptake: 1) The dashboard relied on data that were extracted from a specific, newly introduced application. Thus, the dashboard did not contain information about writing skills from previous years. It only started to contain information after the first half of the course. This may have demotivated students to consult it. 2) Additionally, not all teachers were willing or accustomed to the use of the new application. As a result, there was no data available for some students about their writing assignments. 3) Using the dashboard was voluntarily and not incorporated into the curriculum, which led to low awareness of the existence of the dashboard despite the introduction session.

4 CONCLUSION

In this paper, we described two LA projects in a University where top-down facilities are combined with the bottom-up initiation of LA projects. The advantages of this approach are that there is both management buy-in and stakeholder involvement from the start. This ensures that the LA projects are directly based on the wishes from the end-users, such as teachers and study advisors. However, the evaluation of the projects showed a number of challenges. For example, both projects required a change in the daily practice or workflow of the involved stakeholders, i.e., study advisors (project 1) and teachers and students (project 2). This turned out to be difficult to achieve, even though the end-users were involved in the design process of the LA. One of the reasons may be that the initiators of the projects were enthusiastic early adaptors who were dealt the task of "selling" the idea to their colleagues. Although enthusiasm from an end-user is often the key to unlocking buy-in from others in the long run, at the start of a project it requires considerable time to convince colleagues that the use of LA will have benefits that outweigh the effort required for implementing the change (Charleer et al., 2016). Our evaluations show that support from a team leader or program director – in other words, the meso-layer between institutional leaders (macro) and end-users (micro) – is essential to provide the required time, resources, and motivation to implement an LA project on a larger scale. Moreover, the employed approach has a downside for the central LA team that supports LA initiatives. For the team, it is hard to anticipate training needs of end-users to work with the LA application because there is no long-term planning for the kind of LA projects that will be initiated. In project 1 for example it would have helped if a more elaborate data literacy training had been offered.

Based on our experiences, we recommend the following:

- To include an *implementation plan* as part of the strategy, regardless of the approach (top-down or bottom-up). In this plan, specific attention needs to be paid to how the LA tool will be integrated into the existing workflow, and how end-users will be “convinced” to use it. Avoid introducing the LA tool as a voluntary, separate activity.

- Our approach already included the development of top-down facilities that are needed concerning pedagogy, privacy, and ethics for LA projects, such as a LA policy. Through two projects, we recognized the importance of institution-wide staff capabilities. Therefore, we would recommend investing in *professional development programs*, which could either be offered to all staff independent from the LA projects, or by including a specialist in the LA team that can offer ad-hoc training when implementing an LA initiative.

To conclude, while our approach so far ensures top-down support (macro-layer) and bottom-up involvement of stakeholders (micro-layer) for LA, challenges remain in terms of the required skills and time to achieve impactful implementation of LA tools. Investing in implementation strategies by involving the *meso-layer* of team leaders may be a promising direction forward.

REFERENCES

Baars, G. J. A., Schmidt, H. G., & Hermus, P. (2022). Early Identification of Successful and Unsuccessful Students in the First Year at the University. *Health Professions Education*, *8*(1), 17–26. https://doi.org/10.55890/2452-3011.1016

Broos, T., Hilliger, I., Pérez-Sanagustín, M., Htun, N., Millecamp, M., Pesántez-Cabrera, P.,…& De Laet, T. (2020). Coordinating learning analytics policymaking and implementation at scale. *British Journal of Educational Technology*, *51*(4), 867-1435. https://doi.org/10.1111/bjet.12934

Chareer, S., Klerkx, J., Duval, E., De Laet, T., & Verbert, K. (2016). Creating Effective Learning Analytics Dashboards: Lessons Learnt. In: Verbert, K., Sharples, M., Klobučar, T. (eds) Adaptive and Adaptable Learning. EC-TEL 2016. Lecture Notes in Computer Science, vol 9891. Springer, Cham. https://doi.org/10.1007/978-3-319-45153-4_4

Grann, J. & Bushway, D. (2014). Competency map: visualizing student learning to promote student success. Proceedings of LAK2014, 168–172. https://doi.org/10.1145/2567574.2567622

Sharkin, S. (2004). College Counseling and Student Retention: Research Findings and Implications for Counseling Centers. *Journal of College Counseling*, *7*(2), 99–108.

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

Van Leeuwen, A., Goudriaan, M., Aksu, Ü., & Yousufzai, M. (2024). A Roadmap for Implementing Learning Analytics Projects: taking into account Pedagogy, Privacy, Ethics, and Technical Infrastructure. Companion Proceedings of the Fourteenth International Conference on Learning Analytics & Knowledge (LAK24), p. 9-12.

Leveraging Learning Analytics to Support North Carolina's Advanced Teaching Roles Program

Shaun Kellogg and Robin Anderson
North Carolina State University
shaun.kellogg@ncsu.edu, randers6@ncsu.edu

Thomas Tomberlin
North Carolina Department of Public Instruction
thomas.tomberlin@dpi.nc.gov

ABSTRACT: The North Carolina Teacher Compensation and Advanced Teaching Roles (ATR) program allows public school units to develop innovative teacher compensation models designed to improve student and teacher outcomes. The program enables highly effective teachers, known as Advanced Teachers, to either take responsibility for more students or lead small teams of teachers by providing professional development, coaching, and instructional support. The North Carolina Department of Public Instruction selected the Friday Institute for Educational Innovation at North Carolina State University as their research partner. This partnership has two primary goals: 1) to assess the academic and professional impact of ATR programs, and 2) to understand and improve their implementation. Using a collaborative data-intensive improvement research framework, the research employs a variety of methods, including both conventional qualitative and statistical methods, as well as more novel approaches drawn from the field of learning analytics such as data dashboards, epistemic network analysis, and machine learning.

Keywords: teacher leadership, performance-based compensation, epistemic network analysis, machine learning, data dashboards

1 INTRODUCTION

In 2016, the North Carolina General Assembly enacted legislation to create the Teacher Compensation Models and Advanced Teaching Roles (ATR) program. The ATR program enables local school administrative units to create innovative models that allow highly effective classroom teachers to impact an increased number of students. Broadly defined, Advanced Teachers are highly effective classroom teachers who are provided salary supplements and reach an increased number of students by either 1) assuming academic accountability for an increased number of students, or 2) becoming a lead classroom teacher accountable for the student performance of all students taught by teachers on that Advanced Teacher's team.

1.1 Researcher-Practitioner Partnership Goals

To support these efforts, legislation directs the North Carolina State Board of Education to contract with an independent research organization to evaluate what ATR has accomplished. The Friday Institute was selected as the program's research partner to assist the North Carolina Department of Public Instruction (NCDPI) and Public School Units (PSUs) with two overarching goals for this researcher-practitioner partnership (RPP):

1. to assess the academic and instructional **impact** of ATR programs, as well as their impact on the teaching profession; and
2. to understand and improve the **implementation** of these programs and identify factors supporting or impeding their success.

In July 2024, the Friday Institute was awarded additional funding by the NCDPI to expand upon legislatively required program evaluation efforts and conduct a mixed-methods measurement study focused on the selection and evaluation of Advanced Teachers.[1] Guided by a collaborative, data-intensive improvement research model (Krumm et al., 2018), this study aims to:

- document current selection and evaluation criteria and processes used for ATR.
- examine the relationship between selection criteria and outcomes and program impacts.
- develop and validate new measures for selection and evaluation of advanced teaching roles.
- explore the practical and ethical implications of new and existing measures and models.
- provide data-informed recommendations and resources for teacher selection and evaluation.

2 THE ADVANCED TEACHING ROLES PROGRAM

The purpose of ATR is to allow highly effective classroom teachers to impact an increased number of students by assuming accountability for additional students. In addition, the program enables PSUs to create innovative compensation models that focus on classroom teacher professional growth and that lead to measurable improvements in student outcomes. Per section 2.6.(b) of NC Session Law 2020-78, the intent of the ATR programs is to allow highly effective classroom teachers to reach an increased number of students by either 1) teaching an increased number of students and assuming accountability for their performance; or 2) becoming a lead classroom teacher accountable for the student performance of all of the students taught by teachers on that lead classroom teacher's team. These Advanced Teachers are designated as Classroom Excellence and Adult Leadership teachers respectively. Furthermore, PSUs receive funding from the state to provide salary supplements of $10,000 for Adult Leadership teachers, and $3,000 for Classroom Excellence teachers.

2.1 2024 Evaluation Findings

During the 2023-24 school year, 17 PSUs implemented ATR programs across 277 schools. PSUs employed 849 Advanced Teachers who supported 2,461 classroom teachers, with schools averaging three Advanced Teachers and nine supported teachers per school. Most PSUs, 13 out of 17, currently partner with – or launched their initial ATR work via partnership with – Public Impact, a third-party vendor for ATR programs. ATR schools produced significant effects on students' math test scores and positive but not significant results in ELA and science; these effects grew over time, and teachers in ATR schools were more likely to have higher value-added scores after implementing the program (Kellogg et al., 2024). In addition, PSU case studies highlighted how ATR has provided students receiving Tier 2 and 3 services through MTSS with greater access to effective teachers, and demonstrated how ATR serves as both a career lattice and ladder for professional advancement.

[1] https://fi.ncsu.edu/projects/the-selection-and-evaluation-of-advanced-teachers/

3 LEARNING ANALYTICS METHODS EMBEDDED

Data collection and analyses follow a mixed-methods sequential design (Creswell & Clark, 2017), which brings together the differing strengths and nonoverlapping weaknesses of quantitative methods with those of qualitative methods. To complement more traditional qualitative and quantitative methods employed in this RPP, the Friday Institute is incorporating several approaches drawn from the field of Learning Analytics. While approaches such as data dashboards have already gained widespread adoption among practitioners, this RPP is also embedding Epistemic Network Analysis (ENA) and Machine Learning. These methods are relatively novel for programs like ATR, and their utility has yet to be determined for the program.

3.1 Data Dashboards & Reporting Tools

The Friday Institute has worked with practitioners to share relevant, context-specific data intended to provide state and district leadership with informative and actionable data. Data reporting tools include online interactive dashboards and other reporting tools that include programmatic summaries of NCDPI and PSU-provided data tailored for each partner. For example, Figure 1 shows a data dashboard that provides a statewide summary for ATR programs across NC, including information about 3,310 Advanced Teachers and the teachers they support. These dashboards include data such as salary supplements, subject areas supported, PSU position titles, geographic location, release time, and size of programs and teams. Each data point also serves as a filter allowing practitioners to explore their data at a more granular level and answer questions they may have about their specific programs.

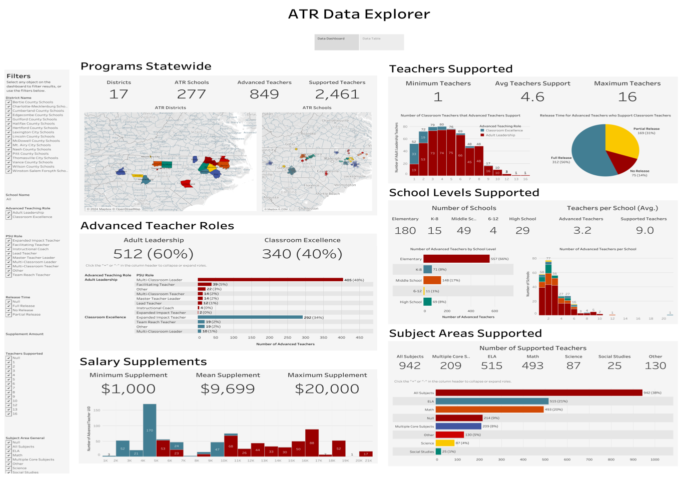


Figure 1: Interactive ATR data explorer

3.2 Epistemic Network Analysis

The RPP also leverages Epistemic Network Analysis (ENA) to analyze coded interview data collected for traditional qualitative analyses. ENA has emerged as a novel and promising ethnographic method for identifying, quantifying, and visualizing connections among elements in coded data, such as text-based transcripts of stakeholder interviews (Shaffer & Ruis, 2017). Within this RPP, the research team is using ENA to visualize, quantify, and compare the epistemic frames or "mental models" of ATR stakeholders with respect to the attributes perceived as essential for effective Advanced Teachers, the expected impacts of Advanced Teachers, and the interdependencies among and between these attributes and outcomes. The goal of this analysis is to better understand the complexities and interdependencies that characterize effective Advanced Teachers and help identify a common and connected set of criteria for the selection and evaluation of these roles.

3.3 Machine Learning

The RPP is also exploring the use of machine learning (ML) algorithms to supplement inferential statistical methods. The purpose of using ML is to further explore the relationships between selection criteria and reported impacts of the programs. While traditional statistical models are more appropriate for classical inference and hypothesis testing, some strengths of supervised learning algorithms are their ability to capture complex interactions and nonlinear relationships between variables, efficiently handle large-scale and high-dimensional datasets, and outperform traditional statistical models in prediction accuracy. The primary goal of embedding ML methods is to identify and develop acceptably accurate predictive models to aid in selection of teachers, as well as important selection criteria and program factors that traditional methods may have overlooked.

4 RESEARCHER-PRACTITIONER CONVENING & DISSEMINATION

On February 27-28, 2025, the Friday Institute will host a two-day convening for ATR researchers and practitioners to share and discuss research findings and practitioner lessons on the selection and evaluation of teachers serving in advanced teaching roles. As part of this convening, the research team will also engage practitioners in understanding the affordances and limitations of both commonplace (i.e. dashboards) and more novel learning analytics methods (i.e. ENA, ML). The research team has committed to external-facing deliverables for education practitioners and researchers as part of this RPP. Practitioner-focused deliverables (e.g., practitioner whitepapers, policy briefs, and a summative report) will be completed by June 30th, 2024. Dissemination of research findings at state and national conferences will begin in March and continue through December 31st. These findings will provide valuable insight into the program as well as the added value of Learning Analytics above and beyond more conventional methods typically used to inform educational programs and policies.

REFERENCES

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199-231. https://projecteuclid.org/journals/statistical-science/volume-16/issue-3/Statistical-Modeling--The-Two-Cultures-with-comments-and-a/10.1214/ss/1009213726.pdf

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Kellogg, S., Bausell, S., Pham, L., Thrasher, E.,& Young, T. (2024). *Advanced Teaching Roles: Evaluation Report*. Prepared for the North Carolina Department of Public Instruction. Raleigh, NC. https://fi.ncsu.edu/wp-content/uploads/sites/175/2024/11/ATR-Evaluation-Report-2024-FINAL.pdf

Krumm, Andrew, Barbara Means, and Marie Bienkowski. 2018. *Learning Analytics Goes to School*. Routledge. https://doi.org/10.4324/9781315650722.

Shaffer, D., & Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. *Handbook of learning analytics*.

A Human-centered Approach on Collecting Learning Analytics Insights in GitLab

Matthias Ehlenz
RWTH Aachen University
ehlenz@lbz.rwth-aachen.de

René Röpke
TU Wien
rene.roepke@tuwien.ac.at

ABSTRACT: This practitioner's report presents a human-centered approach to the collection of learner data from version control systems in software projects. With platforms like GitLab being widely used in computer science courses in higher education, instructors have access to the platform's usage data and can incorporate it in their grading. So far, metrics like lines of code, number of commits and commit message quality are used to draw conclusions on learners' performance, especially in comparative analyses for group projects. Meanwhile, GitLab offers way more functionalities and information for data analysis, e.g., for project management. This work aims at making this information more accessible and usable in the context of learning analytics. It focuses on the stage of data collection and the implementation of the necessary software components, enabling data analysis in the future.

Keywords: Software Projects, Version Control Systems, Data Collection, Learner Behavior

1 MINING REPOSITORY DATA FOR LEARNING ANALYTICS

Software Engineering is an important branch of Computer Science (CS) and plays a vital role in CS education. While there are always different flavors of tools, patterns, and principles in different domains, there are some core concepts, essential to CS curricula in higher education. Among these is the use of (decentralized) Version Control Systems (VCS) such as Git, provided through platforms like GitLab or GitHub. Oftentimes, these are taught in a practical, implicit fashion, where learners employ them in group projects. Here, programming skills are in focus and the VCS plays a subordinate role.

Meanwhile, instructors enjoy the benefits of the VCS when it comes to grading. For example, a git repository of a group project reveals a lot about individual contributions to the code as well as learners' activities in project management. Various approaches exist to garner information out of commits, merges, and other data inherent in any git repository. But in modern software projects, instructors should consider more metrics than mere lines of code, numbers of commits, and commit message quality when individual contribution to a project is to be measured or even graded. And while there is not much more to obtain from the raw data of the repository, used platforms offer much more information about the process, especially regarding issue management and teamwork.

Previous studies have explored the use of VCS data for educational analysis with varying objectives (e.g., Macak et al., 2021; Gitinabard et al., 2020; Matthies et al., 2018; Jagelid & Kindberg, 2018; Putra et al., 2018. For instance, Jagelid & Kindberg (2018) analyzed GitHub repository data from three programming courses but only examined a narrow subset of git activity (number of commits, number of issues, comments on issues). Similarly, Macek et al. (2021) performed post-project analysis and utilized course-specific data extraction methods looking at number and sizes of commits as well as the commit messages. The review of related literature shows that most works focus on commits and issues as they are core features of the VCS. Meanwhile, branching and merging in collaborative projects was not yet analyzed systematically. Further, no related work has used xAPI statements as format for the extracted logs; they either used custom formats or did not specify it in their publications. Lastly, none of the examined works implemented a human-centered design approach guiding the process for data collection, transformation and analysis (Buckingham Shum et al., 2019).

In this practitioner's report we present our approach to make this information more accessible and usable for learning analytics (LA). The goal is to harness all information from GitLab on learner's interaction with a project into a Learning Record Store (LRS) as xAPI statements, where it is available for further analysis and aggregation with additional learner-related data (e.g., learning management system logs or data obtained for multi-modal LA). In this paper we focus on the human-centered approach for data collection and the implementation of the necessary software components for data collection and transformation rather than the stage of data analysis and visualization. Results from the human-centered design workshop, the source code and architectural diagrams are published as open research data: https://doi.org/10.17605/OSF.IO/AWVXM

2 INSIGHTS FROM THE HUMAN-CENTERED DESIGN WORKSHOP

There are multiple options and approaches to collect learner data beyond the actual git repository and possibly also different strategies to map those interactions to meta-data definitions as blueprints for xAPI statements. Thus, we decided to include possible stakeholders in the design process early on. For this human-centered design approach, we gathered a focus group from multiple universities, institutions and disciplines and aimed to include participants with both differing scientific backgrounds as well as different levels of experience with using git in student projects. Eight researchers of different disciplines in Computer Science from three universities participated.

We designed an online workshop format of 120 minutes in June 2023 and prepared a collaborative board using the Miro platform as a creative space for this session. The session was structured in three distinct phases: First, a collection of potential (research) questions the stakeholder might strive to answer in software engineering education. In a follow-up brainstorming in the second phase, after a short introduction on xAPI with a sample statement, individual data units required to answer those questions were collected in the form of post-its to be sorted into the appropriate categories of verbs, activities and extensions as appropriate. For users not too savvy with the xAPI specification, a fourth category of "uncertain" has been provided to encourage everyone to contribute. In the final phase, a structured approach on a definitions set was unveiled, compared to the collected definitions and subsequently refined. Finally, the results were critically discussed and the workshop concluded. An overview after sorting, grouping and structuring the participants contributions is shown in Fig. 1.

In the end, the proposed structure with amendments can serve data sufficient for responding to nearly every question and idea raised in the design-thinking workshop by harnessing the capabilities of GitLab's webhooks, with two limitations: First, this approach is not exactly suited for semantic analysis, i.e. the analysis of commit content quality. This would have required to include actual commit payload within the statement, which would stretch the LRS capacity pretty fast and would require a huge overhead in resources. Still, should this be required in any later analysis, the statements contain all necessary information to pull the specific commit directly from the repository by conventional methods for further analysis. Second, there is currently no sufficient data (or scientific interest in our focus group) to collect insights on pipelines used in Continuous Integration/Continuous Development using the onboard means of GitLab. While potentially available, the mapping to xAPI would be challenging, as many of those events are actually system-triggered (or at least indirectly triggered by some other action recorded by the system).

Figure 1: Results of Human-Centered Design Workshop (High Resolution in OSF Repository)

Based on the requirements derived from the HCD workshop, GitLab Webhooks provide the most convenient way to gather the interactions required for further processing: Webhooks provide a minimally invasive option, are easily configured without the need to alter any content of the repository itself and offer convenient payload with all required information without the need to pull the repository each time. Still, webhooks come with a caveat: The end point has to be highly available, as an unsuccessful webhook might result in a lost payload. In response to this challenge, the software components described in the next section have been designed and implemented.

3 SOFTWARE ARTIFACTS TO COLLECT AND TRANSFORM THE DATA

For the data collection multiple scripts have been implemented and a set of meta data definitions derived from the requirements. The scripts can be roughly divided in two components: First, there is a receiver which does not much more than providing a web endpoint for the webhooks and storing the received payload in a Redis database. This database serves merely as a FIFO queue and in intended to increase scalability of the entire stack. The other scripts form the “worker” side and can be spawned in multiple instances in the long-term if required. We suppose this is only the case in large-scale instances with many repositories while needing close to real-time monitoring. For most use cases in academic settings, the current implementation should be sufficient.

The worker component pops data sets from the Redis queue and invokes further components based on the type of event that triggered the webhook. This has intentionally been implemented in a modular pattern with further extension in mind, as GitLab and other platforms provide further webhooks than the currently implemented, which focus on code and collaboration. Currently, all parts are provided as python scripts along with a complementary docker compose file for straight-forward deployment. Additionally, in future releases this might also decrease configuration efforts and help with scalability. As our aim is to contribute to Open and FAIR data, a metadata vocabulary for all required xAPI statements has been stored in our open registry at [blinded.forreview.com]. It is based on the results of the HCD workshop and available via a web frontend and as machine-readable JSON-API for interoperability. This registry is based on git itself, and thus is versioned and extendable.

4 CONCLUSION AND OUTLOOK

In this paper we presented a practitioner approach on collecting data on students' interactions in git-based software projects. We involved stakeholders early in the process and derived metadata on the data of interest first. Based on that, we evaluated webhooks as the most efficient way of accessing the data, due to their event-driven and minimally invasive nature. We implemented a potentially scalable and highly available software stack to harness the data, convert it into a domain-specific format for LA and aggregate it with other data by storing it in a common LRS. Currently, we are in our first field tests, using the described components in multiple student projects of different formats. We took this opportunity to report on the tool stack itself, as per usual, follow-up papers may focus on the data and respective analysis instead of technical details of the collection process.

This contribution focused solely on stage of data collection, but we conclude it with an outlook on (potential) next steps: The primary reason to collect data as we did is to improve our teaching concepts of and learning experiences in collaborative software development processes. More insight into all the important skills beyond code, like the use of branching, regular commits and pushes, issue management and so on provide a foundation for a fairer grading process. Individual contributions can be analyzed more fine-grained than before and thus be honored. But beyond that are even more opportunities. By collecting meta data in collaboration with stakeholders, we lay a foundation for truly findable, interoperable, and reusable data, that can be used to enhance research of success factors of software development in teams, of the learning process in software engineering, project management and use of VCS, and combined with other data even on student collaboration in a more general sense. We are sure that there are more research questions, where such data should be helpful. And we hope this work contributes to the long-term goal of FAIR data in LA; this can be considered a foundation. To collect data and make it accessible at scale is a community effort.

5 REFERENCES

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, *6*(2), 1–9. https://doi.org/10.18608/jla.2019.62.1

Gitinabard, N., Okoilu, R., Xu, Y., Heckman, S.S., Barnes, T., & Lynch, C. (2020). Student Teamwork on Programming Projects. What can GitHub logs show us? In *Proceedings of the 13th International Conference on Educational Data Mining* (pp. 409-416).

Jagelid, N., & Kindberg, L. (2018). Examination of the effects of student-teacher interactions on student commit patterns - In the GitHub environment. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-229720

Macak, M., Kruzelova, D., Chren, S., & Buhnova, B. (2021). Using process mining for Git log analysis of projects in a software development course. *Educ. Inf. Technol.*, 26, 5939-5969.

Matthies, C., Teusner, R., & Hesse, G. (2018). Beyond surveys: Analyzing software development artifacts to assess teaching efforts. In *Frontiers in Education Conference* (pp. 1-9). IEEE.

Putra, F., Santoso, H., & Aji, R. (2018). Evaluation of learning analytics metrics and dashboard in a software engineering project course. *Global J. Eng. Educ.*, 20(3), 171-180.

Adopting Learning Analytics in a Business Intelligence Framework

How a Small Public Institution in Rural America Implemented Learning Analytics

Carleigh B. Young
Uniformed Services University
carleigh.young.ctr@usuhs.edu

ABSTRACT: This presentation explores how a small rural institution adopted Learning Analytics within a Business Intelligence framework to address enrollment challenges amidst restricted budgets, outdated systems, and limited resources. By focusing on cultural transformation, data preparation, and systems integration, this approach highlights strategies, challenges, and recommendations for establishing sustainable analytics practices in resource-limited contexts.

Keywords: Reimagining Learning Analytics, Adopting Learning Analytics, Strategies for Scalable Learning Analytics, Student Information Systems, Technological Foundations, Business Intelligence Integration, Enrollment Management, Implementation Case Study

1 THE INSTITUTIONAL LANDSCAPE

As a small public institution in a rural community, the college is vital in providing educational access to students facing significant barriers. Its mission of inclusivity supports a diverse population, including first-generation college students, working adults, and economically disadvantaged individuals. However, the institution's rural location presents distinct challenges, such as geographic isolation limiting access to collaborative networks and advanced technologies and economic pressures straining resources. These difficulties, compounded by the looming enrollment cliff and declining revenue streams (Campion, 2020), have heightened financial and operational pressures, threatening the institution's ability to sustain its mission.

Operational inefficiencies have further hindered the college's effectiveness. Legacy systems operating in silos prolonged application processing times and resulted in frequent errors due to manual data entry. Misaligned departmental initiatives and resource constraints exacerbated these issues, leading to high staff turnover and instability during a period of rapid change. Recognizing the urgent need for transformation, the institution has prioritized campus-wide operational improvements, focusing on enhancing efficiency, addressing bottlenecks, and leveraging data to guide decision-making.

2 LEARNING ANALYTICS IN BUSINESS INTELLIGENCE

As the administrator overseeing enrollment management systems, I was tasked with addressing institutional challenges that directly impacted prospective and incoming students. My role was newly created as part of the institution's broader goal of improving systems and technology to enhance operational efficiency and better serve its mission. Recognizing the institution's unique challenges—limited institutional knowledge of analytics, fragmented systems, and resource constraints—I applied Learning Analytics (LA) principles within the existing Business Intelligence (BI) framework to address these needs. My leadership's trust in my judgment supported this approach and aligned with the enrollment team's goals for improving processes and outcomes. Concurrently pursuing a graduate degree in Learning Analytics, I was able to directly translate advanced methodologies into actionable strategies that optimized enrollment processes and established a scalable foundation for future applications in student success.

While BI tools provided essential operational clarity—tracking metrics like application processing times and enrollment targets—their scope was limited to aggregate reporting. To address this limitation, I integrated LA methodologies into the BI framework, aligning institutional goals with actionable insights into student behavior during enrollment. This integration bridged the gap between operational efficiency and a data-driven approach, setting the stage for future applications in student success. While these efforts primarily focused on prospective and incoming students, the data infrastructure created during this phase supported the future integration of Learning Management System (LMS) data. This design empowered student success teams to analyze academic engagement and retention patterns, further extending the impact of the foundational work.

3 THREE PILLARS OF DIGITAL TRANSFORMATION

Addressing the challenges of implementing LA within a BI framework at a small, resource-limited institution required focusing on three core areas: Cultural Transformation, Systems Transformation, and Implementation & Adoption. These pillars were selected to build a sustainable foundation for digital transformation, reflecting insights from organizational change frameworks like Kotter's Change Model (1996) and research on innovation management (Appio et al., 2021). Each pillar represents an interconnected step necessary to align culture, systems, and processes for effective analytics adoption.

3.1 Cultural Transformation

The adoption of analytics at the institution required a significant cultural shift from intuition-based decision-making to data-driven strategies. Resistance stemmed from fears of increased workload, potential misuse of data, and skepticism toward new technologies, compounded by a history of frequent leadership changes and evolving systems. Building trust and fostering collaboration were essential to overcoming these barriers. Transparent stakeholder engagement demonstrated how analytics aligned with institutional goals, such as enrollment growth and student success. Department-specific workshops and personalized onboarding highlighted practical benefits, including reduced administrative burdens, while data literacy programs empowered staff to view analytics as a valuable resource. Despite these efforts, challenges persisted, including mistrust in data accuracy and concerns about the adequacy of training, exacerbated by the overwhelming pace of institutional changes.

3.2 Systems Transformation

Fragmented, outdated systems posed a significant obstacle to analytics adoption, leaving insights incomplete and inaccessible. When I assumed the role, system documentation was outdated, confidence in existing technologies was low, and institutional knowledge about operations was minimal. Addressing these barriers required consolidating legacy systems into centralized platforms to streamline workflows and improve data accessibility. Data cleansing ensured accuracy, while workflow automation reduced manual errors and increased efficiency. The integration of new technologies with legacy systems involved extensive use of APIs, middleware, and tailored custom solutions. However, resource constraints, including limited personnel and funding, made overhauling systems challenging. Iterative testing was crucial to ensure compatibility and maintain operational stability throughout the transition.

3.3 Transformative Implementation & Adoption

Even with cultural and systems readiness, the success of analytics adoption depended on a carefully planned implementation strategy. Iterative development, stakeholder feedback, and continuous refinement ensured that tools were both user-friendly and relevant. High-impact use cases, such as improving enrollment workflows and identifying at-risk students, served as starting points. Tools were refined based on user feedback, and comprehensive training programs built user confidence. Establishing support channels provided stakeholders with ongoing resources, while regular feedback loops allowed tools and processes to evolve with institutional needs. Challenges included balancing short-term wins with the long-term goal of creating a sustainable analytics framework. Furthermore, varying levels of data literacy among stakeholders necessitated tailored communication and education to ensure engagement and effective adoption.

4 OUTCOMES

Key advancements included the development of enhanced dashboards to track applicant engagement metrics such as submission patterns, communication responses, and decision timelines. These tools supported data-informed recruitment strategies and predictive yield modeling, resulting in higher conversion rates and more efficient processes. Centralized and standardized enrollment systems provided a scalable foundation for future analytics efforts. Quantifiable outcomes included a 33% reduction in admissions processing time, a 60% increase in lifecycle efficiencies, significantly higher campaign engagement rates, and administrative tasks reduced by at least five full-time equivalent hours.

While students were not directly involved in the implementation, the operational improvements significantly enhanced their academic journeys. Faster admissions processes reduced delays, allowing students access to advising and academic planning resources earlier. Improved data accuracy and streamlined workflows ensured institutional resources were deployed more effectively, enhancing service quality. These outcomes demonstrate how operational advancements focused on prospective students can indirectly support broader student success initiatives, laying the groundwork for future, more student-centered analytics.

5 LESSONS LEARNED

Several key lessons emerged from this initiative, providing valuable guidance for future analytics projects. Early successes, such as streamlining admissions processes, showcased the tangible value of analytics and helped build momentum for broader adoption. These initial wins were instrumental in fostering trust and enthusiasm among stakeholders, paving the way for further advancements.

Stakeholder engagement proved critical in overcoming resistance and cultivating trust in analytics tools. Identifying departmental champions to advocate for analytics adoption strengthened cultural buy-in and ensured a collaborative approach to implementation. Additionally, phased implementation and iterative development were essential in minimizing disruptions, enabling tools and processes to evolve based on stakeholder feedback and institutional priorities.

As the sole individual overseeing both technical development and strategic execution, the extensive scope of responsibilities presented significant challenges, including periods of burnout. While enabling innovation, the high degree of autonomy granted by leadership occasionally led to role ambiguity and misunderstandings about decision-making authority. These dynamics required consistent communication to clarify responsibilities and ensure alignment across stakeholders. Relying on a single individual for such a critical initiative highlighted the vulnerabilities associated with limited personnel capacity in resource-constrained environments. Despite these challenges, collaborative problem-solving, transparent communication, and leadership support were instrumental in mitigating resistance and fostering a more cohesive and productive project environment.

Sustainability emerged as a cornerstone of long-term success, requiring ongoing investment in training, system maintenance, and data governance. Developing and maintaining dashboards was particularly resource-intensive, necessitating iterative refinement to balance stakeholder needs with institutional constraints. Regular maintenance protocols ensured sustained accuracy and relevance, while continuous updates kept tools user-friendly and impactful. These efforts underscored the importance of strategic planning and resource allocation to address immediate operational needs while supporting long-term institutional goals.

6 CONCLUSION

The implementation of Learning Analytics within a Business Intelligence framework at a resource-limited institution demonstrated the transformative potential of data-driven strategies in higher education. By focusing on cultural transformation, systems transformation, and implementation and adoption, the institution overcame challenges such as fragmented systems, limited resources, and resistance to change. These efforts prioritized operational improvements, aligning analytics goals with institutional priorities to enhance efficiency and indirectly improve student experiences. Key lessons include the importance of streamlining workflows and integrating systems to ensure data accuracy, engaging stakeholders to foster a collaborative culture, and adopting a gradual, feedback-driven approach to implementation. Sustainability emerged as a critical focus, requiring ongoing investment in training, maintenance, and governance to support scalable and impactful analytics frameworks.

REFERENCES

Appio, F. P., Frattini, F., Petruzzelli, A. M., & Neirotti, P. (2021). Digital Transformation and Innovation Management: A Synthesis of Existing Research and an Agenda for Future Studies. *Journal of Product Innovation Management*, *38*(1), 4–20. https://doi.org/10.1111/jpim.12562

Campion, L. L. (2020). Leading Through the Enrollment Cliff of 2026. TechTrends, 64(3), 542–544. https://doi.org/10.1007/s11528-020-00492-6

Kotter, J. P. (1996). *Leading Change*. Harvard Business School Press.

Evaluating learning analytics implementations: three approaches for practice

Tom Olney
The Open University, UK
tom.olney@open.ac.uk

Approaches for evaluating learning analytics implementations are as varied as the definitions of success they attempt to measure. Practically, some have the potential to provide bounded insight and can be simple to apply, whilst others offer deeper, holistic understanding but can be complicated to manage. In order to extend practice, and develop more effective, responsible, and successful implementations it is suggested that it would be advantageous for practitioners to have broad knowledge of several methodologies that could be used for evaluating learning analytics implementations.

Keywords: learning analytics implementations, evaluation approaches

1 INTRODUCTION

If, as Motz & Morrone (2023: 147) claim, institutional learning analytics (LA) implementations 'tend to be characterized by ambitious but uniformed and uncarefully planned initiatives' which would benefit from more critical and academic assessments then a consideration of a variety of approaches to the evaluation of LA implementations would seem appropriate and hopefully informative for practitioners. This report briefly presents three such approaches, chosen because of their familiarity to the author. It describes their use in practice, and provides a short discussion about the advantages and disadvantages of each for further consideration.

2 THREE APPROACHES AND EXAMPLES OF USE IN PRACTICE

2.1 Technology Acceptance Model

The Technology Acceptance Model (TAM) claims that the intention to use, or not use, technologies is influenced by two main factors: the *perceived usefulness* and the *perceived ease of use* of said technology. *Perceived usefulness (PU)* is defined as 'the degree to which a person believes that using a particular system would enhance their job performance', and *perceived ease of use (PEU)* refers to 'the degree to which a person believes that using a particular system would be free of effort'. (Davies, 1989; 320) Therefore, it follows that, the extent to which teachers start and continue to use technological innovations such as learning analytics dashboards is positively correlated with their perceived agreement with these two factors. The influence of PU and PEU has been shown to predict user acceptance of technological initiatives over several years in a wide variety of fields that include education.

In practice: It is not hard to see how a simple instrument could be constructed to collect TAM data from users using a Likert scale and a set of sample statements inspired by reference to Davies (1989).

For example, statements to agree/disagree with could be as straightforward as ‘using the learning analytics dashboard will improve my teaching practice’ (PU), and ‘I find it easy to get the learning analytics dashboard to do what I want it to do’ (PEU). Findings could conceivably be used for a number of purposes such as, justifying the scaling of a pilot LA implementation, reviewing the use of a chatbot, or correlating against another factors. For example, Rienties et al (2016) utilised TAM to develop a seven-question feedback instrument and used it to survey 95 teachers after an opportunity was provided for them to explore a set of LA implementation dashboards in groups and construct their own understanding and knowledge about them. The findings from the survey suggested that whilst the teachers responded positively to the *perceived usefulness* of the LA implementations, they were less positive about the *perceived ease of use*. This finding helped to justify the design and refinement of ongoing professional development activities for LA implementations at that institution.

2.2 Shadow practices

The concept of *shadow practices* has roots in the tradition of social informatics. Social informatics is an approach that views the users of technologies as active social actors whose technological practices constitute a wider socio-technical system. In this view actions are not reliant solely, or even primarily, on LA implementations themselves, but on the entire sociotechnical network, which includes colleagues, access to resources, previous practice, institutional culture and so on, in which the users are located.

In such a context *shadow practices* are described as ‘undesired or unanticipated interactions’ between a user and a LA implementation such as a LA dashboard (described by the authors as a ‘decision support system data dashboard’ or DSS-DD). When a comparison of the users' anticipated practices (typically defined by the design of the dashboard or the expectation of the developers) and the users’ actual practices (as reported by teachers) reveal a difference or disconnect, *shadow practices* emerge to fill this space and ultimately result in ‘the limited or non-use of DSS-DD for decision making processes’ (McCoy & Rosenbaum, 2019:371).

In practice: Designing an evaluation approach around the concept of *shadow practices* could be utilised for almost any LA implementation. For example, Olney et al (2021) used *shadow practices* to explain the limited or non- use of a new LA dashboard introduced into a distance learning setting. Firstly, a review was conducted on the design of the dashboard as well as the collection of reports, literature and communications between the institution and the teachers to establish the expected or anticipated practice. Then, 30 higher education teachers who had been using the LA dashboard were interviewed using a semi-structured instrument constructed on social informatics principles. This provided information about the actual practices. Comparing these two led to the identification of three shadow practices, also observed in the interviews, which could be viewed as either undesired or unanticipated, and some candidate mechanisms that help to explain them.

2.3 Theory of Practice Architectures

The Theory of Practice Architectures (TPA) is described as ‘an account of what practices are composed of and how practices shape, and are shaped, by the arrangements in which they are enmeshed in a site of practice’ (Mahon et al, 2017). As such, TPA takes a site-orientated, ontological approach to investigating practices. In education, practitioners engage in practices which contain specialist discourse (*sayings*), activities and work (*doings*), and exist in a complex ecology of power structures and individuals (*relatings*). Further, such practices are prefigured and shaped by *arrangements* that exist across three mediums: material-economic, cultural-discursive & social-political.

In practice: In the language of TPA the wider field of education is defined as a *practice* that can contain smaller, more discreet *projects*, of which implementing LA could be considered one, and is located in one or more *sites of practice*. For example, Olney & Wood (2023) used TPA to investigate the use of LA in the faculty of a large HEI to try and explain what it meant to 'do' learning analytics there and identify the *arrangements* that enabled or constrained that work. Using semi-structured interviews and document analysis approaches they uncovered the *material-economic arrangements* that made the *doings* of the *project* possible. This included such things as the time available to teachers or the actual availability of the LA dashboards and spreadsheets themselves. Similarly, they identified *cultural-discursive* arrangements by exploring the specialist language or discourse that prefigured, constrained, or enabled the *sayings* of the *project*. Like many other *projects* within education, LA implementations have developed their own specific references and language that is used by practitioners to describe and justify what it is, and how it is practiced. This included how dashboards were referred to, and how the definitions of certain data sources were shared. These sayings were contained in documentation as well as live in discussion between those engaged in seeking meaning from LA. Thirdly, they investigated what *social-political arrangements* shaped and prefigured the *relatings.* This was concerned with how humans related to one another, behaved in the roles they were representing, existed in the power structures that the organization provided, and the experience they brought to group or team environments. Since LA dashboards are not usually built by the same people that are required to use them, and the responsibility for their use has not necessarily been clearly articulated, this interaction is relevant.

Analysis of how the *practices* and *arrangements* come to hang together in a particular *site of practice* and *project* under investigation can allow for an exploration of how one impacts on another and the identification of new and progressive approaches.

2. COMPARISON & DISCUSSION

TAM has been criticised for being too simplistic and although only the original conceptualisation of TAM is presented here, successive models, referred to as TMA2 and TMA3, have introduced more antecedents in order to try and refine the model (Mariykan & Papagiannidis, 2023). Yet, the simplicity of the model, particularly in data collection, is also perhaps it's greatest asset. A recent systematic literature review showed PU and PEU continue to be widely accepted 'to be antecedent factors that have affected acceptance of learning with technology' (Granić & Marangunić, 2019). However, despite this, it is hard to escape the fact that TAM does not allow much room for an interpretation of LA implementations that questions the intrinsic value of the technology itself, or the complex contexts into which it is often being introduced. TAM also makes no claim as to identifying or explaining the kinds of unexpected outcomes that often occur when humans interact with technology, result in *shadow practices,* and contribute to explaining the low acceptance or take up rate of LA. Evaluating LA implementations using social informatic approaches such as *shadow practices* provides a more nuanced set of findings but can require more collaboration between developers and researchers to be effective, and more detailed and time consuming data collection. One step further is utilising TPA which, as a theoretical lens, can be used to develop a situated view of LA practice that is contextualised and far more robust. This report concludes that TPA is the gold standard, addressing questions that have not yet been properly answered about what it means to 'do' learning analytics, from a personal and organizational point of view, as well as how it is being done, and why it is done in that way (Bennett et al, 2018).

REFERENCES

Bennett, S., Lockyer, L., & Agostinho, S. (2018). Towards sustainable technology-enhanced innovation in higher education: Advancing learning design by understanding and supporting teacher design practice. *British Journal of Educational Technology*, 49(6), 1014-1026. https://doi.org/10.1111/bjet.12683

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340. https://doi.org/10.2307/249008

Granić, A. and Marangunić, N. (2019), Technology acceptance model in educational context: A systematic literature review. British Journal of Educational Technology, 50: 2572-2593. https://doi.org/10.1111/bjet.12864

Mahon, K., Kemmis, S., Francisco, S. & Lloyd, A. (2017). Introduction: Practice theory and the theory of practice architectures. Chapter 1 in K. Mahon, S. Francisco & S. Kemmis (Eds.) (2017) *Exploring Education and Professional Practice: Through the lens of the theory of practice architectures*. Singapore: Springer

Marikyan, D. & Papagiannidis, S. (2023) *Technology Acceptance Model: A review*. In S. Papagiannidis (Ed), TheoryHub Book. Available at https://open.ncl.ac.uk / ISBN: 9781739604400

McCoy, C. and Rosenbaum, H. (2019), Uncovering unintended and shadow practices of users of decision support system dashboards in higher education institutions. Journal of the Association for Information Science and Technology, 70: 370-384. https://doi.org/10.1002/asi.24131

Motz, B. A., & Morrone, A. S. (2024). Wild brooms and learning analytics. *Journal of Computing in Higher Education*, *36*(1), 145-153. https://doi.org/10.1007/s12528-023-09353-6

Olney, T., & Wood, C. (2023). The evolution of learning analytics practice over six years at the Open University, UK: what are the arrangements that enable or constrain this practice? *ICERI2023 Proceedings*, 449-455. https://doi.org/10.21125/iceri.2023.0173

Olney, T., Walker, S., Wood, C., & Clarke, A. (2021). Are we living in LA (P) LA Land? Reporting on the practice of 30 STEM tutors in their use of a learning analytics implementation at the open university. *Journal of Learning Analytics*, *8*(3), 45-59. https://doi.org/10.18608/jla.2021.7261

Rienties, B., Herodotou, C., Olney, T., Schencks, M., & Boroowa, A. (2018). Making sense of learning analytics dashboards: A technology acceptance perspective of 95 teachers. *International Review of Research in Open and Distributed Learning*, *19*(5). https://doi.org/10.19173/irrodl.v19i5.3493

Learning Analytics as a Measure of Educational Gain

Authors: Camille Kandiko Howson and Charlotte Whitaker
Imperial College London, UK
c.howson@imperial.ac.uk

ABSTRACT: A national review of measurement of students' learning gain in England identified student engagement as the greatest challenge. Students did not see the value, have the time or interest, or were sufficiently made aware of opportunities to complete additional tests and surveys. To be able to explore the educational gain of students, without additional burden upon them, at Imperial College London we have explored over the past four years how to get the most out of the data we already have about students—primarily through their data trails across the institution and engagement with virtual learning platforms. This presentation explores the potential role of AI in supporting this and the engagement with students and academic staff to capture data relevant to them.

Keywords: educational gain; learning gain; outcomes; teaching; learning; assessment; partnership

1 ARTIFICIAL INTELLIGENCE AND LEARNING ANALYTICS

Artificial Intelligence (AI) and analytics are often thrown together as solutions to a range of higher education problems, but often with little specific detail. But AI and analytics can offer the potential to address an ongoing challenge in higher education: what are students gaining from their time and investment in higher education?

A decade ago, efforts to measure learning gain in England commenced, focusing on assessing the changes in students' knowledge, skills, work-readiness, and personal development, as well as improvements in specific practices and outcomes within particular disciplinary and institutional contexts (Kandiko Howson, 2019). These initiatives were driven by the government which sought to determine the value it was deriving from the investment in higher education (Department for Business, Innovation and Skills, 2016). The work was originally overseen by the Higher Education Funding Council for England (HEFCE) and later transitioned to the Office for Students (OfS).

Through a series of pilot projects, three key dimensions of learning gain were identified: (1) measures of general cognitive development, encompassing students' knowledge and critical thinking; (2) measures of soft skills development, including affective indicators of attitudes, how students feel, and behavioral measures of their engagement; and (3) measures of employability and career readiness, primarily focusing on behavioral indicators of students' activities in preparation for the workforce.

Despite these efforts, significant challenges arose in the measurement of learning gain, including low student participation in supplementary assessments and surveys, variations in students' starting points, and the lack of a standardized baseline across different courses and institutions. These difficulties led the OfS to discontinue the learning gain program. Nevertheless, the concept has resurfaced in the latest iteration of the national Teaching Excellence Framework, which now emphasizes educational gain, although no specific methodology for measuring this has yet been defined (OfS, 2022).

AI and advanced analytics present a potential solution to addressing the challenges in measuring learning gain. Significant progress has been made in the field of learning analytics, which involves the measurement, collection, analysis, and reporting of data related to learners' progress and the contexts in which learning occurs. With the emergence of generative AI models, learning analytics can expand further, incorporating a broader range of data sources to enhance the understanding of student learning.

Evaluation of the learning gain pilot projects indicated that multiple metrics are required to capture the full diversity of student learning in higher education. These measures—affective, behavioral, and cognitive—are highly interrelated, underscoring the need for integrated approaches to assessment. Advanced analytics can support the development of multiple models of learning gain, revealing relationships and patterns across these diverse metrics.

The pilot projects also demonstrated that student learning is multidimensional, varying not only across different domains but also over time and direction. To effectively assess learning gain, it is essential to track student progress throughout their academic journey at multiple points. Analytical models offer the capability to account for initial entry measures and diverse learning pathways, moving beyond a simplistic, linear conception of learning to a more nuanced understanding of individual student progress.

2 ANALYTICS AND EDUCATIONAL GAINS AT IMPERIAL

We attempted to tackle this challenge at an institutional level. Imperial College London, an urban, research-intensive Science, Technology, Engineering and Mathematics (STEM)-based university. To enable measuring students' educational gain and progress in their learning across the institution, Imperial is investing in its institutional data infrastructure and analytical capability. As part of wider institutional data strategy, a Unified Data Platform is being developed to link data across the institution and support the development of learning analytics to offer data-derived insights to enhance learning, teaching, assessment and the experience of staff and students.

"Learning analytics is the application of analytic techniques to analyze educational data, including data about learner and teacher activities, to identify patterns of behaviour and provide actionable information to improve learning and learning-related activities" (van Harmelen & Workman, 2012, p. 5). Higher education institutions can leverage analytics to transform many activities, including enrolment, student support, alumni engagement, financial aid administration and other learning and operational functions. An institution-wide approach is necessary to ensure that data subject rights are respected; data is used appropriately, ethically and transparently; shared with permission at appropriate levels; and to deliver parity of experience for all students.

Most higher education learning analytics dashboard systems are predicated on predicting drop-out and creating early warning systems; to streamline services and minimise costs; or to support regulatory reporting. By contrast, Imperial has aimed to use learning analytics to offer an enhanced student experience, and to better know and support our students.

Analytics systems have provided a unique opportunity for Imperial to develop and deliver on its strategic priorities for education: to empower students; to facilitate high quality staff-student interactions in order to maximise student success; and to offer a world-leading evidence-based educational experience. The STEM-based academic staff at Imperial are uniquely qualified to utilise their disciplinary analytical and mathematical modelling skills to gain insights from educational data to research and evaluate their own teaching and learning contexts. Similarly, our students have opportunities to reflect and gain insights into their own educational experience, as well as engage in opportunities to design research projects using learning analytics data for use in course projects.

Applying analytics to wider questions of outputs from higher education allows us to evidence student educational gain, engagement and progress and show the data in dashboards to both staff and students, allowing them to be active agents in their own learning. This initiative echoes our educational approach and integrates educational expertise, disciplinary research and methodological skills from our academic faculty in areas such as machine learning and AI, in partnership with students.

3 PARTNERSHIP APPROACH TO USING LEARNING ANALYTICS DATA

The ethical use of learning analytics is essential. We developed a partnership project with staff and students to develop guidelines and policies for the ethical use of learning analytics and the application of AI. We received institutional funding for a year-long project supporting students as co-researchers. Staff and students worked in collaboration to conduct focus groups with students about data use, analytics and interventions. The focus group protocol was adapted from the SHEILA project student instrument (Tsai, Moreno-Marcos, Tammets, Kollom, & Gašević, 2018) and applied to the institutional context. The protocol followed the original ten question prompts under eight themes about students' awareness of data collection and processing by the institution, how learning analytics might support them as students, how staff should act on analytics, managing control over their data and ethical concerns. Institutional ethical clearance was obtained to collect the data.

Six focus groups were conducted, involving three to seven students. While the focus groups produced extensive data, here we highlight that students' main preferred uses of learning analytics data. These were to check their progress on their educational goals and to gain wider understanding of their learning, including patterns and about their development in relation to their peers. Students were supportive of AI-based analysis to explore trends and patterns but not to replace human interventions. These insights helped to set boundaries on appropriate uses of AI (e.g. mapping and analysis) and where staff should have an active role (e.g. discussions implications of data insights, pastoral support). This provided insight for the use of learning analytics data, but students' also noted the need for boundaries, and the desire to limit the data to their learning environment, and not to include data beyond that, for example extracurricular activities or their wider social lives.

4 USING EDUCATIONAL GAINS

Data about learning does not inherently provide benefit to students; it depends on how the data is used for enhancement. The findings from the study with students provided insight into how, in partnership with students, we could use student learning data to support students to understand what they have gained from their higher education experience. We identified the importance of making the data understandable and presented in ways a broad audience could understand. If the analytics and outputs are too complex and confusing for staff and students to understand and apply them, they will have limited impact. Students and academic staff need support, advice and guidance to use educational gain data.

Students reported the importance of integrating measures in disciplinary contexts—which fits with findings from research on learning gains that there are wide variations in engagement across disciplinary, professional and regulatory bodies. Fields such as Medicine with standardised outgoing exams are ahead of non-professionally oriented subjects with less prescriptive outcome goals.

Drawing on AI models and the feedback from students at Imperial, the linking of educational gains and learning analytics are underway. This includes capturing baseline data on students' AI skills and understanding and tracking this over time. AI tools are being used to link students' individual assessments with their wider course engagement, including attendance, virtual learning platform and video recording use. Research is also underway capturing affective measures of learning gain, including soft skills such as resilience, through exploring students' engagement with online coursework and automated feedback platforms. Future work is planned for a 'live' syllabus with clearly mapped intended learning outcomes that can be linked to data from virtual learning platforms to show students explicitly what skills they have gained. Further connections with sites such as LinkedIn can more directly link students and their skills with employers and the labour market.

This is work in progress. The technology is already largely in place; however, as seen across higher education, Imperial is still developing connected, up to date student information systems and learning platforms. However, using the data trails left by students may provide a more sustainable way to capture gains from higher education, bypassing the need for additional surveys or exams. This also offers a scalable approach across whole institutions, delivering on greater parity of experience across the student body.

REFERENCES

Department for Business, Innovation and Skills. (2016). *Success as a knowledge economy: Teaching excellence, social mobility & student choice*. London: Department for Business, Innovation and Skills.

Kandiko Howson, C. B. (2019). *Final evaluation of the Office for Students learning gain pilot projects*. Bristol: Office for Students.

Office for Students (2022). Regulatory advice 22. Guidance on the Teaching Excellence Framework (TEF) 2023. OfS 2022.60. www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

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).

van Harmelen, M., & Workman, D. (2012). Analytics for learning and teaching. JISC Cetis Analytics Series, 1(3). Retrieved from http://publications.cetis.org.uk/wp-content/uploads/2012/11/Analytics-for-Learning-and-Teaching-Vol1-No3.docx

STELA Live
Implementing Learning Analytics for Student Success

Author(s): Angelica Risquez, Teresa Curtin, Chris McInerney, Michael P O'Brien, Donal Palcic, John Walsh, Mohd Fazil, Sarah Gibbons
University of Limerick
angelica.risquez@ul.ie

ABSTRACT: This presentation describes a large cross-institutional project run at an Irish university which used learning analytics (LA) capabilities to enhance student engagement. The project was a first for the institution in several ways, as it took a centralized and coordinated approach to the utilization of LA for the first time. The purpose of this project was to establish the infrastructure and framework necessary to provide learning interventions that could mitigate the risk of students underperforming in selected first-year modules. The predictive models utilized a combination of demographic data, continuous assessment scores, and student engagement data derived from the university's Virtual Learning Environment (VLE). After building these models with a cohort of 8000 students over four academic years, a pilot intervention was designed in which about 2000 students were notified about their likelihood of success based on the model's predictions and were directed toward available academic support services. The outcomes of this pilot intervention were evaluated and the findings are shared to offer insights on how learning analytics can be applied in higher education to support student success, particularly in large, diverse cohorts.

Keywords: student engagement, blended learning, machine learning, academic performance, first year classroom, large classes, at-risk students

1 INTRODUCTION

Disengagement among students in large, first-year cohorts has become a significant issue (Bowden, 2022). However, while vast amounts of data are collected in academic institutions, the challenge lies in effectively utilizing that data to have a tangible, positive impact on students' academic experiences. Educational institutions curate huge amounts of data in the form of 'learning traces' (Gašević et al., 2015) to enable this, but we need to analyse what variables are relevant in each context and find solutions that are simple and scalable enough in practice.

1.1 Background

In Ireland, sectoral guidelines for the application of learning analytics (O'Farrell, 2017) stress the importance of developing learning analytics strategies in partnership with students and staff to ensure that the primary focus remains on benefiting learners. Following the approval of a learning analytics policy at the university where the project was conducted, the project was initiated with three main objectives: (a) Data Access and Scoping Phase: this involved consultations with data owners to explore potential data access, define data management plans, and ensure compliance with data protection regulations; (b) Baseline Analysis: this phase involved analyzing historical data from four large first-year modules, combining demographic information from student information systems with continuous assessment and VLE engagement data to identify the key variables that predict student performance; and (c) Design of the Intervention: based on the baseline analysis, a protocol was developed to provide personalized feedback to students that supported their academic journey and encouraged them to engage with relevant resources.

2 METHODOLOGY

The baseline analysis phase was centered around four first-year cohorts, each representing different disciplines across Business, Science and Humanities, with between 350 and 600 students in each module, and chosen from the emerging community of practice in learning analytics. The data collected spanned four academic years and included demographic data, continuous assessment scores, and engagement metrics from the VLE that included information on frequency of access logs and performance in mid-term examinations. Each module was assessed differently, with three modules using traditional grading scales and one module assessed on a pass/fail basis. The data was preprocessed, and variables were encoded to create a model that could predict whether a student was likely to perform at a high, moderate, or at-risk level. The attention-aware BiLSTM-CNN model was particularly suited to this task because it could account for the sequential nature of student engagement data over time, identifying patterns of behavior that correlate with academic performance (Fazil et al, 2024). For the implementation phase with the 2022 cohort (Table 1), the predictive models were trained only using five weeks of student engagement data from the VLE, along with demographic and academic information from the university's academic registry. Students were then notified mid-semester about their likelihood of succeeding in the module and were referred to appropriate academic supports, such as tutoring services, learning centers, or meetings with their instructors. Finally, students were surveyed at the end of the semester to assess their level of satisfaction with the intervention and to collect their suggestions for improvement.

Table 1: Dataset implementation phase

Module	Student Category			Total Student
	High-performer(H)	Moderate-performer(M)	At-risk (F)	
Module-1	373	958	219	1550
Module-2	208	1689	169	2066
Module-3	399	1468	538	2405
	Pass		Fail	
Module-4	207		1590	1797

Notably, the model's predictive accuracy was higher in some modules compared to others. In one side of the spectrum, Module 4 achieved an accuracy of 86% thanks to it being based on a pass/fail basis. Of the modules based on a more nuanced performance category, Module 2 achieved an accuracy of 73%, and saw the best balance between student engagement with assessments and the VLE. On the other side, Module 3 recorded the lowest accuracy of 58%, attributed to the fact that students in this module engaged primarily with third-party tools outside the VLE, making it harder to predict their performance based on VLE interactions.

3 RESULTS

3.1 Student Engagement Patterns

A major focus of the intervention was on improving student engagement, especially among at-risk students. The project monitored the frequency and quality of student interactions with the VLE before and after the intervention, and the results varied across the modules. The results show that following the intervention, student engagement increases significantly for courses 1 and 3, and it also increases moderately for course 2. However, student engagement slightly slows down following the intervention in course 4.

The timing of the intervention was key. For most modules, the feedback was provided between weeks 7 and 10 of the semester. This period was carefully chosen as it coincided with mid-semester assessments or deadlines for major assignments, when students would be more receptive to feedback about their performance. More detailed analysis showed that prior to receiving feedback, the at-risk students engaged with the VLE significantly less than their high-performing peers, and as expected, the high-performing students were generally more proactive in using the VLE to access course materials, submit assignments, and review their progress. In contrast, at-risk students often lagged in engagement, which was identified as a key predictor of their eventual performance in the course. The frequency of VLE interactions for at-risk students was often half that of high-performers. After the intervention, high-performing students showed a marked increase in VLE engagement following the intervention. These students responded well to the feedback that predicted their success, using it as motivation to continue their high level of engagement. The intervention served as a confirmation of their efforts, prompting them to stay consistent or even increase their participation. Moderate-performing students also increased their engagement, though to a lesser extent. The intervention seemed to alert these students to their potential to improve, leading them to seek additional academic resources. At-risk students, while benefiting from the feedback, did not show the same level of increased engagement. Although the intervention prompted some engagement, their participation in VLE activities remained notably lower than that of their peers. This is a critical finding, as it suggests that while predictive feedback may motivate some students, a more targeted approach may be required to encourage at-risk students to engage more fully with the learning resources.

3.2 Student Feedback on the Intervention

The response rate to the survey (n=239) was highest among high-performing students and lowest among at-risk students (which ranked between 6% and 19.5% in different courses). This finding is consistent with the engagement trends seen earlier—high-performing students were more likely to interact with all aspects of the course, including feedback mechanisms, while at-risk students were more disengaged overall. A significant majority of students, particularly those who were high performers, found the feedback to be useful and motivating. These students appreciated the clarity and timeliness of the feedback, which helped them stay on track academically. Moderate-performing students were slightly less enthusiastic but still found the feedback beneficial. They noted that the feedback helped them identify areas where they could improve, though some expressed a desire for more specific guidance or felt that it did not represent an accurate picture of their level of engagement. At-risk students, while the least likely to respond to the survey, provided mixed feedback. Some at-risk students indicated that the feedback was useful, but others expressed frustration, feeling that the feedback did not offer them the support they needed to make significant improvements. Students were also asked whether the feedback motivated them to engage with specific academic resources, such as lectures, VLE materials, and tutoring services. Over 80% of students reported that the feedback motivated them to use the VLE more frequently. This aligns with the earlier findings that post-intervention VLE activity increased, particularly among high and moderate performers. Approximately 75% of respondents indicated that they were more likely to attend lectures and tutorials after receiving the feedback. Interestingly, fewer students (around 25%) reported that the feedback encouraged them to seek additional support from academic advisors or learning centers. This indicates that while the feedback successfully motivated students to engage with core academic resources, more targeted efforts are needed to increase the use of supplementary support services. In terms of future interventions, many students expressed interest in receiving similar feedback in other modules. However, some students suggested improvements to the feedback mechanism, such as providing more detailed, personalized advice on how to improve in specific areas of the course.

4 CONCLUSIONS AND IMPLICATIONS

The results of this project show that learning analytics, when applied thoughtfully, can have a significant positive impact on student engagement and academic performance. Yet, the differences in predictive accuracy suggest that predictive models work better in certain academic settings than others, particularly when the blended learning and assessment approach heavily relies on the use of the VLE. Also, future work could focus on refining the predictive models to better account for moderate performers, who often fluctuate between success and failure. The intervention led to increased engagement with the VLE and other course materials, particularly among high-performing and moderate-performing students. However, engaging at-risk students remains a challenge, and more personalized, targeted interventions to help these students improve their performance are called for. This could involve real-time feedback throughout the semester and adaptive learning resources that are responsive to the individual student’s needs and engagement patterns, and a tighter integration with the personal advisor system.

Finally, the survey results suggest that students appreciate receiving feedback on their academic performance and would like to see such interventions expanded to other courses. The feedback provided was particularly useful for motivating students to stay engaged with the course, although more work is needed to encourage students to take full advantage of available support services.

REFERENCES

Bowden, J. L. H. (2022). Analogues of engagement: Assessing tertiary student engagement in contemporary face-to-face and blended learning contexts. *Higher Education Research and Development*, *41*(4), 997-1012. https://doi.org/10.1080/07294360.2021.1901666

Fazil, M., Rísquez, A., & Halpin, C. (2024). A Novel Deep Learning Model for Student Performance Prediction Using Engagement Data. *Journal of Learning Analytics*, 1-19. https://doi.org/10.18608/jla.2024.7985

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

O’Farrell, L. (2017). Using Learning Analytics to Support the Enhancement of Teaching and Learning in Higher Education. National Forum, Dublin.

Leveraging Knowledge Graphs and Large Language Models to Track and Analyze Learning Trajectories

[1]Yu-Hxiang Chen, [1]Ju-Shen Huang, [1]Jia-Yu Hung, [1]Chia-Kai Chang *
[1]National Central University
ckchang@ncu.edu.tw

ABSTRACT: This study addresses the challenges of tracking and analyzing students' learning trajectories, particularly the issue of inadequate knowledge coverage in course assessments. Traditional assessment tools often fail to fully cover course content, leading to imprecise evaluations of student mastery. To tackle this problem, the study proposes a knowledge graph construction method based on large language models (LLMs), which transforms learning materials into structured data and generates personalized learning trajectory graphs by analyzing students' test data. Experimental results demonstrate that the model effectively alerts teachers to potential biases in their exam questions and tracks individual student progress. This system not only enhances the accuracy of learning assessments but also helps teachers provide timely guidance to students who are falling behind, thereby improving overall teaching strategies.

Keywords: Learning Trajectory, Knowledge Graph, Large Language Model

1 INTRODUCTION

Tracking and analyzing students' learning trajectories has become crucial in contemporary education (Ellis et al., 2014). Educational service platforms have already been implemented in industry, and academic research focuses on developing tools to explain and observe learning behaviors. For example, in 2012, Anna Lea Dyckhoff, Dennis Zielke, and others proposed the Exploratory Learning Analytics Toolkit (eLAT), which provides teachers with a user-friendly interface to explore students' learning activities and assessment results through data visualization, allowing them to reflect on and improve teaching strategies.

Similarly, José Michel Fogaça Vieira et al. proposed various methods of representing learning trajectories. However, these approaches are often limited to data display and are not widely applicable across different academic subjects. Therefore, we devised a strategy based on knowledge graph analysis that enables teachers to grasp students' learning progress better. For instance, it can monitor the extent of students' curriculum coverage and observe changes over time, providing insights into their learning trajectories.

This study introduces a system that leverages knowledge graphs built from large language models (LLMs) to analyze learning materials and track students' progress. By transforming the materials into a structured list of nodes and relationships, individualized knowledge graphs are generated for each student, integrating exam data to assess academic performance and teaching effectiveness. Applied to an introductory Python programming course at a national university in Taiwan, the system identified gaps in exam coverage and student progress, helping teachers adjust the scope of exams and providing targeted support to students lagging.

2 METHOD

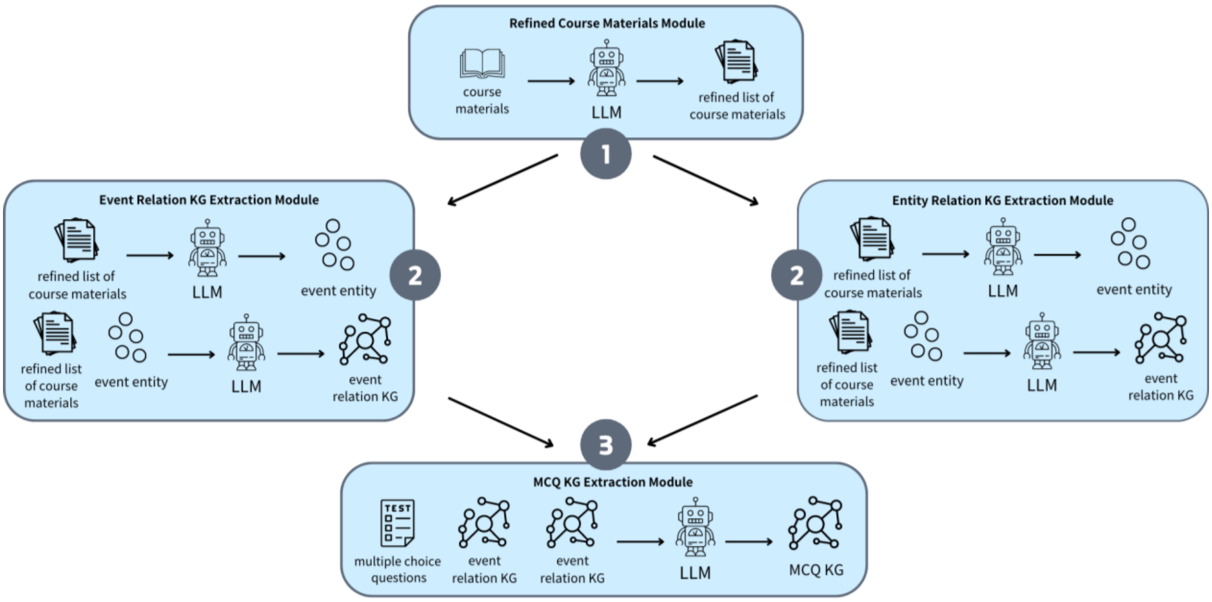


Figure 1: Multiple-choice Questions Knowledge Graph Construction Framework Diagram

As shown in the first step of Figure 1, we converted the course materials into text files. We used a Large Language Model (LLM) to create a refined list by removing unnecessary conjunctions and particles. This structured the content into concrete and meaningful data, making it more organized and interpretable for the subsequent construction and analysis of the knowledge graph, thus improving the accuracy and effectiveness of learning trajectory measurement.

In the second step of Figure 1, following Ling Feng Zhong et al., we extracted nodes (entities) from the refined list of materials, categorizing them as general nodes (people, objects, time, places) and event nodes (specific events). Using Shuang Yu et al.'s method, we utilized the LLM to automatically generate relationships (entity relations) between nodes, enhancing efficiency and accuracy. We then input both the general nodes and refined list into the LLM to create relationships, mainly verbs or prepositions connecting two nodes. Event nodes were also processed similarly to establish causal and sequential connections, forming a complete knowledge graph.

In the third step of Figure 1, we tracked students' learning trajectories using quizzes, midterms, and final exams to build their knowledge paths. Multiple-choice questions and answers from these assessments were input into the LLM and knowledge graph. The Chain-of-Thought (COT) process in Prompt Engineering enabled the LLM to match these questions to the corresponding edges in the knowledge graph, allowing a detailed mapping of student learning progress.

Once we had determined which edges in the knowledge graph corresponded to each question, we could create a personalized knowledge graph for each student to record their learning trajectory. When students correctly answered a question, we marked the corresponding edge in their knowledge graph. Through this complete learning trajectory-building process, we could study the changes in students' knowledge paths and use them to evaluate their abilities and the effectiveness of the course. Below are four aspects that can be explored in research:

- Changes in the knowledge graph correspond to different score groups.
- Identifying key knowledge points to determine which gaps lead to difficulty answering specific questions, thus causing learning bottlenecks.
- Changes in the coverage of knowledge points across the class are needed to assess whether students have mastered all the knowledge covered by the course after the teacher's instruction.
- Evaluating whether the test comprehensively assesses the knowledge students learn in the classroom.

3 CASE STUDY

3.1 Knowledge Node Coverage Warning and Cognitive Bias System for Instructors

Our framework model was implemented in an experimental research study on an introductory Python Programming course at a national university in Taiwan. A total of 47 students participated fully in the study. During the course, each student completed three standardized and unbiased multiple-choice assessments designed by the course instructor based on the curriculum and related to fundamental Python programming skills. In Figure 2, the color differences reflect knowledge point coverage across testing phases. Green dots in the Pre Test represent foundational knowledge assessed before instruction, while purple and blue dots in the Midterm and Post Test indicate knowledge introduced or reinforced during teaching. This highlights curriculum progression and helps identify gaps or newly emphasized concepts.

The study results are shown in Figure 3, which highlights part of the knowledge graph depicting the distribution of knowledge nodes across the three assessments. The percentage of knowledge nodes covered was 6.1% in the pre-test, 8.8% in the mid-term exam, and 6.1% in the post-test. It was noted that the knowledge nodes in the pre-test and post-test overlapped significantly with those in the mid-term exam. This suggests that the instructor’s selection of knowledge points may have been influenced by selective attention, a phenomenon where focus is unintentionally directed toward specific areas, potentially overlooking other important knowledge nodes. As a result, the three assessments did not adequately cover the entire scope of the course content. Our system allows for early detection of such gaps, helping instructors adjust the scope and content of future assessments.

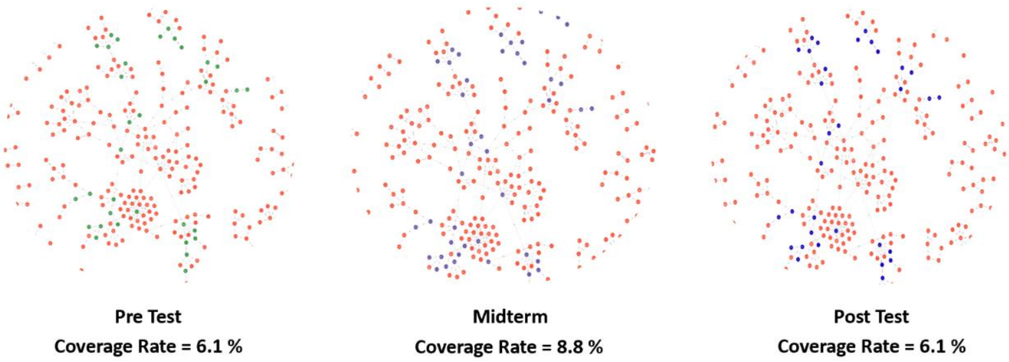


Figure 2: Intersection of Knowledge Graph Coverage for Three Assessments and the Overall Python Course.

3.2 Student Performance Warning System

We also analyzed the growth rate of knowledge nodes for individual students across the three assessments and compared it to the class average. Figure 3 compares a student's knowledge graph coverage with the class average. The student's coverage rate for knowledge nodes in the mid-term exam was 79.4%. In the figure, the red areas represent the course's knowledge points, the yellow areas indicate the knowledge points already mastered by the student, and the green areas indicate the knowledge points the student lags the class average in mastering. Using our system, students can identify areas requiring improvement and receive targeted alerts based on their level of knowledge deficiency.

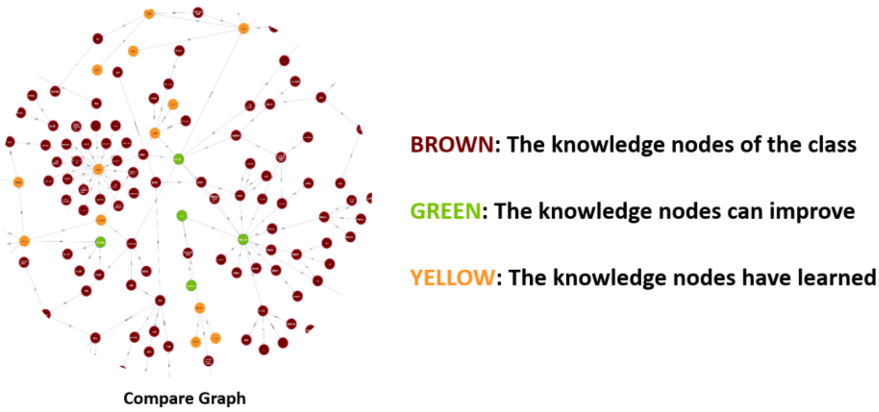


Figure 3: Knowledge Graph Comparison Between an Individual Student's Mastery and the Class Average.

4 CONCLUSION

This paper proposes a knowledge graph-based method using LLM to track and analyze students' learning trajectories, addressing the issue of incomplete coverage in traditional assessments. Individualized knowledge graphs are generated by converting teaching materials into structured data and integrating them with students' test results, mapping learning progress, and identifying gaps. The system helps educators adjust exam content, track performance, and support students, improving assessment accuracy and teaching strategies.

REFERENCES

Ellis, A. B., Weber, E., & Lockwood, E. 2014. The case for learning trajectories research. In *PME-38 and PME-NA-36 Joint Meeting Proceedings*, Vancouver, Canada, July 15–20, 2014.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U.2012. Design. and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society*, 15(3), 58-76.

José Vieira and Luciana Zaina. 2021. Learning Trajectories Visualizations of Students Data on the Computational Thinking Context. In *Proceedings of the 32nd Brazilian Symposium on Computers in Education*, November 22, 2021, Online, Brasil. SBC, Porto Alegre, Brasil, 705-717.

Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. 2023. A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys*, *56*(4), 1-62.

Yu, S., Huang, T., Liu, M., & Wang, Z. 2023. Bear: Revolutionizing service domain knowledge graph construction with llm. In *International Conference on Service-Oriented Computing* (pp. 339-346). Cham: Springer Nature Switzerland.

GenAI for teaching and learning: a Human-in-the-loop Approach

Wei Qiu
Centre for the Applications of Teaching & Learning Analytics for Students (ATLAS)
Nanyang Technological University (NTU), Singapore
qiuwei@ntu.edu.sg

Maung Thway
ATLAS NTU, Singapore
maung.thway@ntu.edu.sg

Joel Weijia Lai
ATLAS NTU, Singapore
joellai@ntu.edu.sg

Fun Siong Lim
ATLAS NTU, Singapore
LIM_Fun_Siong@ntu.edu.sg

ABSTRACT: This paper presents a human-in-the-loop development and implementation of a Socratic generative artificial intelligence (GenAI) tutor for undergraduate statistics courses. GenAI has potential to personalize and encourage desired deep learning behaviors in a diverse student population. However, thorough planning and evaluations are essential to ensure responsible use of AI. Our systematic approach started with a GenAI tutor designed with course coordinators and instructors, followed by a trial phase involving student volunteers and instructors. The GenAI tutor was piloted in a real class setting, with data collected on the conversation logs, the experiences of both students and instructors, as well as the resulting outcomes. This approach fosters trust in GenAI and facilitates continuous improvement. The findings contribute to the ongoing discourse surrounding the use of AI in learning environments, with a particular focus on enhancing human capabilities.

Keywords: Human-in-the-loop, generative AI, pedagogically designed chatbot, analytics dashboard, technology-enhanced learning

1 INTRODUCTION

The integration of generative artificial intelligence (GenAI) into education is transforming the way learning support can be designed and delivered. GenAI shows huge potential to offer personalized learning to students at scale. However, ensuring that these agents effectively contribute to the intended learning outcomes requires thoughtful design and continuous human involvement. Before any large-scale deployment, their effectiveness in terms of accuracy of responses, quality of engagement, and learning gains must be rigorously designed and studied.

2 THE CHATBOT DESIGN APPROACH

Figure 1 depicts our approach to implementing GenAI tutors responsibly. Our journey begins with identifying the courses most in need of additional teaching support. Statistics is a core subject for the large and diverse undergraduate student population at our university. In the statistics course of this paper, 90% of the students have historically requested for tutoring support. There are approximately 600 students in each cohort, making it a prime candidate for experimentation.

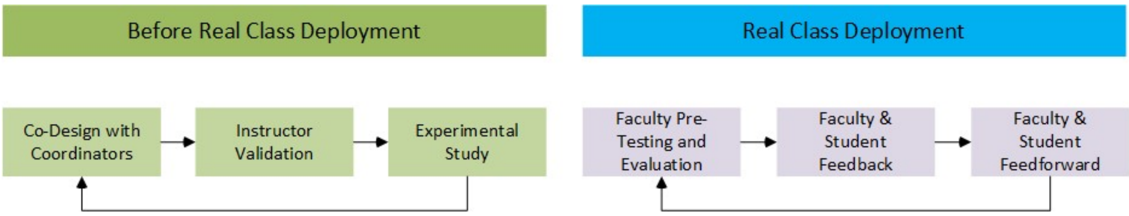


Figure 1: Human-in-the-Loop GenAI Deployment Approach

Co-Design with Coordinators: The coordinators wanted the GenAI tutor to be pedagogically designed as a Socratic GenAI tutor to facilitate inquiry-based, step-by-step problem-solving to deepen students' conceptual understanding of statistic concepts taught in the course. As the university's learning analytics center, the development team also wanted to give the instructors the ability to monitor their students' use of the GenAI tutor in their learning.

As such, the design of the GenAI tutor comprised two main parts: a student-facing GenAI tutor and a faculty-facing dashboard. Developed as a prototype using Streamlit web application framework (https://streamlit.io/), the GenAI tutor aided students in learning statistics through a ChatGPT-like interface. As shown in Figure 2, it was prompt engineered and grounded with curated course materials to enable the desired Socratic pedagogical behavior and learning experience.

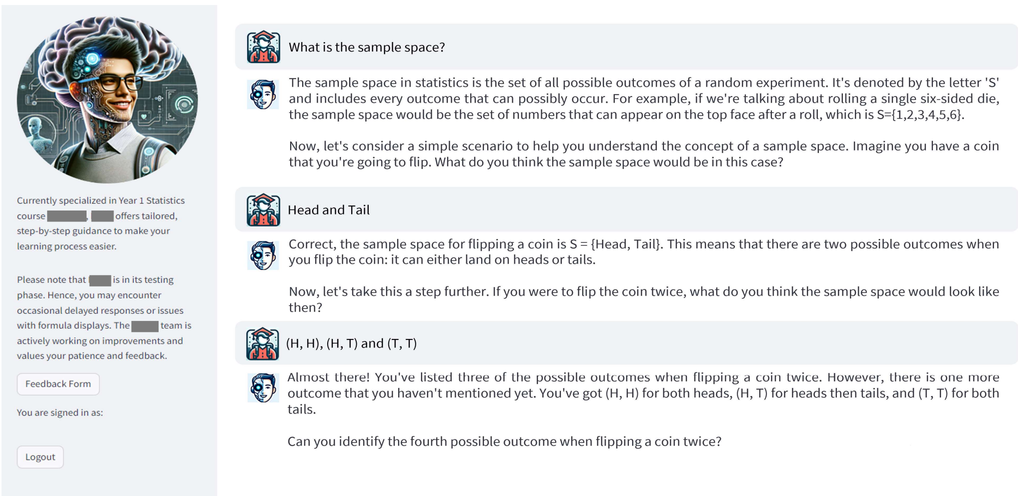


Figure 2: GenAI tutor which is 1) grounded with course materials, and 2) prompt-engineered to deconstruct questions and guide students.

Instructor Validation: Before experimenting the GenAI tutor with students, course instructors were recruited to assess the accuracy and the quality of the GenAI tutor's response (Qiu W. , et al., 2024). These subject matter experts were asked to evaluate the chatbot's answers against two predefined criteria identified by the course coordinators – accuracy of the response and quality of the engagement. The evaluation revealed an accuracy rate exceeding 80%, with a high level of agreement among evaluators, reflected in a Randolph kappa score of 0.704. At the same time, we noted from the instructor validation and feedback that the Socratic engagement needed to be further refined with prompt engineering.

Experimental Study with Students: Before piloting the GenAI tutor in a real class, we conducted an experimental study to understand its impact on student learning. Following approval from the

Institutional Review Board, we invited students who were either weak in statistics or had never taken statistics courses before to participate in an experimental study. Students were asked for their consent at the beginning of the study and were informed that they could withdraw from the study at any given time. Moreover, participants who provided feedback were made aware that their responses would be recorded for research purposes and were assured that their personal information would remain confidential throughout the process.

The three-week study randomly assigned 45 student volunteers into a control and experimental group. The control group interacted with the baseline chatbot using GPT-4-Turbo. The experimental group used our Socratic GenAI tutor with the same GPT-4 model. Both chatbots are grounded with the same curated materials. The key difference was that the baseline chatbot was not prompt engineered, and students were unaware of their assignment. Pre and post-tests were administered on 12 topics learned and post-study feedback was collected along with conversation log data. The experimental group showed higher learning gains compared to the control group (Qiu W. , et al., 2024), especially when they questioned the GenAI tutor's response and asked for application examples to test their knowledge (Lai, et al., 2024). The Socratic approach encouraged students to engage more often and on more complex topics. Finally, student feedback indicated a preference for our GenAI tutor, given its explanations and guidance, but suggested that its response time can be improved. Overall, the findings gave us the confidence to deploy the GenAI tutor in a real class with some refinements to ensure faster performance.

At the same time, the student-chatbot interactions were collected, analyzed, and presented in a faculty- dashboard. In essence, we wanted the dashboard to be both a learning support tool and a resource for improving teaching strategies based on real-time analytics. We surveyed the coordinators and other faculty (n = 15) on the measures that they want presented on the dashboard. The popular choices included the frequency students use the GenAI tutor, the topics asked, the cognitive level of student questions, and student feedback on the GenAI tutor's answers. These feedback were incorporated into the design of the faculty dashboard as shown in Figure 3.

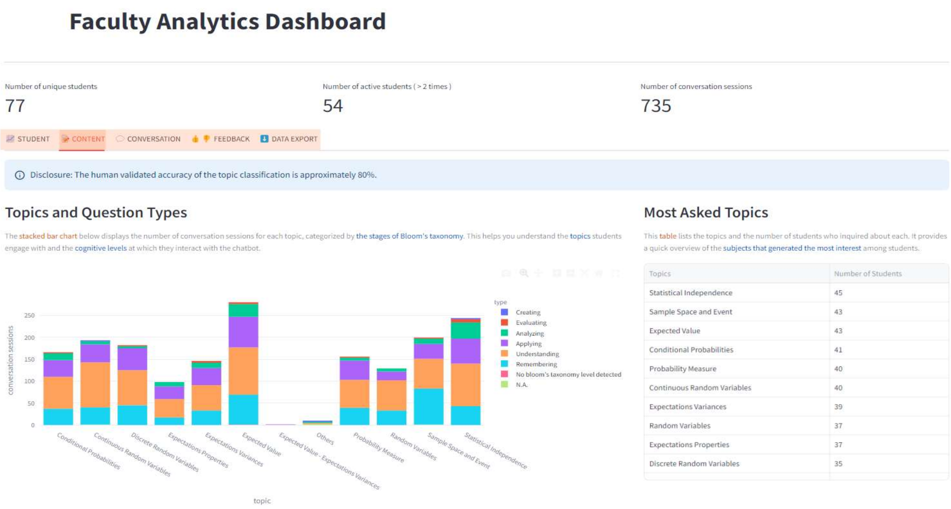


Figure 3: The Content section view of the faculty analytics dashboard.

3 THE EXPERIMENT IN PRACTICE

In preparation for a large-scale rollout, we collaborated extensively with course coordinators and instructors to conduct a pilot study within real classroom environments for a few weeks. This pilot was an essential component of our human-in-the-loop approach, ensuring that the GenAI tutor was not only effective in a controlled environment but was also aligned with both faculty expectations and student learning needs in a real class.

Faculty Pre-Testing and Evaluation: The course coordinator and instructors were again invited to thoroughly test the GenAI tutor's performance since a new model, GPT-4o, was released. They assessed its responses across a variety of content areas, ensuring that the GenAI tutor's outputs were accurate, contextually appropriate, and pedagogically sound. They also gave feedback on the faculty dashboard, which led to the addition of a new data export function and several quality-of-life improvements.

Faculty/Student Feedback and Feedforward: Throughout the pilot, the course coordinators and instructors worked with the development team to address major technical and performance issues. At the same time, they monitored and provided elaboration on responses from the Socratic GenAI tutor that students downvoted. This feedback loop was essential for not only addressing hallucinations but also ensuring the completeness of the responses, with the ultimate goal of improving the quality of the responses generated by the chatbot.

At the time of writing, the GenAI tutor is still being piloted. We intend to collect student and faculty experience as well as performance data to improve its usefulness for statistical learning before rolling out for the entire semester and other similar courses. We anticipate that feedforward will be an ongoing iterative process as the technology advances and as learning needs evolve.

4 CONCLUSION

Our experience in implementing a pedagogically designed GenAI tutor demonstrated both the potential and importance of involving humans when using AI to support learning. The encouraging findings from the various stages of implementation may have shown the power of AI, but this is only possible with human involvement every step of the way. We hope that our approach offers ideas for other institutions to discuss how best to implement GenAI for education responsibly.

ACKNOWLEDGEMENTS

We want to express gratitude to our ATLAS team members Ms. Nurabidah Binte Jamil, Ms. Chit Lin Su, Mr. Samuel Ng Soo Hwee, Mr. Lei Zhang, and Mr. Daniel Hernandez for their invaluable contributions to the development and study of the GenAI tutor.

REFERENCES

Lai, J., Qiu, W., Thway, M., Zhang, L., Jamil, N., Chit, L., . . . Lim, F. (2024). Leveraging Process-Action Epistemic Network Analysis to Illuminate Student Self-Regulated Learning with a Socratic Chatbot. *EdArXiv Preprints*.

Okonkwo, C. A.-I. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*. doi:https://doi.org/10.1016/j.caeai.2021.100033

Qiu, W., Chit, L., Jamil, N., Ng, S., Chen , C.-M., & Lim, F. (2024). "I Am Here To Guide You": A Detailed Examination of Late 2023 Gen-AI Tutors Capabilities in Stepwise Tutoring in an Undergraduate Statistics Course. *18th International Technology, Education and Development Conference*.

Qiu, W., Chit, L., Jamil, N., Thway, M., Ng, S., Zhang, L., . . . Lai, J. (2024). A Systematic Approach to Evaluate the Use of Chatbots in Educational Contexts: Learning Gains, Engagements and Perceptions. *EdArXiv Preprints.*

Developing Metrics to Evaluate Student Engagement in MOOCs: A Learning Analytics Approach

Inma Borrella
Massachusetts Institute of Technology
inma@mit.edu

Eva Ponce-Cueto
Massachusetts Institute of Technology
eponce@mit.edu

ABSTRACT: This report presents the development and implementation of learning analytics metrics to evaluate student engagement in a MOOC-based program. The initiative aims to address the limitations of traditional evaluation methods by introducing a three-tiered system of metrics: course monitoring metrics, course evaluation metrics, and program-level metrics. These metrics offer practical insights into student behavior, support timely interventions, and guide course design improvements. Our findings highlight the critical impact of data-driven decision-making in online education, with implications for improving student outcomes and program management in MOOCs.

Keywords: Learning analytics, online education, MOOC performance evaluation, student engagement analytics, educational data dashboards, program management.

1 INTRODUCTION

Massive Open Online Courses (MOOCs) have expanded educational access, providing flexible learning opportunities to diverse global audiences. However, tracking student progress and engagement within these environments remains a challenge, as traditional metrics offer limited insights into the nuanced learning pathways students navigate in these courses (Hadi & Gagen, 2016).

To address these gaps, we developed and implemented a set of metrics designed to capture real-time, post-course, and program-wide performance data. This initiative builds on data-driven approaches to provide educators, course designers, and program administrators with comprehensive insights into student engagement, ultimately supporting timely interventions and long-term course improvements.

2 DEVELOPMENT OF METRICS

2.1 Course Monitoring Metrics

The first set of metrics provides real-time insights into course activity. MOOC instructors value good visualizations that provide information beyond just grades (Stephens-Martinez et al., 2014). By tracking enrollment, verification status (students who paid for the option to earn a certificate), weekly graded assignment completion, average grades, and forum activity, course administrators can monitor student participation as it happens. This real-time data helps identify immediate issues, such as declining assignment submission rates or low forum engagement, allowing for swift interventions, such as sending targeted communications to re-engage students. They also enhance visibility of course performance and student progress and provide a reporting standard for the whole team.

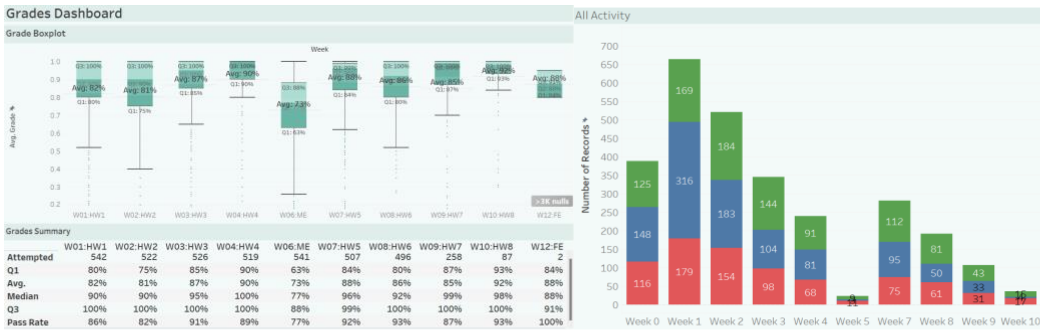


Figure 1: Course Monitoring Metrics Dashboard

2.2 Course Evaluation Metrics

After course completion, post-hoc metrics offer a deeper analysis of student behaviors and outcomes. Key metrics include the number of enrollments, verified students, active participants, and "zombies" (a zombie is a student who paid for the option to get a certificate but did not complete any graded assignment in the course). We also measure completion and certification rates, and calculate important ratios such as conversion, retention, and pass rates. Additionally, we include metrics that reflect students' perception of the course, such as NPS, collected through feedback surveys. This set of metrics is vital for reflecting on course performance and identifying areas for improvement in future iterations.

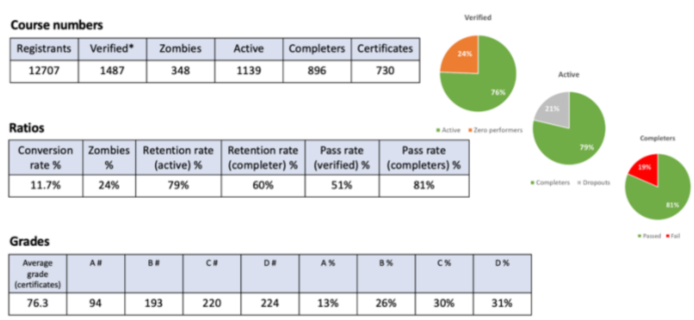


Figure 2: Course Evaluation Metrics dashboard

2.3 Program-Level Metrics

The third set of metrics supports program-wide management by tracking trends across multiple course runs and cohorts. Delivered through an interactive dashboard, these metrics include enrollments, verification rates, pass rates, program credentials, and student demographics (e.g., geography, gender, age). This dashboard provides program administrators with the ability to monitor trends over time, such as variations in enrollment and verification numbers, and completion and passing rates per course, allowing for data-driven adjustments at the program level.

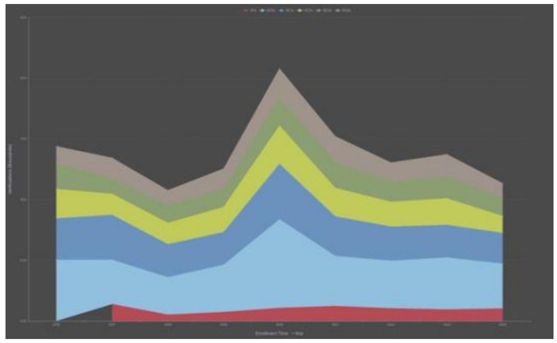


Figure 3: Verified learners per course (color-coded) over the years

3 IMPLEMENTATION

The metrics were implemented in a MOOC-based program hosted on the edX platform, consisting of five MOOCs and a final exam leading to a program credential. Data on enrollment, verification, progress in graded assignments, and demographics were extracted from edX and processed using Python and Excel. Visualizations and dashboards were created with Tableau. And an interactive dashboard was created supported by technology from the MIT CAVE Lab, to offer retrospective insights into course and program performance.

Instructors and course administrators were actively involved in shaping these metrics to align with real-world instructional needs. Regular feedback loops between staff and the metrics development team ensured that the data captured was relevant and actionable. Despite the successful implementation of these metrics, data latency remained a challenge, as data feeds from the edX platform require time to refresh, affecting the timeliness of real-time in-course interventions.

4 RESULTS

4.1 Course Monitoring Metrics

Real-time metrics enabled us to identify and address issues as courses were running. For instance, unusually low grades on specific assignments prompted immediate reviews and corrections, improving student outcomes. These metrics also allowed for early detection of disengagement, such as drops in forum activity or assignment completion. In response, targeted communication campaigns to re-engage students were deployed. Results from these interventions revealed that that while modifications to course content had a significant positive impact on students' engagement (Borrella et al., 2022), targeted email communications had no impact on dropout rates (Borrella et al., 2019).

4.2 Course Evaluation Metrics

Post-hoc analyses provided comprehensive insights into course performance. One notable finding was the high "zombie" rate of 30-35% in the first course, compared to 20-25% in subsequent courses. (A "zombie" is defined in section 2.2.) This led to a review of the first course structure, transitioning from an instructor-paced to a self-paced course, and the development of strategies to enhance early engagement through the implementation of a student-facing learning analytics dashboard. Further research could explore other targeted interventions, such as more interactive content or personalized learning pathways, to reduce early dropout rates.

Additionally, a “funnel effect” was observed in all courses, with engagement declining after the midterm exam, which accounted for 35% of the final grade. To address this, the midterm exam content was thoroughly reviewed, making sure it covered foundational knowledge rather than marginal topics (Borrella et al., 2022). Post-midterm pacing was also adjusted, improving retention and completion rates.

4.3 Program-Level Metrics

Program-wide metrics revealed trends in student demographics and outcomes. For instance, students from Europe, South America, and Oceania had higher pass rates (around 66%) compared to those from North America and Asia (around 50%). These regional differences suggest variations in student commitment, which may be influenced by cultural or educational factors, and it underscores the need for tailored support based on student demographics, further enhancing the global accessibility of MOOCs. The dashboard also highlighted that, despite declining overall enrollments, the verification rate had increased, indicating sustained interest in obtaining formal credentials. This trend may be driven by the increasing number of companies using microcredentials to upskill their workforce.

5 CONCLUSION

The development and application of comprehensive learning analytics metrics in this MOOC program has significantly improved our ability to track and analyze student behavior, engagement, and performance. By leveraging real-time, post-hoc, and program-wide data, we were able to implement timely interventions, enhance course design, and improve program management. While challenges remain, particularly around data latency and early-course engagement, our approach demonstrates the value of data-driven decision-making in online education.

Our metrics-based approach to course monitoring and program management provides a model for other institutions seeking to enhance their online programs through data-driven insights.

REFERENCES

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2022). Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education*, *179*, 104412.

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2019). Predict and intervene: Addressing the dropout problem in a MOOC-based program. *In Proceedings of the sixth ACM conference on Learning @ Scale* (pp. 1-9).

Hadi, S. M., & Gagen, P. (2016). New model for measuring MOOCs completion rates. *Proceedings of the European MOOC Stakeholder Summit 2016*, Research Track 95.

Stephens-Martinez, K., Hearst, M. A., & Fox, A. (2014). Monitoring MOOCs: which information sources do instructors value?. In *Proceedings of the first ACM conference on Learning@ Scale* (pp. 79-88).

How engagement analytics is helping retain Foundation Year students at Keele University

Simon Rimmington
Keele University
S.Rimmington@keele.ac.uk

Rachel Maxwell
StREAM by Kortext
Rachel.Maxwell@kortext.com

ABSTRACT: The Keele University Foundation Year (KUFY) is a route to undergraduate study typically for students from underrepresented backgrounds with mixed experiences in their journeys through the education system. The KUFY recognises the strategic value of learner analytics in addressing student engagement and retention. Analysing student data facilitates implementation of targeted interventions that boost engagement over the entire student journey. This **presentation** considers how data-driven insights enhance student success and learning experiences, examining the implementation of learner analytics, considerations surrounding student data, and their impact. These initiatives lead to evidence-based strategies that enhance inclusivity and student support, aligned to wider university priorities.

Keywords: data-driven insights, engagement analytics, foundation year, inclusivity, retention, student engagement, targeted interventions

1 ENHANCING STUDENT SUCCESS USING ENGAGEMENT ANALYTICS

Over the past two decades, the UK's higher education sector has undergone significant transformations, driven by a steady increase in student numbers and a shift in demographic composition (Wong & Hoskins, 2019, 2022). This dynamic landscape has created a more complex and competitive environment (Jewitt 2020), highlighting the need for universities to adapt their strategies and leverage innovative tools to maximise the best possible student experience and outcomes to meet the diverse needs and expectations of all student groups and cohorts. Engagement dashboards and learner analytics are increasingly seen as powerful tools to enhance student success with academic institutions adopting early warning systems to efficiently identify students at risk (Rimmington, 2024).

In this context, student engagement encompasses active involvement in educational and social activities, representing the practices and attitudes that contribute to successful teaching and learning in higher education, "a desirable set of practices and orientations in students which should be worked towards or encouraged for teaching in higher education to be deemed successful" (Gourlay and Oliver, 2018), and significantly impacts student achievement (Kahu, 2013).

The Foundation Year at Keele University is tailored to students from diverse backgrounds, providing the academic and personal skills needed for degree success and bridging the gap between secondary education and university. The size and complexity of the KUFY brings opportunities and challenges in fostering inclusive and equitable learning experiences. Although student engagement has rightly gained prominence in recent years, it remains conceptually blurred (Appleton, Christenson and Furlong, 2008; Reschly and Christenson, 2012; Azevedo, 2015). Engagement is complex and individual, covering behaviour, emotion and cognition. Engagement analytics have been instrumental in developing a proactive and strategic approach to providing targeted and timely student support.

2 THE IMPETUS FOR CHANGE

Although student retention is regulated in the UK, the associated B3 metric (OfS, 2024) around successful outcomes for all fails to capture the multifaceted nature of student success, as retention is often a symptom of deeper underlying issues, particularly those related to mental health. Keele University has strived to use engagement data to meaningfully understand and address these concerns, intervene early where appropriate to meet the moral commitment to provide the support students need and deserve and, additionally, to support university finances.

To overcome the complexities of fragmented student data and poor staff data literacy, an engagement analytics pilot study was launched. The chosen platform addressed the need for data democratisation, enabling frontline staff to easily access and act early based on insights from student engagement data (Poirier et al., 2021, Yang and Li, 2020). To gain stakeholder buy-in, the platform presented clear, user-friendly information and empowers users with varying data literacy levels to make informed decisions.

2.1 Addressing early challenges through a pilot study

In 2021/22, the successful pilot of the StREAM engagement analytics platform enabled Keele to gain confidence in understanding the engagement profiles of the FY cohort and the platform was rolled out university-wide in 2022/23. All Academic Mentors were trained in its usage, and the data, along with other indicators like assessment submissions, were utilised to identify and support students needing academic assistance. Further refinement of engagement profiles followed in 2023/24, alongside targeted interventions for student cohorts based on these profiles. During the pilot, 73% of students with a ‘High’ or ‘Good’ engagement profile passed all their modules; an additional 10% required reassessment in only one or two assessments. In contrast, only 10% of students with a ‘Low’, ‘Very Low’, or ‘None’ engagement profile passed all their modules, with 53% failing all modules.

Effective communication and access to reliable data are crucial for promptly identifying and addressing student needs. However, the scale and diversity of the KUFY programmes pose challenges in maintaining consistent, reliable, and real-time communication and data flow. Regardless of the underlying causes of disengagement, prolonged inactivity risks perpetuating a downward spiral, fostering negative sentiments toward university life.

3 A TWO-PRONGED APPROACH – INDIVIDUAL AND COHORT

To translate this compelling knowledge into effective actions, a data-driven dual framework that combined individual and cohort approaches was employed to address reasons for poor engagement and ensure a nuanced approach to supporting students. The KUFY Centre assessed cohort-level risk and managed communications while, as the first point of contact for academic concerns, Academic Mentors undertook 1:1 meetings and liaised with Student Experience and Support Officers for non-academic issues. Analytics data enabled more transparent, meaningful conversations, fostered student ownership of their learning and enabled mentors to provide personalised feedback. The platform's flexibility in analysing engagement data over different periods allowed mentors to uncover underlying factors influencing student progress and provide timely support.

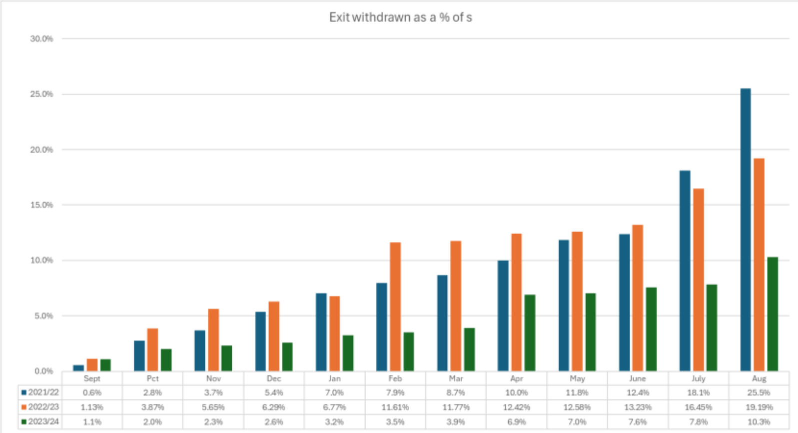


Figure 1: Cumulative percentage withdrawal rates

While the data did not explain the reasons behind a student's early 'disengagement', initiating communication could potentially prompt students to seek support from alternative networks such as friends and family. Comparing monthly withdrawal data, Figure 1 indicates a significant decrease in student withdrawal rates since the implementation of the retention and engagement framework, with the pilot also showing that the engagement data is compelling in identifying students by their engagement profile and relating it to academic performance. This year has seen the largest absolute number of students successfully progressing into their respective schools.

4 LESSONS LEARNED

While cohort-level engagement data provides proxy insights into student behaviour and improvement areas (Muir et al. 2019), this does not fully indicate individual potential or the complexities within individual engagement patterns (Xu et al. 2023). When implementing success and retention strategies, students should not be labelled or stigmatised based on their engagement profiles; the purpose and use of the data must be communicated in a sensitive and empowering manner. Open communication fosters authentic conversations with students and provides the tools and decision-making capabilities to understand student engagement patterns and develop strategies to improve learning.

The presentation will explore Keele University's comprehensive, data-informed intervention strategy, *'what worked and what didn't work'*. This strategy is nuanced to cater to all students and develops a sophisticated, context-laden intervention matrix that is dynamic, and robustly evaluated and refined to ensure effectiveness. Early, supportive, and honest outreach is crucial. The StREAM engagement analytics platform helps, and better tools and methods are positively impacting retention and success.

REFERENCES

Appleton, J. J., Christenson, S. L., and Furlong, M. J. (2008). Student engagement with school: critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*(5), 369–386. https://doi.org/10.1002/pits.20303

Azevedo, R. (2015). Defining and measuring engagement and learning in science: conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Gourlay, L., and Oliver, M. (2018). Student engagement in the digital university: Sociomaterial assemblages. In *Student Engagement in the Digital University: Sociomaterial Assemblages*. Taylor and Francis. https://doi.org/10.4324/9781315647524

Jewitt, K. (2020). Connecting Students with Customized Technology Solutions: Embedding Partnership in a Digital Learning Strategy. *Journal of Higher Education Policy and Leadership Studies*, *1*(3), 16–25. https://doi.org/10.29252/johepal.1.3.16

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758–773. https://doi.org/10.1080/03075079.2011.598505

Muir, T., Milthorpe, N., Stone, C., Dyment, J., Freeman, E., and Hopwood, B. (2019). Chronicling engagement: students' experience of online learning over time. *Distance Education*, *40*(2), 262–277. https://doi.org/10.1080/01587919.2019.1600367

Office for Students, 2024. Regulatory framework for higher education in England - Office for Students. [online] Available at: <https://www.officeforstudents.org.uk/> [Accessed 10 December 2024].

Reschly, A. L., and Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. *Handbook of Research on Student Engagement*, 3–19. https://doi.org/10.1007/978-1-4614-2018-7_1/COVER

Rimmington, S. (2024, April 17). *Creating a sense of empowerment through engagement data - HEPI*. https://www.hepi.ac.uk/2024/04/17/creating-a-sense-of-empowerment-through-engagement-data/

Wong, B., & Chiu, Y.-L. T. (2019). 'Swallow your pride and fear': the educational strategies of high-achieving non-traditional university students. In B. Wong & Y.-L. T. Chiu, British Journal of Sociology of Education (Vol. 40, Issue 7, p. 868). Taylor & Francis. https://doi.org/10.1080/01425692.2019.1604209

Wong, B., & Hoskins, K. (2022). Ready, set, work? Career preparations of final-year non-traditional university students. In B. Wong & K. Hoskins, Higher Education Pedagogies (Vol. 7, Issue 1, p. 88). Taylor & Francis. https://doi.org/10.1080/23752696.2022.2100446

Xu, X., Shi, Z., Bos, N. A., and Wu, H. (2023). Student engagement and learning outcomes: an empirical study applying a four-dimensional framework. *Medical Education Online*, *28*(1). https://doi.org/10.1080/10872981.2023.2268347

Yang, N., and Li, T. (2020). How stakeholders' data literacy contributes to student success in higher education: a goal-oriented analysis. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00220-3

Social annotation: metric and intention matching

Gavin Porter
Harvard Medical School
gavin_porter@hms.harvard.edu

ABSTRACT: Instructors desire quality dialogue and balanced interactions from students in discussion settings. Too many students in a forum may lead to information overload, while too few students may not be productive. A healthy blend of initiative and responsive output from each student in discussions is also a reasonable pedagogical goal. Built-in social annotation platform data organized on a per-document basis is lacking in student-centered longitudinal metrics, which have been suggested in many prior dashboard studies. In this practitioner report, a longitudinal analysis of student social annotation output across numerous documents allowed flagging of those clearly outside an initiative/responsive norm. This suggests an early to mid-course correction and valuable dashboard metric, as these same students would not have been identified by other common metrics. Cycling students through various group sizes also influenced the initiative/responsive balance, with desirable outcomes in two out of three sizes tested. Consideration of how metrics fit pedagogical goals will be essential in building future social annotation dashboards.

Keywords: Social Annotation, Collaborative Discussion, Personalized Feedback

1 INTRODUCTION

For learning analytics information to be actionable, information gathered from students needs to be aligned with an instructor's pedagogical goals (Lockyer et al., 2013; Wise & Jung, 2019). Instructors are faced with a wealth of information, but not always a clear path to future intervention. In a study by Dazo et al. (2017), an instructor shared:

> "In general, initially I was very hungry for analytics. Over time, I realized that…..I have the data, but I don't necessarily know what interventions to use to get the end result that I want."

Through the lens of *social* learning analytics, instructors are prompted to consider how student activity traces will be experienced by other students, and how student behaviors may compare to a norm established by the rest of the class (Shum & Ferguson, 2012). Sometimes, norms may need to be positively nudged by an instructor's intent. Recent work laments the absence of reports on the impact of analytics-informed adjustments on teaching practice, and provides suggestions for a dashboard applicable to social annotation (Hong et al., 2024). In this practitioner report, also dealing with social annotation output, the reader will see how student annotation behavior can be shaped through a class-wide intervention, and longitudinal analysis of individual students will indicate a targeted intervention based on the pedagogical intention of balanced responsive and initiative student output.

Social annotation platforms (i.e. hypothes.is, Perusall) take advantage of the natural inclination to mark up a text (questions, thought prompts, links to other material, etc.), and enable the sharing of annotations to drive further engagement. Students can initiate an annotation thread from any point in the source text and can also respond to the annotations of others to build discussion threads. Group sizes are pertinent to any student discussion forum, whether it occurs on a MOOC, traditional LMS-based forum, or with social annotation. Too many students in a discussion group may leave a student feeling overwhelmed and that everything interesting has already been said by the time they arrive to the forum. On the other hand, too few students in a group may fail to leverage the power of a crowd and diversity of thought from different members. Two heads may be better than one, but are 20 heads better than 10 heads, or 5 heads? There has been scant prior research on this in the social annotation field. The current default for group size on the Perusall platform is set to 20 students, perhaps informed by one prior study examining thread lengths and annotation quality, finding a quality plateau between 10 and 40 students (Miller et al., 2016). It is possible that ideal group size may vary with the level of the students, and the type of document being annotated, but research in this area is lacking.

The Perusall platform provides many metrics on a given annotated document. These include: grade distributions based on machine learning assessment of comments, submission time heat maps, page view reports, and student activity reports. The student activity reports give information such as: total comments, threads started, responses made, upvoting behavior, and average number of words per comment for each student. There is currently no built-in metric that looks at student performance *across* multiple annotated documents. Following an individual's output over multiple documents in a course could yield actionable information for intervention and improvement.

A student who only initiates discussions (makes the first annotation in a thread), may not be consuming subsequent responses, may not be motivated to respond to challenges to their initial post, and may not be consuming content generated by other students in other threads. On the other hand, a student who only responds, may be taking a short cut on reading the source text and orienting their output around what their classmates have said. These are worst-case scenarios, but demonstrate how a blend of responsive and initiative annotation is a reasonable pedagogical goal for an instructor overseeing social annotation.

In this practitioner report, students were cycled through various group sizes to look for an ideal, and longitudinal measurements yielded a helpful student-centered metric based on the pedagogical intention of balanced responsive and initiative output.

2 METHODS

The study took place with first year Master's students, in a university in the United States, in a course focused on the analysis of scientific primary research articles. Two student cohorts participated: a 2022-23 cohort of 21 students, and a 2023–2024 cohort of 19 students, for a total of 40 students. The research proposal was reviewed by the university's Human Research Protection Program, received the lowest risk categorization, and students provided informed consent. The students cycled through various annotation group sizes throughout the academic year, as each cohort analyzed 12 research articles using Perusall (see cycling sequences and representative access level rectangles in Figure 1). All students experienced all conditions (small group/low peer access: 4 groups of approximately 5 students; medium group/medium peer access: 2 groups of approximately 10 students; large group/large peer access: 1 group of approximately 20 students), and the sequences were inverted from one cohort to the next. The author prefers the term access level over group size. Obviously, output by a group of 20 students will be greater than that of one group of 5 students; that type of analysis is not given here. This report focusses on initiative (1st annotation made) versus responsive (2nd or later annotation in a thread) output balance. Responsive annotations were divided by the total number of annotations that each student made to derive a response percentage on a per paper basis. Initiative percentage is the inverse (i.e. if 70% of a student's annotations per paper were responsive, then 30% were initiative). As such, only response percentages are given here. A balanced response percentage (roughly 50%) stood as a reasonable pedagogical goal for individual student output. Data was exported from the Perusall platform, then analyzed in Microsoft Excel and GraphPad Prism.

3 RESULTS AND DISCUSSION

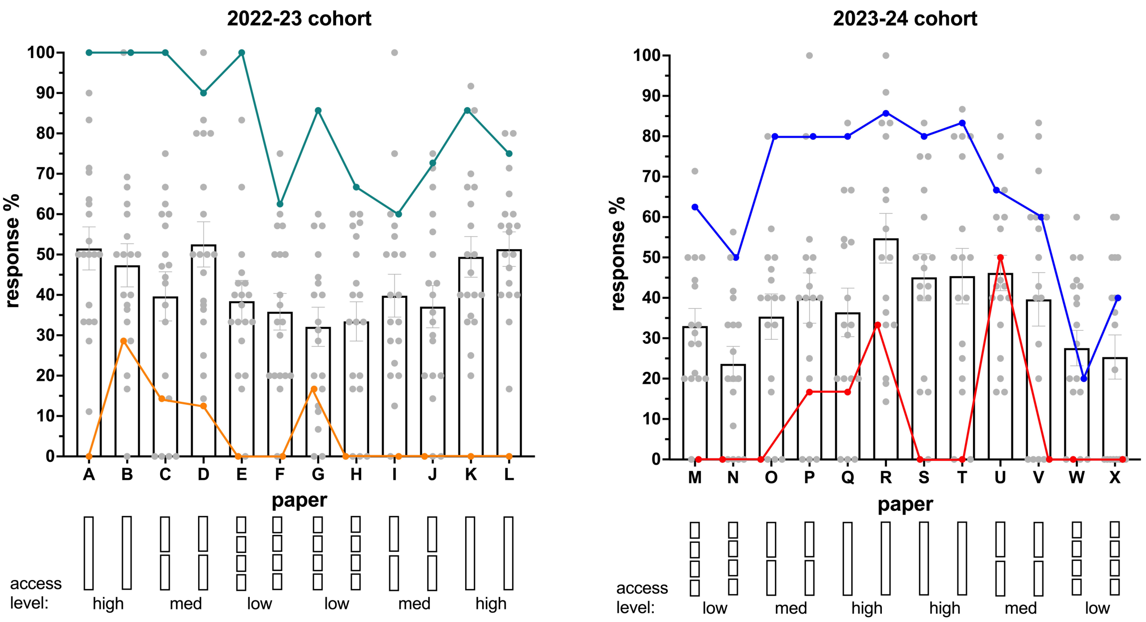


Figure 1: Longitudinal response percentages, flagging high and low cases

Actionable information came from the analysis of individual student traces over time. Students with the highest (teal 83%, blue 66%) and lowest (orange 6%, red 10%) mean response percentages are flagged (Figure 1; black bars: mean of all students in cohort on a given paper +/- SEM). These students are clearly outside the norm for an extended period, on multiple papers. Future interventions could target such students during the first 3-4 assignments (student in red trace had a 0% response rate in the first 3 assignments in the 2023-24 cohort), and prompt students whose output is predominantly responsive to be more initiative, and students whose output is predominantly initiative to be more responsive. The flagged students had all completed the assignments on time, and it is noteworthy that they would not have been flagged by looking at either the machine learning automated grading available on the platform, nor manual grading, as the grading rubric did not take responsive/initiative balance into account.

The effect of the independent variable of *access level*, is also evident. When students are exposed to more annotations from peers, they are more likely to make a response (Figure 2 summarizes both cohorts, n=40; repeated measures ANOVA; Tukey's post-hoc test, p values on figure, grey dots: each student's mean response % from 4 papers in indicated access level, black bars: mean of all students +/- SEM). Mean response percentage was 31% while in the low access condition, 41% in the medium access condition, and 48% while in the high access condition. The medium and high access conditions are closer to the desired pedagogical goal of a 50/50 blend of initiative and responsive output.

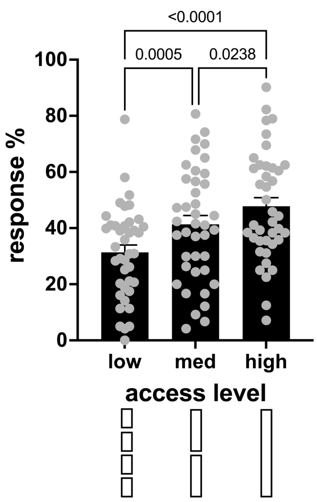


Figure 2: Access level influences response balance

In conclusion, access level is a clear driver for response behavior in students, and longitudinal analysis of response percentage follows from a pedagogical goal. The lowest access level seems to be outside of a *sweet spot*, as it is light on dialogic responses. Not all course settings may have the opportunity to use the largest group size, due to low enrolments. In that case, an instructor may want to take a larger role in promoting student dialogue. This practitioner report may encourage future social annotation analyses to follow longitudinal output, and prompt group size optimization for instructors using social annotation in their own courses.

REFERENCES

Dazo, S. L., Stepanek, N. R., Chauhan, A., & Dorn, B. (2017). Examining Instructor Use of Learning Analytics. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2504–2510. https://doi.org/10.1145/3027063.3053256

Hong, S., Zhu, X., & Chen, B. (2024). Tackling the alignment problem: The design of a learning analytics dashboard for teacher inquiry. *Companion Proceedings 14th International Conference on Learning Analytics & Knowledge (LAK24)*, 239–241.

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

Miller, K., Zyto, S., Karger, D., Yoo, J., & Mazur, E. (2016). Analysis of student engagement in an online annotation system in the context of a flipped introductory physics class. *Physical Review Physics Education Research*, *12*(2), 020143.

Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology & Society*, *15*(3), 3–26.

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.

Early detection of at-risk students in a DPT program utilizing predictive analytics and data visualization

Author(s):

1. Dr. James G. Rauh, Associate Professor and Doctor of Physical Therapy Director, School of Science, Marist University

2. Edward Presutti, Director Data Science and Analytics, Information Technology, Marist University

3. Hongli Xuan, Data Analytics Specialist, Information Technology, Marist University

4. Dr. Julie Fineman, Associate Professor and Director of PT Clinical Education, School of Science, Marist University

Email addresses:

{James.rauh, Edward.Presutti, Hongli.Xuan, Julie.Fineman}@marist.edu

ABSTRACT: This position paper describes the ongoing design and application of sequential data events to determine at risk students in a doctoral of physical therapy (DPT) program. Data elements were aggregated on each student from external and internal data sources that span from application through licensure examination. The goal of this initial project was to aggregate data into cohesive data storage and provide consolidated data visualization to facilitate data-informed decision-making. This will allow faculty to monitor student success, identify when and where students are struggling, develop timely remediation practices and improve future teaching/learning practices. Preliminary analysis links performance below a B in clinical decision-making didactic courses to first-time failure of licensing board exams, National Physical Therapy Exam (NPTE). Further, data analytics are focused on identifying individual assessments early in the curriculum that correlate to poor NPTE performance, in order to flag at risk students and allow for real time educational interventions during specific courses.

Keywords: Learning Analytics, Early Detection, Predictive Analytics, Tableau, Student intervention -remediation, Doctor of physical therapy students

1. PROJECT DESCRIPTION

The cumulative event of any professional education program is the successful completion of a licensing board exam. The average United States (US) first time pass rate for the National Physical Therapy Examination (NPTE) in 2023 was 84.9%. Ultimate (two-year pass rate) in 2023 was 97.5% (The Federation of State Boards of Physical Therapy -Free Basic Pass Rate Report, n.d.). Professional Physical Therapy (PT) programs are evaluated based upon these metrics and other accreditation benchmarks. Numerous studies have determined specific variables that are predictive of students passing the NPTE (Abdolahi et al., 2016; Baldwin et al., 2023; Coleman-Salgado, 2019; Dombkowski et al., 2023; Heath et al., 2020; Kume et al., 2019; Wolden et al., 2020). Data elements of physical therapy program applicants as well as grade point averages of students at various points throughout the course of study have shown correlation with pass scores for first time and ultimate pass rates. Preadmission data of the undergraduate cumulative GPA (UGPA) was more closely correlated with the DPT overall GPA whereas the undergraduate pre-requisite science GPA (SGPA) correlated with

NPTE scores (Fell et al., 2015). The UGPA was also found to be predictive of students NPTE success as well as predictive of students having academic difficulties in the program (Dombkowski et al., 2023; Heath et al., 2020; Utzman et al., 2007). Various studies have also linked the SGPA to passing the NPTE (Fell et al., 2015; Roman & Buman, 2019). Other pre-admission variables such as the GRE have also been linked to student academic difficulties and the need for remediation in programs (Kume et al., 2019). In addition to pre-admission criteria, DPT curricular criteria have also been shown to impact NPTE performance. Time spent in lab working on psychomotor skills was found to be more predictive of NPTE performance than didactic coursework (Maring et al., 2013; Utley et al., 2016).

While the use of pre-admission data elements as well as end of semester or year GPA's may provide insight into potential student poor performance on the NPTE, it does not delineate when the student is having difficulty during the actual programs course of study. Cui and Chen expressed that course-level prediction of students course performance is key so that early intervention for improvement and successful student outcome can occur (Cui et al., 2019). Applying a similar methodology will allow early identification of DPT students who may be at risk for poor outcomes during the course of the program and on the NPTE.

Data elements on DPT students are generated from the moment they apply to the program all the way through licensure examination. This data comes from internal and external sources. There are 7 sources of data. The Physical Therapist Centralized Application Service (PTCAS) is an external database containing the student's data as submitted with their program application. The Clinical Internship Evaluation Tool (CIET) is another external data source that houses clinical rotation performance data for individual students. The Federation of State Boards of Physical Therapy (FSBPT) is the national organization and data repository of the board examination data, including practice board and the NPTE. The FSBPT data is available at multiple levels including school, cohort, student or specific question level data. Exxat Prism provides the source of all data related to the clinical education experiences. ExamSoft encompasses written examination data from DPT program across the curriculum and by all faculty. Banner is the internally deployed Enterprise Resource Planning (ERP) software based on the implementation of the student information system at the college. Banner internally houses the official course grades for every student. Brightspace is a cloud-based Learning Management System (LMS) which accumulates student progress throughout the semester by storing grades for exams, quizzes and assignments and presenting course material.

The goal of this initial project was to aggregate data into cohesive data storage and provide consolidated data visualization to facilitate data-informed decision-making. This will allow faculty to monitor student success, identify when and where students are struggling, develop timely remediation practices and improve future teaching/learning practices. Additionally, this aggregated data can be leveraged to improve admissions practices and enhance the curriculum design. Development of such a robust system of data storage with associated visualization will also expedite the DPT program's formal accreditation process through the automation of data collection, aggregation, and reporting processes.

2. CONSOLIDATING DATA SOURCES FOR EXTRACTION TRANSFORMATION AND LOADING (ETL)

To support the consolidated dashboarding effort it was necessary to provide automated mechanisms that could extract and normalize the data across disparate systems. The normalization process involved the identification of each of the elements that were available and synchronize their data types, lengths and attributes. It was known that each system did not store the data in consistent formats, for example: dates were sometimes stored or reported as text strings with varying formats. The first step in the process was to implement extraction automation wherever possible. To that end automated programs were developed in Python using Linux shell scripts running on an extraction schedule that could interface with the various systems, in some cases direct database connections were created in other cases direct application program interfaces (API) were used. There were also circumstances where no direct access was available nor were there API interfaces available, in these cases, reports were generated from the source systems and placed in a file storage location where automated processes would ingest the data. In each case regardless of the interface the data was taken in its raw form and stored in a “transition data store”, whose sole purpose was to act as a transient location for the data so that it could be further processed. The transition data store is a MariaDb database.

Once the data was extracted to the transition data store internal database procedures were developed using structured query language (SQL). These procedures also ran on a schedule and were triggered after the extractions. The purpose of the procedures is to further transform the data and subsequently store the data in MariaDb relational databases that coordinated the keys from each of the source systems into a single data warehouse identifier. In addition to the data that was specific to the DPT systems, the warehouse also contained institutional data that existed from previous data warehouse activities or data that was stored as part of the core functions of the data warehouse. All the data was then loaded into related databases and database tables.

The final step in the process was to create views into the data that could easily be read by the dashboard tool (Tableau). To facilitate this process a DPT reporting database was created that provided SQL views which could be subsequently read by the bridge software whose purpose was to move the data to the cloud data source for the creation of the dashboards. The data views provided a level of abstraction that isolated the complexity of the underlying data, which provided an ease of implementation when building the dashboard. In addition, using the views provided additional security and reduced the exposure of the database to unwanted access. Data connections directly to the original data sources were forbidden and data was pushed to the cloud dashboard implementation which eliminated the need for inbound connections. Ultimately the Tableau dashboards had relatively simple data sources that created a very responsive interface.

3. OUTCOMES

Preliminary analysis of the aggregated data indicates that grades of less than a “B” in the didactic courses in the examination and evaluation of patients are more indicative of NPTE 1st time failure. These courses have a very strong clinical decision-making component. Further assessment of individual assignments, quizzes and each test are the current focus of ongoing data analysis. These courses occur beyond the second term of the program. The goal is to identify as early as possible, ideally during the first term course work, students who would benefit from interventions and remediation. Analysis of courses assignments, quizzes, lab and written exams in the first term and their association with potential academic difficulties is a primary focus at this time.

REFERENCES

Abdolahi, A., Bull, M. T., Darwin, K. C., Venkataraman, V., Grana, M. J., Dorsey, E. R., & Biglan, K. M. (2016). A feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. *Health Informatics Journal*, *22*(2), 304–311. https://doi.org/10.1177/1460458214556373

Baldwin, J., Schmidt, C., Plummer, L., Gochyyev, P., Battista, J. E., Kaur, S., & Naidoo, K. (2023). How doctor of physical therapy students overcome academic challenges to achieve first-attempt success on the national physical therapy examination: A mixed methods study. *Education Sciences*, *13*(5), Article 5. https://doi.org/10.3390/educsci13050430

Coleman-Salgado, B. (2019). The relationship of preadmission academic variables to academic performance in a doctor of physical therapy program. *Journal of Allied Health*, *48*(1), E9–E14.

Cui, Y., Chen, F., Ali, S., & Fan, Y. (2019). Predictive analytic models of student success in higher education. *Information and Learning Science*, *120*(3/4), 208–227. https://doi.org/10.1108/ILS-10-2018-0104

Dombkowski, R., Sullivan, S., Widenhoefer, T., Buckland, A., & Almonroeder, T. G. (2023). Predicting First-Time National Physical Therapy Examination Performance for Graduates of an Entry-Level Physical Therapist Education Program. *Journal of Physical Therapy Education*, *37*(4), 325. https://doi.org/10.1097/JTE.0000000000000291

Fell, N., Mabey, R., Mohr, T., & Ingram, D. (2015). The preprofessional degree: Is it a predictor of success in physical therapy education programs? *Journal of Physical Therapy Education*, *29*(3), 13.

Heath, A. E., Mahoney, E., Middleton, A., Brown, T., & Fritz, S. (2020). Demographic and admission predictors of students with perceived difficulty in entry-level doctor of physical therapy programs. *Journal of Allied Health*, *49*(4), 279–284.

Kume, J., Reddin, V., & Horbacewicz, J. (2019). Predictors of physical therapy academic and NPTE licensure performance. *Health Professions Education*, *5*, 185–193. https://doi.org/10.1016/j.hpe.2018.06.004

Maring, J., Costello, E., Ulfers, M., & Zuber, E. (2013). Curriculum, faculty, and cohort variables predicting physical therapist assistant program graduate success on the national physical therapy examination. *Journal of Physical Therapy Education*, *27*(2), 33.

Roman, G., & Buman, M. P. (2019). Preadmission predictors of graduation success from a physical therapy education program in the southwestern United States. *Journal of Educational Evaluation for Health Professions*, *16*, 5. https://doi.org/10.3352/jeehp.2019.16.5

The Federation of State Boards of Physical Therapy -Free Basic Pass Rate Report. (n.d.). The Federation of State Boards of Physical Therapy. Retrieved October 5, 2024, from https://school.fsbpt.net/Report/ViewReport?format=web&id=24761

Utley, C., Brown, S. R., & Robel, J. S. (2016). Effect of clinical experience on comprehensive examination performance in a physical therapist education program. *Journal of Physical Therapy Education*, *30*(2), 38.

Utzman, R. R., Riddle, D. L., & Jewell, D. V. (2007). Use of demographic and quantitative admissions data to predict academic difficulty among professional physical therapist students. *Physical Therapy*, *87*(9), 1164–1180. https://doi.org/10.2522/ptj.20060221

Wolden, M., Hill, B., & Voorhees, S. (2020). Predicting success for student physical therapists on the National Physical Therapy Examination: Systematic review and meta-analysis. *Physical Therapy*, *100*(1), 73–89. https://doi.org/10.1093/ptj/pzz145

Deep-Reflect: an LLM-based Reflection Analytics Tool

Nasrin Dehbozorgi
College of Computing and Software Engineering, Kennesaw State University, USA
dnasrin@kennesaw.edu

Mourya Teja Kunuku
College of Computing and Software Engineering, Kennesaw State University, USA
mkunuku@students.kennesaw.edu

ABSTRACT: This practitioner report introduces an AI-based framework for analyzing students' reflections. Integrating Large Language Models (LLMs) into educational tools has revolutionized learning analytics by allowing complex analysis of textual data. Reflective writing is known to promote cognitive and metacognitive skills among students. However, providing timely feedback on these reflections is a time-intensive task for educators, often limiting its practice. This paper introduces Deep-Reflect, an LLM-powered tool designed to automate the analysis of student reflections by extracting learning outcomes and challenges and visualizing them through a dynamic dashboard. This tool enables instructors to provide timely feedback and make data-driven interventions. A case study conducted in a graduate software engineering course showed that using Deep-Reflect significantly improved student performance. This finding highlights the potential for LLM-powered tools to enhance reflective learning and student outcomes in higher education settings.

Keywords: Natural Language Processing, Large Language Models, Learning Analytics, Reflection

1 INTRODUCTION

The rapid advancement of AI has significantly transformed computer science education, leading to increased reliance on AI-based content among students. While this shift offers several benefits, it raises concerns about student engagement, learning outcomes, and retention rates in higher education. Engagement is critical for student success and can be enhanced through formative assessments, critical thinking, and reflective practices [1]. Reflection plays a crucial role in developing metacognitive and critical thinking skills, yet conventional assessment methods—such as quizzes, exams, and surveys—fail to capture real-time learning progress effectively. Manual review of reflections is time-consuming and can also introduce bias, while quantitative methods may overlook the complexity and depth of reflective thought. Although ML-based automated text analysis methods hold promise, these systems require substantial training data and often struggle with context-specific nuances. This limitation underscores the need for AI-powered tools capable of in-depth analysis of reflections to provide educators with real-time insights into students' learning outcomes and challenges. Building on our prior work [2] as the foundation, this study introduces the Deep-Reflect tool, which leverages LLM capabilities for a more comprehensive analysis of student reflections. Additionally, we assess students' performance by comparing grades before and after the tool's implementation. The following sections will review related literature, outline our methodology, present analysis results, and conclude with discussions on our findings and future directions.

2 REFLECTION: CHALLENGES AND OPPORTUNITIES

Reflective writing enhances student learning by promoting self-awareness, critical thinking, and metacognitive skills [2]. Through evaluating their learning experiences, students identify strengths, areas for improvement, and strategies for personal growth. Research indicates that reflective practices increase student engagement and enhance learning outcomes by helping students recognize their knowledge gaps [2]. This active engagement leads to more profound knowledge comprehension, improves academic performance, and fosters students' motivation, perseverance, and self-efficacy [3]. While the benefits of reflective writing are well-documented, traditional assessment methods such as surveys and manual analysis of reflections present significant challenges. These methods often fail to capture the depth and nuances of reflective thought, and manual analysis can be time-consuming and subjective, particularly in large classes. As a result, many educators hesitate to incorporate reflective practices into their curricula despite their potential advantages. However, the emergence of LLM techniques offers promising opportunities to automate student reflection analysis and enables educators to gather insights into students' learning efficiently. Many existing research studies have applied advanced machine learning and natural language processing methods such as topic modeling, text classification, and transformer-based models for automating reflection analysis [2,4]. By harnessing these innovative methods, educators can enhance reflective learning practices and better support students' success. However, gaps still exist in understanding the scalability and generalizability of these models across diverse educational contexts.

3 METHODOLOGY AND OUTCOMES

In this study, we developed a new tool named Deep-Reflect that leverages the strengths of the LLMs by employing the LangChain framework along with the GPT3.5 to analyze student reflections. LangChain is a framework that offers a modular approach to integrating and adopting LLM capabilities into different applications.

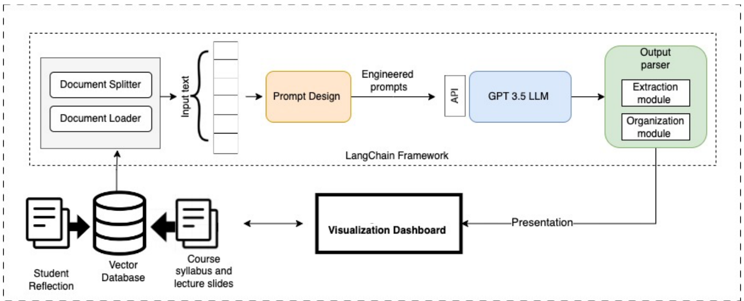


Figure 1: High-level Architecture of Deep-Reflect

The LangChain framework provides three key features: models, prompts, and parsers. Models refer to the foundational LLMs, such as GPT-3.5, used in the framework. Prompts create inputs for the LLMs, and the parsers structure the outputs into a more organized format for downstream tasks. In this work, we adopted the GPT-3.5 model via OpenAI API to enhance the precision and accuracy of our analysis. LangChain prompts are very insightful for in-depth analysis of data. It allows users to interact with existing models more effectively by customizing these prompts, which improves the precision and quality of the generated content. LangChain's output parser offers customization options that extract and organize the LLM output into specific formats based on defined criteria. This flexibility ensures that the LLM-generated content is coherent and relevant to the main system goal. In this study, we crafted LangChain prompts to elicit precise and contextually relevant responses from GPT-3.5 in the analysis of the reflection data. These prompts guided the model's generation process to ensure the extracted topics align with the course subject (i.e., Software Engineering). The framework's output parser played a crucial role in extracting, organizing, and presenting the data generated by GPT-3.5 and improving the readability and usability of the endpoint dashboard. The high-level architecture of our model is illustrated in Figure 1. The process begins with data collection and preprocessing student reflections gathered from the 'Minute Paper' technique where students summarize their learning experiences and challenges encountered in each lecture. This raw input undergoes cleaning, tokenization, and section-specific splitting before being stored in a centralized database. The system's database includes additional data, such as course syllabi and lecture notes, to facilitate the subsequent prompt-generation tasks. The LangChain framework retrieves documents from the database using a Document Loader, and the Document Splitter breaks them into manageable chunks for detailed analysis. The input text undergoes prompt engineering before being processed by the LLM for topic extraction. The pipeline starts with an API call for the LLM execution, where GPT-3.5 analyzes the prompts and generates results. The output parser structures these responses for more precise interpretation.

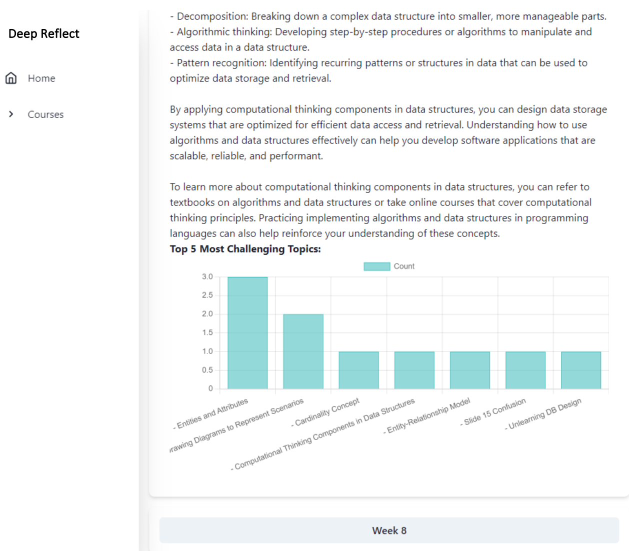


Figure 2: Sample output of the Deep-Reflect dashboard

A distinctive feature of our model is its use of the LLM's language generation feature to provide instructions and additional guidelines about students' challenging topics. The final stage includes the analysis and visualization module, which plots key challenges and learning outcomes for each class session. This module calculates weights based on the frequency and relevance of identified topics and provides instructors with a clear visual representation of areas where students face difficulties.

3. 1 Case Study of Deep-Reflect in a Graduate Software Engineering Course

The primary research question guiding this study is: "How does the application of Deep-Reflect impact students' learning and performance?" To investigate this, we conducted a case study in a Software Engineering (SWE) course and evaluated the impact of the Deep-Reflect tool on students' grades. The study involved two groups: Group A, consisting of 80 students who completed the SWE course in Fall

2022 before the intervention, and Group B, with 80 students who took the same course in Fall 2023 after the intervention. We hypothesized that implementing Deep-Reflect would enhance student performance, expecting Group B to achieve higher grades than Group A. After collecting and analyzing grades from both semesters, we conducted a two-tailed t-test to compare the mean grades of the two groups, setting a confidence level of 0.05. The calculated p-value was 0.034, indicating a statistically significant difference in grades. This result led to rejecting the null hypothesis, suggesting that Deep-Reflect positively influenced students' performance. However, it is important to note that this analysis is confined to a single Software Engineering course. Further studies are required to draw broader conclusions regarding the tool's effectiveness across various disciplines. Figure 2 presents a sample output from the analysis dashboard by showing the frequency of the top challenging topics students encountered in a specific class session. Users can access additional guidelines generated by the LLM by clicking on each challenging topic.

4 CONCLUSION AND FUTURE DIRECTIONS

In this study, we developed Deep-Reflect, a tool that utilizes the capabilities of GPT-3.5 through the LangChain framework by employing a carefully crafted prompt engineering process. This approach enhances topic identification accuracy and yields more customized outputs for student reflections. We Implemented this tool in a college-level Software Engineering classroom to provide insights into students' learning trajectories. Our analysis indicated a statistically significant improvement in grades post-intervention, highlighting the tool's effectiveness in facilitating formative assessments of students' reflections via the Minute Paper technique. In the future, we plan to explore additional LLMs within the LangChain framework to further enhance topic extraction and analysis capabilities in Deep Reflect. Additionally, we aim to track individual students' learning trajectories over multiple classroom sessions, offering personalized insights into each student's progress and learning journey over time.

REFERENCES

[1] Menekse, M., Anwar, S., & Akdemir, Z. G. (2022). How do different reflection prompts affect engineering students' academic performance and engagement? *The Journal of Experimental Education, 90*(2), 261–279. Taylor & Francis.

[2] Dehbozorgi, N., & Kunuku, M. T. (2024, March). An LLM-based Reflection Analysis Tool for Identifying and Addressing Challenging Topics. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2 (pp. 1618-1619).

[3] Butt, A. A., Anwar, S., Magooda, A., & Menekse, M. (2022). Comparative analysis of the rule-based and machine learning approach for assessing student reflections. In *Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022, pp. 1577-1580*. International Society of the Learning Sciences.

[4] Dehbozorgi, N., & Dindu, K. G. (2023, October). Analysis of Learning Outcomes in Software Engineering: an Automated Reflection Analysis Tool. In *2023 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.

From Research to Practice: Translating a Dissertation-based solution into an Enterprise-level System

Wei Qiu
Centre for Applications of Teaching & Learning Analytics for Students (ATLAS)
Nanyang Technological University (NTU), Singapore
qiuwei@ntu.edu.sg

Andy W.H. Khong
School of Electrical & Electronic Engineering and Lee Kong Chian School of Medicine
Nanyang Technological University (NTU), Singapore
andykhong@ntu.edu.sg

Fun Siong Lim
Centre for Applications of Teaching & Learning Analytics for Students (ATLAS)
Nanyang Technological University (NTU), Singapore
LIM_Fun_Siong@ntu.edu.sg

ABSTRACT: Reducing student dropout rates and enhancing academic success are critical challenges in higher education. While predictive machine learning models have shown promise in identifying at-risk students, their practical deployment often remains elusive. This paper presents a scalable data pipeline to operationalize a suite of grade-prediction models developed during a Ph.D. program. By integrating Denodo, Dataiku, Snowflake, and Qliksense, we established a robust and secure data flow, encompassing data collection, transformation, modeling, validation, and visualization. The pipeline automatically updates predictions every six months, enabling timely intervention by student care managers. This successful case study demonstrates the potential of AI and data science to improve student retention and foster academic success.

Keywords: Learning analytics, educational data, data pipeline, grade prediction, early intervention

1 IMPORTANCE OF GRADE PREDICTION IN LEARNING ANALYTICS

The ability to predict grades and identify at-risk students is vital not only for improving individual learning outcomes but also for fostering a more equitable educational environment. With insights from predictive models, instructors can tailor their teaching strategies to meet the specific needs of struggling students. This targeted approach helps institutions optimize resource allocation, directing tutoring, counseling, and other support services to where they can have the most significant impact. Incorporating predictive analytics into education supports data-driven decision-making and enhances student retention efforts. By providing timely interventions such as peer tutoring and personalized learning plans, institutions can provide the necessary support and help students succeed, resulting in reduced dropout rate. Ultimately, predictive models enable institutions to better serve the needs of students, creating a more inclusive, supportive, and successful learning environment.

2 RESEARCH WORK ON GRADE PREDICTION AND AT-RISK STUDENT DETECTION

The research on student grade prediction and the identification of at-risk students was initiated as part of a Ph.D. funding initiative to develop advanced machine learning techniques, with a particular focus on grade prediction models (Qiu, 2023). Student academic challenges are identified by predicting grades based on history. An academic history provides insights into a student's learning trajectory, including their performance relative to peers, the influence of past courses on future ones, and the impact of external factors on learning outcomes. Machine learning algorithms can accurately predict grades. Figure 1 illustrates how the prediction model detects at-risk students.

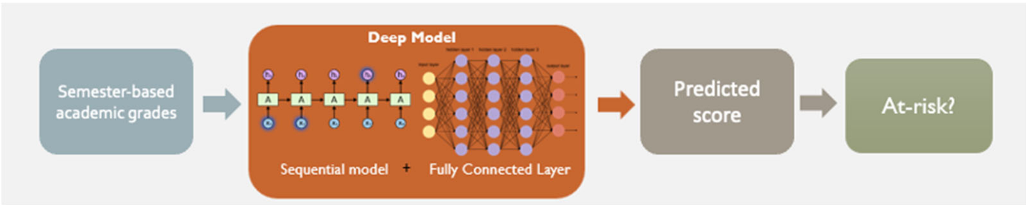


Figure 1: The workflow of how prediction model detects at-risk students.

The Ph.D. student adopted the above framework by analyzing anonymized data and identifying the key components highlighted above. Three new models were subsequently proposed with the first being a two-stage model that incorporates both long-term temporal data and short-term performance fluctuations (Qiu , Khong, Supraja, & Tang, 2023). Modelling short-term fluctuations allows the model to account for performance fluctuations due to external factors such as financial challenges and/or workload demands. The second model introduces a framework (Qiu, Supraja, & Khong, 2022) that captures temporal dynamics of academic performance, short-term performance consistency, and peer-relative performance. This comprehensive approach allows for more precise grade prediction by considering the difficulty of courses in addition to detrimental factors that may affect students within a given semester. The third model employs efficient data encoding techniques in conjunction with contemporary Transformer architectures to address inadequate handling of relative performance and complex data constraints in existing models (Qiu, 2023). The proposed model comprises three core modules: a relative performance module, a logic reasoning module, and a Transformer-powered grade prediction module. Collectively, these modules enhance the representation of student data.

The translation of research outcomes arising from a Ph.D. program into practical, real-world applications requires the adaptation of these sophisticated algorithms to handle diverse and incomplete data. Challenges associated with scalability across the entire institution, user-centric interface design, integration with existing IT systems, and cybersecurity requirements were overcome by leveraging cloud-based data infrastructure, integrating AutoML platform, and implementing robust security measures. In addition, to ensure successful roll out, data governance policies were revised while student care procedures were streamlined to adopt a more preemptive stance.

3 TRANSLATION OF RESEARCH TO PRACTICE

Regardless, the ultimate goal of learning analytics is translational so that students and staff benefit from the insights generated by the models and systems created by researchers even if they have moved on. A successful example of this translational approach is the translation of the mentioned Ph.D. research to practice within our university since 2021 after the first model was deployed. Since May 2023, the team has worked to integrate the Ph.D. candidate's models, along with other baseline machine learning models, into our enterprise solution to support schools in identifying and assisting students who may be at risk of academic challenges before mid-way into the semester. The early alert system now includes a fully-automated data and analytics pipeline, pulling data from our data warehouse to a data science platform, and finally to a dashboard accessible by designated student care managers. This process is scheduled to run at the start of each semester.

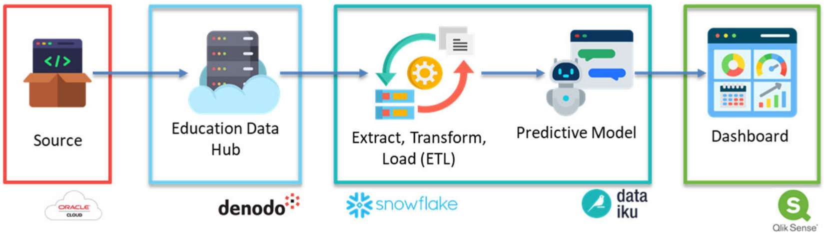


Figure 2: The automated model pipeline for grade prediction.

As show in Figure 2, the pipeline comprises a comprehensive infrastructure designed to support scalable learning analytics and predictive modelling within an institutional context. It begins with data extraction from various sources, such as Oracle Cloud, containing student academic and performance data. This raw data is then integrated through Denodo, a data virtualization platform that provides real-time access to heterogeneous data sources without the need to make duplicate copies to ensure a single source of truth, enhancing flexibility, and reducing complexity. The data is subsequently processed in Snowflake, a cloud-based data warehousing platform that allows for large-scale data processing with the ability to handle the extensive transformations required. Processed data is passed to Dataiku, which hosts machine learning models that are trained on historical data and automatically updated with each new semester's information, making the system adaptive and operationally sustainable. The outputs of these models are then visualized in Qlik Sense, a business intelligence tool that provides stakeholders, such as student care managers, with interactive dashboards that provide a holistic view of each case for efficient diagnosis and decision making. As shown in Figure 3, the dashboard offers real-time insights, allowing for timely interventions that can improve student outcomes. This pipeline exemplifies how learning analytics research can be translated into enterprise-level solutions, integrating advanced technologies to ensure data security, governance, and scalability, while delivering actionable insights that directly benefit both students and staff.

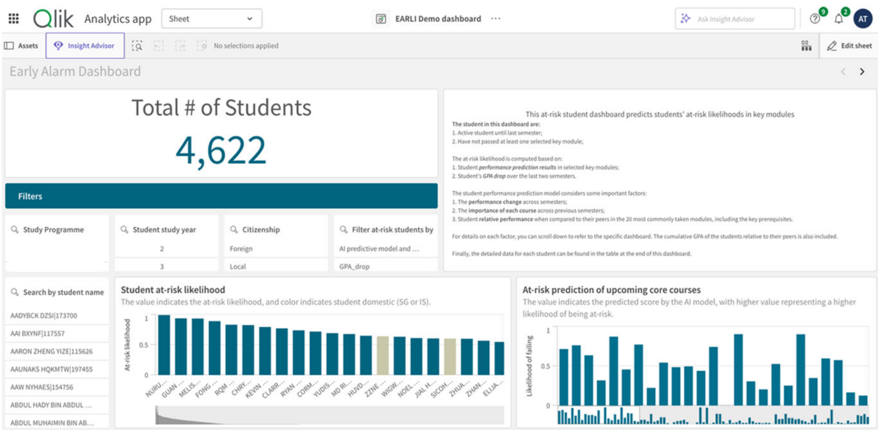


Figure 3: The screenshot of the early alert dashboard - names have been redacted for privacy.

By the current semester, the early alert system has been implemented across five schools within our university. However, the implementation process was not without its challenges. The successful implementation of the system requires the commitment of departments, including the principal investigator and Ph.D. supervisor, the head of the university's [redacted for review] centre, the student data governance lead, and the data services team lead from IT Services Centre. The synergy between these administrative departments facilitates the integration of learning analytics research into real-world practice.

4 IMPACT

Six trials conducted since January 2021, involving 13,944 students. Each semester, the at-risk student is made available via a digital dashboard to student care managers, who are then responsible for contacting the student. At the time of writing, the true-positive rate in the schools surveyed is over 70%. The following is the testimony of an Associate Chair (Students) from one school:

"…… Prior to its implementation, pinpointing students facing potential challenges was difficult…… With this tool, greater attention can be directed towards the identified students to provide them with the necessary support and assistance to succeed."

In the 2023/24 academic year, another school reported that 13 out of 17 students were identified as requiring assistance by the school care manager. The student care manager shared that:

"……in the past, student care managers might learn about academically struggling students a little too late and they didn't have enough time to work with students to try and help them turn things around. That is until [redacted for review] was developed and they could get alerted earlier."

For the 13 students identified as being at risk, the student care manager observed an improvement in their grade point average (GPA) after intervention. One went from failing four modules to achieving an average grade of B. Another said the support gave him optimism and motivation to persevere and complete his degree. The student's feedback is provided below:

"I feel that it is very helpful as it gives students like me confidence and hope to make improvements in my studies knowing the school is supporting and watching over us. Before the call I thought the school simply do not care about students that are performing badly in their studies. The call definitely gave me more drive to achieve my goal of pulling my GPA up to at least 2.5 at graduation."

While many institutions have similarly adopted early alert systems for the benefits of students and staff, our paper shares how data science and machine learning enterprise infrastructure can be leveraged to facilitate a seamless integration of learning analytics research produced at the university. We demonstrate an approach where successful learning analytics research is being translated in a sustainable and scalable way that avoids becoming yet another dissertation shelved within the university repository.

REFERENCES

Qiu , W., Khong, A., Supraja, S., & Tang, W. (2023, 11 15). A Dual-Mode Grade Prediction Architecture for Identifying At-Risk Students. *IEEE Transactions on Learning Technologies, 17*, 803 - 814.

Qiu, W. (2023). *Academic achievement inspired machine learning methods for student grade prediction and at-risk detection.* Singapore: Nanyang Technological University.

Qiu, W., Supraja, S., & Khong, A. (2022). Toward better grade prediction via A2GP - an academic achievement inspired predictive model. *15th International Conference on Educational Data Mining (EDM 2022)* (p. 11). United Kingdom: International Educational Data Mining Society.

Evaluation and Learning Enhancement Via Automated Topic Extraction (ELEVATE)

Aditya Phatak*
University Of California, Irvine
phataka@uci.edu

Suvir Bajaj*
University Of California, Irvine
bajajss@uci.edu

Rutvik Gandhasri
University Of California, Irvine
rgandhas@uci.edu

Pavan Kadandale
University Of California, Irvine
pavan.k@uci.edu

Jennifer Wong-Ma
University Of California, Irvine
jwongma@uci.edu

Hadar Ziv
University Of California, Irvine
hziv@uci.edu

*These authors contributed equally to this work and the ELEVATE tool

ABSTRACT: Every term, instructors receive course evaluations that, in theory, should provide them with insights into student experiences in their course. However, manually identifying recurring themes and extracting actionable insights from potentially thousands of reactions is extremely time-consuming, if not impossible. We present Evaluation and Learning Enhancement Via Automated Topic Extraction (ELEVATE), a topic modeling tool designed to cluster student responses and extract latent themes, topics, and trends from large student evaluation datasets (within a term or across terms). ELEVATE offers an intuitive dashboard for users that effectively integrates qualitative (identification of topics) and quantitative (sentiment type and strength) analyses. Furthermore, this paper presents one case study to showcase its capabilities in learning analytics: an investigation on variations between offerings of a Computer Science course taught by multiple professors.

Keywords: Course Evaluations, Sentiment Analysis, LLM Tools, Topic Modeling

1 INTRODUCTION

Student feedback is crucial to improving pedagogy and fostering inclusive learning environments. Student Evaluations of Teaching (SETs) continue to be the de facto standard for collecting numerical "scores" and written comments about the student experience in a course (Dziuban et al., 2023). In courses with large enrollments, the sheer volume of responses makes accurate interpretation of overall student response challenging. Instructors risk forming a skewed view of their course based on the most vocal feedback. Consequently, faculty often resort to using simple numerical summary scores for evaluating teaching quality - a practice that is fraught with issues, including biases against age, ethnicity, gender, etc. (Heffernan, 2022). Thus, despite the time and effort spent implementing these SETs, extracting meaningful and representative insights from student feedback remains a significant challenge.

2 METHODOLOGY

ELEVATE leverages BERTopic (Grootendorst, 2022), a newer topic modeling approach relying on word embeddings and clustering. Unlike older statistics-driven methods such as Latent Dirichlet Allocation (LDA), which rely on word frequency and co-occurrence probabilities, BERTopic captures the semantic relationship between words in its embeddings. It then clusters similar comments based on their density and distribution, utilizing HDBSCAN and K-Means respectively. This quantitative approach results in consistent and reliable topic modeling. Subsequently, ELEVATE uses Llama 3.1 – an open-source Large Language Model – to generate meaningful topic labels from representative comments and keywords. Furthermore, ELEVATE utilizes RoBERTa to quantify sentiment with numerical scores that reflect the intensity of the emotions expressed in each comment. All of this allows for a more nuanced, consistent understanding of student comments.

Despite the complexity of the approach, what is required of a user is very simple, and consists of merely having to upload a CSV file with student comments (and optionally, any relevant metadata). The ELEVATE pipeline then processes the file and outputs a CSV with the identified topics that a user can download and use for their downstream analysis. This simple workflow is summarized in Figure 1.

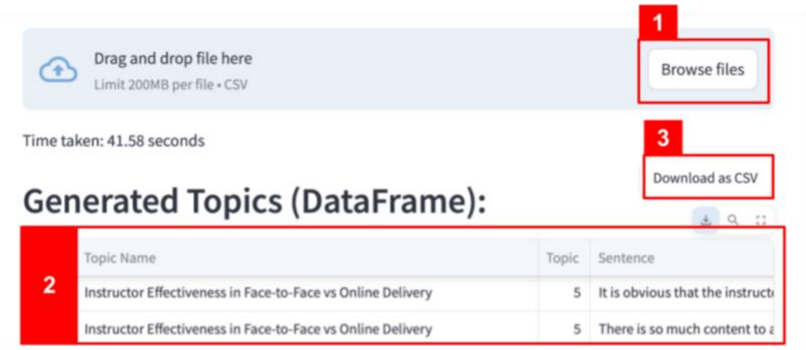


Figure 1: ELEVATE Dashboard Homepage and User flow; 1. Upload a course evaluation CSV file to run ELEVATE on **2.** View output CSV with topic assignment **3.** Optionally download CSV file for further analysis

3 COMPARISON WITH LLMS

Table 1: Comparison of themes generated by GPT-4o versus ELEVATE

GPT-4o	ELEVATE
Coding and Languages	MIPS Programming Experience and Problem Solving
	Low-Level Programming Experience and Coding Labs
Q&A and understanding	Structured Lecture Materials and Support
	In person Q&A sessions and Clarification opportunities
	Instructor Effectiveness in Face-to-Face vs Online Delivery

Given the ready accessibility of LLMs (Large Language Models), we compared the outputs of GPT-4o and ELEVATE for a file of 213 student evaluations from a Computer Science course. The prompt used for GPT-4o was "*Determine the common themes that occur within the student evaluation file and use the themes to categorize the student comments.*" Table 1 shows examples of how GPT-4o overgeneralized themes, producing broad topics, while ELEVATE's output broke down these broader themes into detailed, contextually meaningful categories. Thus, ELEVATE provides more defined topics, allows instructors to better pinpoint specific areas of the student experience and makes analyzing feedback more actionable. Further, although GPT-4o is capable of sentiment analysis, its outputs are qualitative labels (positive, negative, or neutral) whereas ELEVATE outputs a score from -1 to 1 (most negative to most positive), offering a more precise measure of emotional analysis and allowing for tracking of theme-specific sentiment trends when combined with other metadata (such as time). GPT-4o does not perform topic modeling in a transparent manner. Due to its inaccessible embeddings (which is the case for all LLMs) which are used internally during individual chat sessions to find topic clusters, the analyses are not reproducible, and the output quality relies heavily on prompt engineering. ELEVATE uses the same algorithm for embeddings every time, does not require the user to engage in iterative rounds of prompt engineering and can be represented in a two-dimensional space (Figure 2), providing a clear picture of how documents are clustered and their coherence within each cluster.

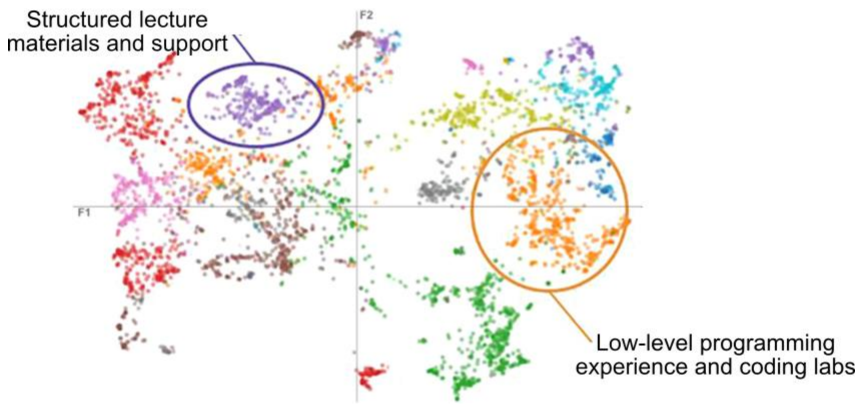


Figure 2: Example showing representation of ELEVATE-generated clustering of topics in two-dimensional space. This allows for visually inspecting the generated clusters for coherence within each cluster.

Thus, by simply navigating to the publicly hosted site, uploading a CSV file, and pressing submit to run the pipeline, ELEVATE provides the user with accurate, robust, reproducible, fine-grained results, obviating the need for any coding and/or prompt engineering. ELEVATE is also the superior option in terms of privacy for a variety of reasons: a) ELEVATE does not store any user data; b) ELEVATE does not use any user data for training; c) ELEVATE can be run locally and/or offline.

4 CASE STUDY

Similar to how undergraduate courses are taught in a variety of institutions and contexts, the introductory computer science course at a large R01 university is taught by multiple instructors using standardized content and structure. Previous analytics methods could not differentiate between student experiences that were rooted in the common course structure and content versus those that were specific to an instructor. We used ELEVATE to investigate 4,440 student responses collected over 9 course offerings from four different instructors. Further analysis of the ELEVATE output identified two major patterns within the student comments: instructor (or pedagogy)-dependent topics (for example, topics 8 and 37 in Figure 3) and course-wide topics (for example, topic 42 in Figure 3).

For instructor-dependent topics, we see an uneven distribution of the comments, with most of the comments coming from classes taught by one (topic 37) or two (topic 8) instructors. In contrast to this, topics that are more about course-content or course-wide structure (topic 42) have a much more even distribution, indicating that these are not about individual instructor practices or pedagogical strategies. These insights can be used to promote discussions amongst the instructors about how to change overall course content or structure, or to increase adoption of successful pedagogies or strategies used by individual instructors. For example, the other instructors might consider increasing their use of group discussions and active learning strategies (like instructor D), based on the student comments seen for Topic 37.

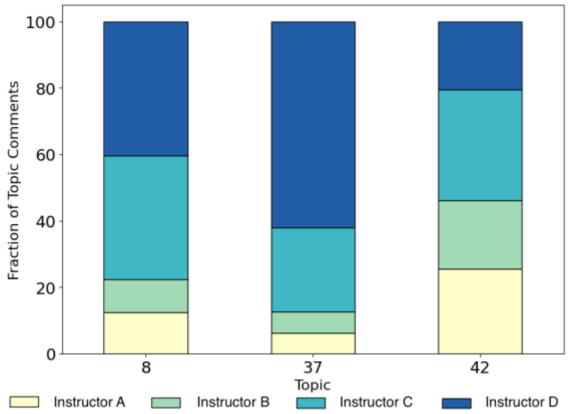


Figure 3: Distribution of comments within a specific topic across four course offerings. Topic 8 is *"Quiz and Assignment Issues."* **Topic 37** is "*Interactive Learning Experience with Group Discussion.*" **Topic 42** is "*Cognitive Challenge and Creative Thinking.*"

5 CONCLUSION

ELEVATE expands the learning analytics toolbox, allowing users to obtain new insights at the individual or institutional level that would have previously been impossible. By enabling the identification of key themes in large and very large SET datasets, instructors and programs can derive actionable insights to improve the student experience and better assess the impact of any changes to the curriculum or pedagogy. The case study presented here illustrates the benefits and potential applications of ELEVATE and highlights its power as a learning analytics tool. We expect to iteratively improve ELEVATE's performance while allowing others to access it to derive insights into the student experience in a variety of ways and across diverse contexts and disciplines.

REFERENCES

Dziuban, C., Moskal, P., Reiner, A., Cohen, A., & Carassas, C. (2023, November). Student Ratings: Skin in the Game and the Three-Body Problem. *Education Sciences, 13*, 1124. doi:10.3390/educsci13111124

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. Retrieved from https://arxiv.org/abs/2203.05794

Heffernan, T. (2022). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education, 47*, 144–154.

Sun, J., Yan, L. Using topic modeling to understand comments in student evaluations of teaching. *Discov Educ* 2, 25 (2023). https://doi.org/10.1007/s44217-023-00051-0

The Birth of an institutional Learner Analytics Platform

Author(s): N.K. Wilkin, C. Greenway & K. Hamilton
University of Birmingham, UK
n.k.wilkin@bham.ac.uk

ABSTRACT: If you do not have a learner analytics platform as an institution. Should you? And how might you design and implement it? This case study reviews a journey to launch of a learner analytics platform, at an institution with c40000 students. It considers the diverse needs and ambitions of all stakeholders. Central to all design decisions has been "how will this benefit the students?" We review our stepwise implementation, where each step considers the scale of change in terms of the combined parameters of awareness of the platform, digital literacy of stakeholders and signposting supportive actions.

Keywords: Learner Analytics launch, stakeholder analysis, student centered, data literacy.

1 WHO ARE THE PRACTITIONERS?

This case study reflects on the 'birth' of a learner analytics platform at a large comprehensive university. It will illustrate how via an inclusive design thinking process: simple yet strong foundations, can be created upon which an institution can iteratively design and build – with stakeholders actively engaged such that they themselves become practitioner-researchers in learner analytics.

Defining practitioner in the context of Learner Analytics is of itself a helpful exercise (Viberg & Grönlund, 2023). Who is invested in understanding the stakeholders needs, with what motivation and how is their expertise objective or one of lived experience? Figure 1 (*Becoming a practitioner-researcher*, n.d.), provides a method for charting the practitioners. The categorization will shed light on individual drivers for defining a successful or useful platform.

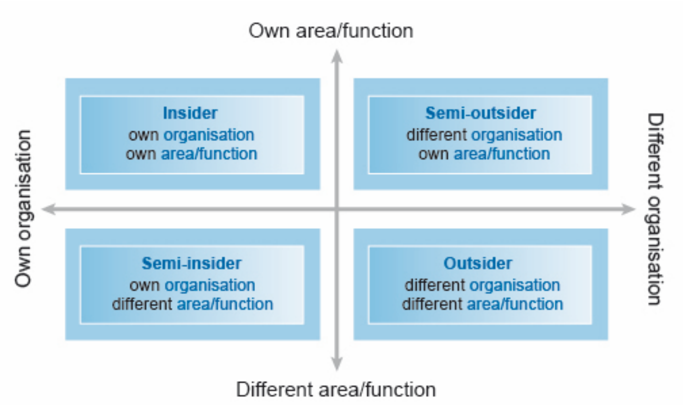


Figure 1: Classifying different types of practitioner-researcher, (*1 Becoming a practitioner-researcher*, n.d.)

Within a university community the categories of semi-insider, and insider are first identified.

Insiders, Practitioners of learning: students, are invested in their own journey and should also care about their data from a privacy perspective – thus learner analytics is a representation of their individual life journey but alongside this they must be empowered with the data literacy and signposting to the support their profile indicates, in order that they are best enabled to achieve.

Semi-insiders Practitioners, student facing 'education advisors', will be primarily concerned with understanding an individual student's circumstances and opportunities – although depending on the organizational set up they may be responsible for many individuals. These roles can be academic or professional services. Practitioners who are University leaders, will have strategic institutional priorities for learner analytics: outcomes from a cohort perspective; effective use of resources for a given discipline's delivery; and for ensuring compliance with national regulatory requirements.

Via conferences, and research papers, '*Outsiders*' expertise can be readily accessed. It requires an internal project lead to realize these resources exist, and to be academically curious to access them. Academic Researchers of Learner Analytics may focus on evaluating existing platforms where their research may lead to enhanced learning outcomes. Alternatively, they may be constructing a prototype with a tightly controlled student numbers, where practicalities of scale-up are not a requirement.

Remaining, are the '*semi-outsiders'*. One could consider them colleagues from different educational institutions who are able and willing to articulate their journey, successes, and challenges in constructing and driving adoption of a LA platform. This case study itself, seeks to be a '*semi-outsider*' for a university at the early stages of a LA platform implementation plan. We ourselves have proactively reached out and benefitted from expert advice from other institutions and individuals who have been generous with their time and advice (*Berkeley Online Advising (BOA): Transforming Data for Student Success*, n.d.; *LEARN Lab at NYU*, n.d.; *UTS:CIC*, n.d.)

2 WHAT ALREADY EXISTS LOCALLY?

The impetus for the project was needing to understand, via the data, how student cohorts were faring in the COVID pandemic from an educational engagement and perseverance perspective. The senior leaders realized that the data existed within the institution, but via many disparate sources with varying levels of accessibility for different stakeholders. Hence investment was authorized. For our institution, the choice was made to surface and summarize via a dashboard, the acknowledged complexity of (messy) data via an internally developed solution. The alternate route could have been to purchase a solution, e.g. specific software. Both have inherent challenges and benefits, the commercial, brings the '*outsider*' voice explicitly into the conversation, alongside additional drivers of commercial viability for the supplier and the challenge of further data complexity.

That the institutional education data was siloed is a long-standing issue, known by the education advisors across the university. Local learner dashboards had evolved, designed by frustrated '*semi-insiders*', who found the time to make a solution work for their local user case, but had no awareness of other practitioners who were doing similar across the university. Further, because this wasn't a commissioned project, knowledge of the existence of the local learner analytics (LA) processes, and their effectiveness was confined to the local departmental needs and from a technological perspective the systems are reliant on individuals remaining in place, and their technological competence.

A further group of academics were individually mining the self-serve analytics data from the virtual learning environment, and designing interventions based on their reading of the data, to support their own cohorts. Because this was considered educational support, as opposed to practitioner-research or ‘scholarship’, no time or process existed to bring these early adopters together to discuss their findings or how to enable the less data-confident to benefit from the self-serve resources. Alongside the mechanisms described above, students individually either self-requested support, spoke to their tutor, recorded low attendance, or failed formal examinations, and these became the ‘actionable’ insights to which the extensive university support services responded.

In summary, the non-existence of a ‘learner analytics dashboard’ did not correspond to a blank slate. Recognizing this and designing the project from a people rather than a technological perspective is seen as key to the project’s success. Driving the design from an actionable insights perspective, at the LA discipline origin(Siemens, 2013), *not*, ‘how complex or interesting is this data stream?’ *but* ‘how can the data surfaced be used to support an individual student?’ Further, by explicitly framing it as ‘student at the heart’, the project has to date been received positively by students.

3 PILOTING THE POSSIBLE

Initially, an internal technology development team was empowered to build a dashboard. They demonstrated that despite the complexity of the university data architecture it was possible to construct a LA dashboard with relevance to individual students. However, the project delivery, scoped from a technical perspective, did not consider how it would become embedded in people’s workflows from a supporting student perspective. At senior level the project was rescoped, with the paper authors as leads representing the people and culture first, technology enablement second, approach. Two are academics (Wilkin & Greenway), with successful research in their home disciplines, and can be considered practitioner-researchers in the LA field. They bring expertise of STEM research and education research within the school sector. Respectively they have responsibility for the digital ecosystem overall from an academic perspective, and for creating an effective manner of working for senior tutors across all disciplines. [1] The third (Hamilton) has a senior role responsible for student data systems and education support. Key to success is their combined prior skills, networks across the organization and silo-free team work (Tett, 2015). The decision was made to proceed via two routes.

Pilot the product designed ‘tech-first’ with a cohort of students who are in a preparatory university year. The associated education advisors and university support systems could be identified and supported to become digitally literate in the nuance of the data they were being presented with. This route led to the further development of the product that was much appreciated by the education advisors, and the students. Inevitably, the close working with the developers led to bespoke features that could not be universally rolled out because of the staff training that would be required in order that the data insights were correctly interpreted within different disciplinary norms.

Create a `no-ambiguity product’: if one considers that the actions taken from the data, are the important criteria, rather than complexity of the data, it becomes evident that ‘consistent actions

[1] The senior tutor is responsible for overseeing, supporting, and monitoring the tutorial system.

from the data' by all education advisors is paramount. Hence, a university strategy was agreed that requires that at each step of the LA journey, the whole university product requires an underpinning data-literacy plan for all users (students and staff) that is commensurate with the complexity of the system. This inevitably required a movement of people-resource from the technical, to digital adoption. The overarching strategy was to ensure that at 'birth' no data was visible for which statistical inference was required*. This led to the launch of a product whose original intent was 'learner analytics' with no 'analytics'.* However, the consolidated data, enables facilitated student support and actions that had not been easily possible previously. It has as an initial objective of data-facilitated tutor meetings, particularly from advisors who would not have self-served data previously.

4 FUTURE DEVELOPMENT

The launch of the easy to use 'foundational platform' is leading to the owners of the local learning analytics dashboards, those who successfully self-served data from the LMS and experts in inclusion from an education perspective, to enthusiastically become part of a community of practice across the university. Working with them, will enable the creation, in small, well-defined steps of a platform that will support all students. Speed of technical delivery will be throttled by ensuring that data-ethics (Li et al., 2022), data-literacy(Wolff et al., 2016) and signposting of support for students are prioritized. There is also opportunity, and a chance to engage from the LAK community(*Conceptual Framing*, n.d.). How would you, as '*outsiders*' (researchers) advise and evidence our future platform development steps? Given this case study how would you suggest your research conclusions are applied?

REFERENCES

Becoming a practitioner-researcher. (n.d.). Open Learning. Retrieved October 5, 2024, from https://www.open.edu/openlearn/money-business/using-data-aid-organisational-change/content-section-1

Berkeley Online Advising (BOA): Transforming Data for Student Success. (n.d.). EDUCAUSE. Retrieved October 5, 2024, from https://events.educause.edu/eli/annual-meeting/2021/agenda/berkeley-online-advising-boa-transforming-data-for-student-success

Conceptual Framing. (n.d.). Retrieved October 6, 2024, from https://sites.google.com/view/actionablela-lak24/conceptual-framing

LEARN Lab at NYU. (n.d.). Retrieved October 5, 2024, from https://wp.nyu.edu/learnlab/

Li, Q., Jung, Y., d'Anjou, B., & Wise, A. F. (2022, March 21). Unpacking instructors' analytics use: Two distinct profiles for informing teaching. *LAK22: 12th International Learning Analytics and Knowledge Conference*. LAK22: 12th International Learning Analytics and Knowledge Conference, Online USA. https://doi.org/10.1145/3506860.3506905

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *The American Behavioral Scientist*, *57*(10), 1380–1400.

Tett, G. (2015). *The silo effect: The peril of expertise and the promise of breaking down barriers*. Simon & Schuster.

UTS:CIC. (n.d.). Retrieved October 5, 2024, from https://cic.uts.edu.au/

Viberg, O., & Grönlund, Å. (Eds.). (2023). *Practicable learning analytics*. Springer International Publishing.

Wolff, A., Moore, J., Zdrahal, Z., Hlosta, M., & Kuzilek, J. (2016). Data literacy for learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. the Sixth International Conference, Edinburgh, United Kingdom. https://doi.org/10.1145/2883851.2883864

Inferring the Causal Relationships among Influence Factors on Student Confidence in Foreign Language Acquisition by Causal Discovery and Explainable AI approaches

Setsuro Matsuda[*1]
[1]Prefectural University of Kumamoto, Japan
[*]smatsuda@pu-kumamoto.ac.jp

ABSTRACT: In our university's questionnaire survey, the 4th graders almost always feel less confident in foreign language acquisition. Clarifying the reason for this phenomenon is difficult because the questionnaire was not directly related to language learning. Even if it were clarified, inferring the causal relationships among the factors influencing student confidence would be difficult without prior domain knowledge. Thus, an automatic causal discovery method, LiNGAM was applied to the questionnaire data analysis. After identifying promising influence factors and grasping their causal relationships, a Gradient Boosting Decision Tree (GBDT) model was trained to predict student confidence level from the influence factors. As the GBDT is a black-box AI model, Individual Conditional Expectations (ICEs) were used for explaining the relationship between the inputs and output of the model. The combination of the causal discovery and the explainable AI has revealed that students' self-evaluation for practical skills positively influences their confidence in foreign language acquisition and the GPA in the fall semester of the 4th grade does not.

Keywords: Foreign language acquisition, Influence factor, Gradient Boosting Decision Tree, Causal discovery, LiNGAM, Explainable AI, Individual Conditional Expectation

1 INTRODUCTION

Our university has conducted three types of questionnaire surveys, targeting the 1st, 2nd & 3rd, and 4th graders. At enrollment, the students agreed to their data being used for the purpose of research and education. The surveys last for more than a decade and almost always reveal that the 4th graders lack their confidence in foreign language acquisition. This is a serious problem because the university has striven to foster the students' global mindset and foreign language skills are a prerequisite for communicating with people all over the world. In the field of foreign language learning, researchers have reported that variables such as willingness to communicate, anxiety, locus of control, and self-efficacy are powerful predictors of foreign language performance (e.g., Yashima, 2002). However, our questionnaires were not designed to measure the influence of those variables on foreign language acquisition. Namely, the items in the questionnaires ask the students about more general topics in their school days. Therefore, the present study aims at checking if the quantitative analysis of data not directly related to foreign language learning can serve as a tool to know the reason why the students feel less confident about their foreign language skills.

2 QUESTIONNAIRE DATA AND AI MODEL

The subjects in the surveys is 444 students enrolled in 2016 and 110 of them positively answered to the question, "Have you improved your foreign language skills?" The 110 students were categorized into Group 1 and the remaining 334 students with negative answers into Group 0. The statistical analysis to check the significant difference between Groups 0 and 1, and the calculation of feature importances based on the binary classification between the two groups extracted seven promising predictors of the response to the question above: 3rd_use_eng, study_meaning, practical_skills, 4th_fall_GPA, 2nd_work_abroad, 1st_fall_eng, and international_activities. The 1st_fall_eng means the average score of English classes in the fall semester of the 1st grade. The question about international_activities asks if students considered how active international exchange is when choosing the university. The 2nd_work_abroad and 3rd_use_eng measure how much emphasis students put on opportunities to work abroad and to use English for their future career. The response to each question was rated on a Likert scale ranging from 1 (most negative) to 5 (most positive) and the GPA varies from 0 to 5. The author set a regression task where student confidence in foreign language acquisition (eng_prof_level) is an objective variable and the seven variables above are explanatory ones. In this task, Gradient Boosting Decision Tree (GBDT) was used as it has high predictive power without hyperparameters tuning and can handle missing values without any preprocessing.

3 INFERENCE OF CAUSAL RELATIONSHIPS

Because constructing the structural causal model for questionnaire data is difficult without prior domain knowledge, an automatic causal discovery method, LiNGAM (Shimize et al., 2006) was employed. Figure 1 shows an estimated causal graph among eight variables, including the objective variable. The numerals in the figure represent the path coefficient between two nodes. Edges with a coefficient less than 0.1 were removed. Considering the temporal relationship between variables, the direction of the edge between 4th_fall_GPA and 1st_fall_eng, and the one between 3rd_use_eng and 2nd_work_abroad appear to be reversed. However, most of the relationships among the variables seem valid. Particularly, the sequential relationship from study_meaning to practical_skills and toward eng_prof_level would be natural in terms of human motivation. Interestingly, 4th_fall_GPA has negative influence on eng_prof_level and so does study_meaning on 2nd_work_abroad. It should be said that the causal discovery by LiNGAM is helpful to grasp the overview of causal structure.

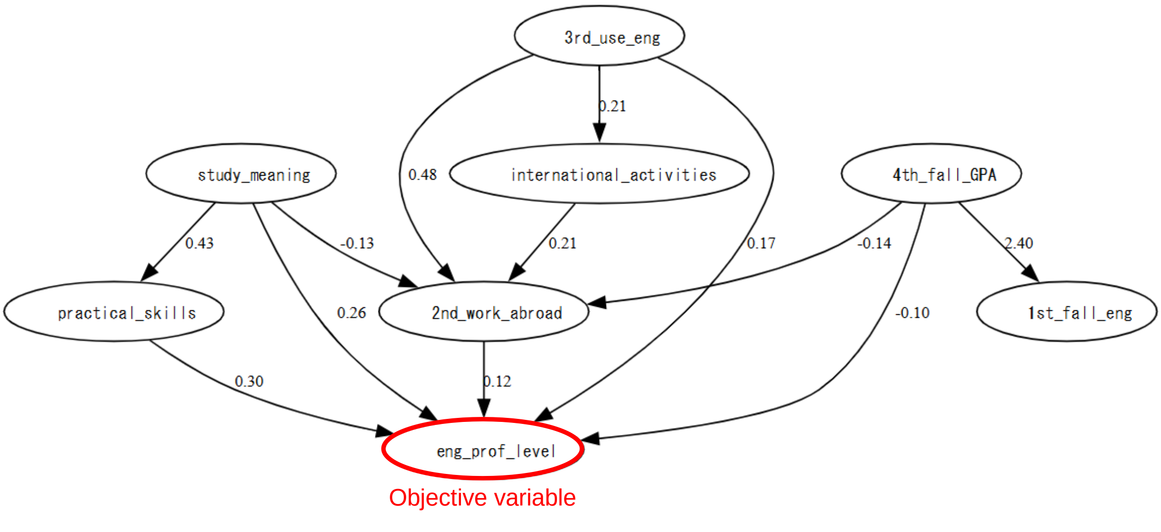


Figure 1: Causal relationships between the objective variable and the seven explanatory variables in the regression task.

4 EXPLAINING THE OUTPUT FROM AI

Although decision tree is a white-box model, GBDT is a black-box model because ensemble learning with multiple tree models has low interpretability due to the non-linearity of prediction. To explain how the GBDT-based model predicted the objective variable, the author used a model-agnostic explanation: Individual Conditional Expectation (ICE), applying the trained model to test data. The ICE represents the relationship between the output and the inputs for an individual instance. Figure 2 shows the ICEs of two variables: (a) practical skills and (b) 4th fall GPA. Figure 2(a) indicates that the higher the self-evaluation for practical skills becomes, the larger the output gets. It is noteworthy that there are three jumps in the output at the values of 2.0, 3.0, and 4.0 for practical skills. On the other hand, Figure 2(b) shows that the output remains flat or gets even smaller as the GPA becomes higher. This is surprising because other five variables except international activities are roughly in a proportional relationship with the output. The 4th graders with high GPAs may feel less confident about their English proficiency in contrast to their excellent graduation theses.

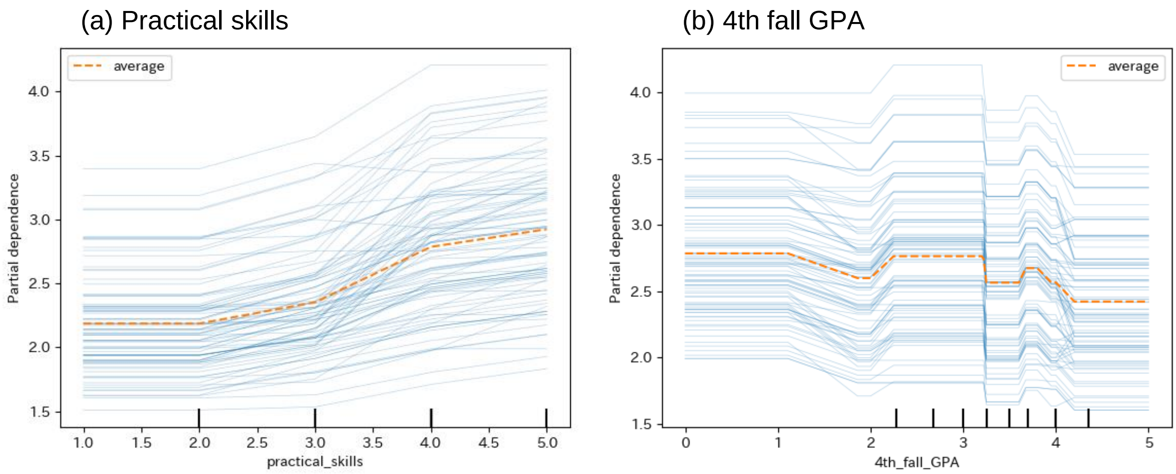


Figure 2: Individual Conditional Expectations of two variables: (a) practical skills and (b) 4th fall GPA. The dotted line in orange shows the average and corresponds to Partial Dependence.

5 CONCLUSION

The combination of the causal discovery by LiNGAM and the explanation of GBDT model by ICEs has clarified the influence factors on student confidence in foreign language acquisition and their causal relationships. The findings obtained can provide teachers with clues to improving their foreign language classes. For future work, SHAP (another explainable AI) is being applied to the predictive responses of students with considerably high and low confidence, connecting the explanation to the type of job they got after graduation.

REFERENCES

Shimize, S., Hoyer, P. O., Hyvärinen, A. & Kerminen, A. J. (2006). A linear-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7, 2003-2030.

Yashima, T. (2002). Willingness to communicate in a second language: the Japanese EFL context. *The Modern Language Journal*, 86, 54-66.

Digital Reading Affordances in E-Books: Effects of Highlighting, Annotation, and Tooltips on Student Comprehension

Anna Eunji Kim
The Pennsylvania State University
aek5559@psu.edu

ABSTRACT: Digital reading has become an intrinsic part of student learning. One of the key benefits of digital reading as compared to print reading is access to various affordances, including highlighting, annotation, and tooltips. However, studies investigating the effectiveness of these functionalities have been surprisingly limited. In this study, we examine what kinds of affordances in digital reading improve comprehension. In particular, we focus on two prevalent forms of affordances, (1) highlighting and annotation, and (2) tooltip access. We used a two-by-two experimental design with a sample of 179 undergraduate students at a large Northeastern University and measured students' comprehension with multiple-choice and open-ended questions. Results showed that affordance availability was not associated with multiple-choice comprehension performance, though tooltip access showed a significant effect on open-ended question performance. When students used available affordances, highlighting key words relevant to post-test questions was associated with better performance on multiple-choice questions. Students also had better performance when including summaries or restatements in annotations. Moreover, students' comprehension was positively related to the number of tooltips accessed, with significant correlations for both multiple-choice and open-ended questions. The implications of our findings and future research directions are discussed.

Keywords: digital reading, e-book affordances, reading comprehension, learning analytics

1 INTRODUCTION

Digital texts or e-textbooks have become widespread in higher education, often providing functionalities like highlighting, annotation, and tooltips to support active reading comprehension. Advocates of digital reading highlight these affordances for their potential to aid memory retention, engagement, and deeper understanding. However, research on the effectiveness of these affordances has yielded mixed results, especially in a digital context, where student interaction patterns may differ from traditional print-based methods. This study focuses on two common affordances: (1) highlighting and annotation; which allow students to mark important content and write notes, facilitating a deeper interaction with the material, (2) tooltip access; which provides supplemental definitions or explanations to enhance understanding without interrupting reading flow. Prior research on e-reading has mainly focused on comparing print and digital reading (Ben-Yehudah & Eshet-Alkalai, 2018) or has examined contextual affordances (e.g., highlighting) in isolation. Goodwin et al. (2020) examined highlighting, annotating, and students' use of online dictionaries in print and digital reading, reporting mixed results on the impact of digital affordances. This study aims to examine whether and how these tools support reading comprehension. We focus specifically on undergraduate students' use of these affordances, analyzing how different interaction types and frequencies affect comprehension outcomes. We have the following research questions: (1) What are the effects of highlighting/annotation tools and tooltips access on students' comprehension? (2) What is the association between highlighting and annotation use and students' comprehension? (3) What is the association between the number of tooltips accessed and students' comprehension?

2 METHODS

2.1 Participants

A total of 179 undergraduates enrolled in an introductory Educational Psychology course participated, with 163 students' data analyzed post-outlier removal. The sample comprised mainly freshmen (60.1%) and sophomores (27.6%), predominantly female (77.9%) and White/Caucasian (84%).

2.2 Procedures

The study had three main parts: a pre-test, a reading task, and a post-test. During the pre-test, participants were asked about their e-textbook use and prior knowledge of the two reading topics (i.e., parenting styles and peer social status). Students were then asked to read two textbook excerpts about two topics in developmental psychology, not covered in the course they were taking. Students were randomly assigned to one of four experimental conditions prior to reading. We used a 2 × 2 fully-crossed design (tooltip access available vs. unavailable; highlighting and annotation tools available vs. unavailable). Participants were assigned to one of four conditions: control, tooltip-only, highlighting/annotation-only, or combined (where both tooltips and highlighting/annotation tools were available). After reading the texts, participants were asked to answer reading comprehension questions. Examples of each type of affordance are shown below (Figure 1).

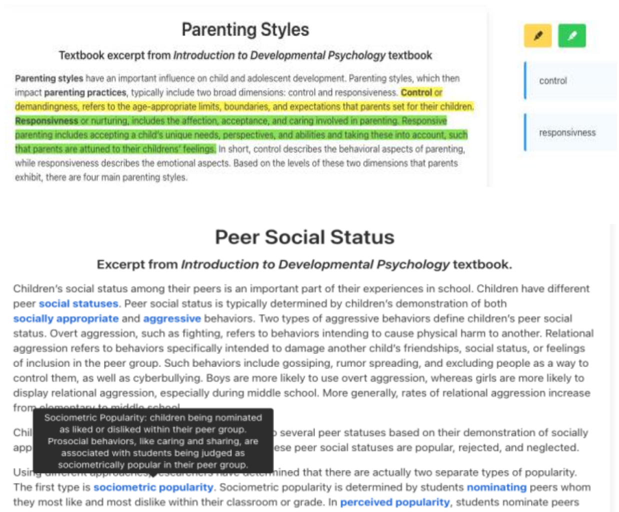


Figure 1: Examples of Affordances in Experimental Conditions (Highlighting and Annotation, and Tooltips)

We used ANOVA and regression analyses to examine the effects of conditions and specific affordance interaction on comprehension scores.

3 RESULTS

3.1 Effect of Affordance Access on Comprehension

No significant main effect was found for either the tooltip or highlighting/annotation conditions on multiple-choice comprehension scores. A significant main effect was observed for tooltip access on open-ended questions (F [1, 159] = 4.09, $p < .05$). Students with access to tooltips scored higher on open-ended questions, suggesting that additional contextual information facilitated a deeper understanding.

3.2 Highlighting and Annotations

Students who highlighted keywords relevant to post-test questions tended to perform better on multiple-choice questions ($r_{(87)} = .30$, $p < .01$), supporting the idea that focused highlighting correlates with better comprehension. Only certain types of annotations, such as summaries or restatements, were associated with higher comprehension scores, especially on multiple-choice questions ($r_{(87)} = .22$, $p < .05$).

3.3 Tooltip Access and Engagement

The number of tooltips accessed was positively correlated with both multiple-choice ($r_{(79)} = .28$, $p < .05$) and open-ended scores ($r_{(79)} = .25$, $p < .05$), indicating that accessing additional definitions and explanations contributes positively to comprehension.

4 DISCUSSION AND CONCLUSION

Our findings indicate that while simply having access to digital reading affordances did not automatically improve comprehension, students who actively used these tools showed better learning outcomes. Specifically, students who highlighted keywords relevant to post-test questions and used summarizing annotations showed better comprehension on multiple-choice tasks. Additionally, both the availability and frequency of tooltip use were associated with better performance, particularly on open-ended questions, suggesting that contextual aids can foster understanding by offering immediate, relevant information that reduces cognitive load. These results underscore the importance of not only providing digital tools but also guiding students in effective usage strategies. For educators and digital learning designers, the study highlights a need for adaptive e-book features that encourage meaningful interactions, potentially through prompts or recommendations for affordance use. Future research should explore the long-term effects of digital affordance use across varied content areas to better understand its impact on comprehension and retention.

REFERENCES

Ben-Yehudah, G., & Eshet-Alkalai, Y. (2018). The contribution of text-highlighting to comprehension: A comparison of print and digital reading. *Journal of Educational Multimedia and Hypermedia*, *27*(2), 153-178.

Goodwin, A. P., Cho, S. J., Reynolds, D., Brady, K., & Salas, J. (2020). Digital versus paper reading processes and links to comprehension for middle school students. *American Educational Research Journal*, *57*(4), 1837-1867. https://doi.org/10.3102/0002831219890300

Can We and Should We Support Student Well-being with LA?

Marije Goudriaan
Utrecht University
m.goudriaan1@uu.nl

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

Ünal Aksu
Utrecht University
u.aksu@uu.nl

ABSTRACT: Educational institutions are increasingly incorporating health and sustainability into their mission, thereby aiming to support student well-being. With the advances in learning analytics, researchers have begun to examine the potential of automatically collected study data as a source for monitoring and predicting well-being. This could enable pro-active interventions for a wide range of students, such as informing students about their well-being and offering suggestions to improve their wellbeing. In this paper, we describe the steps that have been taken so far at a Dutch University towards the development of a student-facing dashboard that supports student well-being. The paper consists of two elements: 1) We provide an overview of issues that have been raised in the literature concerning data, validity, and bias, potential negative effects, and student agency. 2) We provide a description of the initial prototype of the dashboard.

Keywords: Well-being, Student-facing dashboard, Ethical considerations, LMS trace data.

1 INTRODUCTION AND BACKGROUND

Well-being is a state "in which every individual realizes their potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to their community" (WHO, 2024). Increasingly, students in higher education experience low well-being, with detrimental effects on study success (Storrie et al., 2010). For example, in the Netherlands almost half of higher education students experience psychological problems such as anxiety and depression[1]. Therefore, educational institutions increasingly incorporate health and sustainability into their mission and strategic plans (Ahern, 2018). The heightened focus on well-being often translates to interventions such as taskforces, courses, and additional advisors for students to reach out to. These interventions can be characterized by a "on-demand" approach: it is up to the student to monitor their own well-being and to reach out for support. The disadvantage of this approach is that students are often not aware of such programs or are hesitant to reach out (Storrie et al., 2010). With the advances in learning analytics (LA) - the analysis of students' learning processes - the potential of

[1]https://www.trimbos.nl/aanbod/webwinkel/af2137-monitor-mentale-gezondheid-en-middelengebruik/.

automatically collected study data as a source for monitoring and predicting well-being is recognized (Kuijpers, 2022). Tracing well-being with the help of study data could enable pro-active interventions for a wide range of students, such as informing students about their well-being and offering individualized suggestions to improve their wellbeing. However, employing LA to support student well-being also raises a number ethical issues (Cormack & Reeve, 2022). We are investigating the possibilities of employing LA for supporting student well-being, with the ultimate aim to develop and evaluate a LA well-being intervention at our University.

2 LITERATURE EXPLORATION

2.1 Data, validity, and bias

Trace data collected from an LMS have been suggested to reflect changes in a student's well-being. However, previous research is limited and shows mixed results, indicating that LMS data may not be valid or sufficient to capture well-being (Kuijpers, 2022). Other studies have used self-reported well-being (Hossain et al, 2023). These self-reported measures might be more valid, but also have disadvantages. By simply asking how a student feels, the student might become more aware of their mood and act on it. Regardless of what data is used, higher education data are at risk of being biased, for example due to selection procedures, or because the data are initially collected with a different purpose (Ahern, 2018).

2.2 Potential negative effects

Potential negative effects of monitoring and visualizing student data related to well-being might also have unwanted effects. It could for example induce a feeling of unease or being monitored and result in a reduced sense of well-being (Cormack & Reeve, 2022). Furthermore, a well-being intervention might become a self-fulfilling prophecy, especially when a prediction is made for the student's progress (Prinsloo & Slade, 2016), thereby leading to counter-productive effects.

2.3 Role divison

The extent to which educational institutions should actively monitor and support well-being is also subject to discussion. It is not always possible educational institutions to take a larger responsibility in caring for students due to legislation, especially when it concerns medical data as could be the case for well-being (Ahern, 2018). Asking students to consent to participate in a well-being intervention seems a viable option. However, students are not always aware what they consent to, which can result in a biased dataset (Cormack & Reeve, 2022). Even if consent is used as a legal basis, it still requires a well-thought through intervention and support system. More research is needed to established recommendations for the issues outlined above. Also, educational institutions need to establish a code of practice in which they outline their viewpoints, preferably developed in collaboration with privacy officers and mental health care professionals.

3 DESIGN OF A STUDENT-FACING WELL-BEING LA DASHBOARD

Figure 1 shows a prototype of our student-facing well-being dashboard in which we tried to account for the challenges described in section 2. The dashboard is to be embedded into the landing page of the LMS for easy access. The basis for this prototype is our LA policy[2] in which the importance of validity of data metrics and student agency is emphasized. Based on the LA policy and previous research, we decided not to use individual clickstream data, but to offer well-being quizzes (self-reported well-being; panel A in figure 1). LMS data is only used on the level of planned activities (panel B). Panel C provides tips and courses on how to deal with periods of higher workload (from panel B) and well-being quiz outcomes (from panel A), thereby providing individualized suggestions for each student. The student well-being taskforce is currently writing a code of practice to clarify the role division and responsibilities in our University. During LAK25 we would appreciate input on how to further develop the dashboard, evaluate its impact, and mitigate potential negative effects.

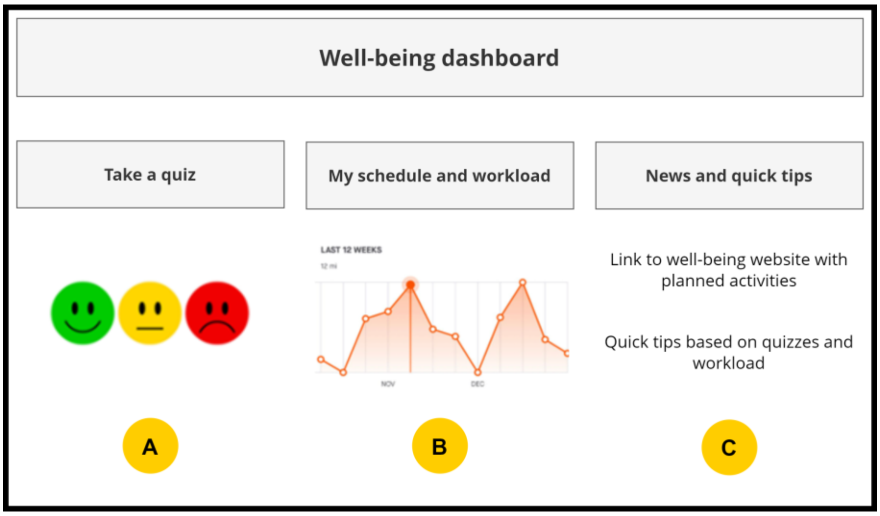


Figure 1: Prototype of a student-facing dashboard that supports well-being

REFERENCES

Ahern, S. J. (2018). The potential and pitfalls of learning analytics as a tool for supporting student wellbeing. *J. Learn. Teach. High. Educ.*, *1*, 165–172. https://doi.org/10.29311/jlthe.v1i2.2812

Cormack, A. N. & Reeve, D. (2022). Developing a Code of Practice for Using Data in Wellbeing Support. *Journal of Learning Analytics*, *9*(2), 253–264. https://doi.org/10.18608/jla.2022.7533

Hossain, A., O'Neilll, S., Stranova, I. (2023). What Constitutes Student Well-Being: A Scoping Review Of Students' Perspectives. *Child Indicators Research, 16,* 447 – 483. https://doi.org/10.1007/s12187-022-09990-w

Kuijpers, R.A.J. (2022). A New Step Towards Increased Mental Health of University Students Using Learning Analytics to Model Student Well-Being. Retrieved from https://research.tue.nl/

Prinsloo, P. & Slade, S. (2016). Student Vulnerability, Agency, and Learning Analytics: An Exploration. *Journal of Learning Analytics*, *3*(1), 159–182. https://doi.org/10.18608/jla.2016.31.10

Storrie, K., Ahern, K., & Tuckett, A. (2010). A systematic review: Students with mental health problems-A growing problem. *International journal of nursing practice*, *16*(1), 1-6. https://doi.org/10.1111/j.1440-172X.2009.01813.x

WHO (2024). https://www.who.int/data/gho/data/major-themes/health-and-well-being

[2] https://www.uu.nl/en/education/learning-analytics/la-policy-and-privacy-statement

Evaluating Feedback Strategies for Promoting Self-Learning

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

Frederic Tao Yi Chang
National Taichung University of Education
*snoopyfred@kimo.com

Chih Wei Yang
National Taichung University of Education
yangcw@mail.ntcu.edu.tw

ABSTRACT: This study compares the effectiveness of different feedback strategies provided by a learning platform through multiple-choice math tests on pre-class self-learning. A total of 460 5th-grade students participated using the Taiwan Adaptive Learning Platform (TALP), divided into four groups. All students completed a pre-test before the intervention. After watching instructional videos, all groups completed a multiple-choice test. TALP provided three types of feedback during the multiple-choice test: Group 1 received verification feedback indicating whether their answers were right or wrong; Group 2 was allowed to answer until correct with hints for incorrect responses; and Group 3 used S-TALPer, which delivered adaptive feedback generated by AI based on students' responses. The control group received no feedback on their answers. After the intervention, all students completed a post-test. The findings showed that all groups demonstrated improvement from pre-test to post-test, with the control group showing the smallest gain of 12.38%. The effect sizes, when compared to the control group, were 0.21 for Group 1, 0.45 for Group 2, and 0.86 for Group 3. These results suggest that feedback, particularly AI-generated adaptive feedback, significantly enhances self-learning. S-TALPer, combining GPT-4 and TALP's cross-grade diagnosis system, had the greatest impact, especially for low-achieving students.

Keywords: Feedback, Adaptive Feedback, Digital Learning, Self-learning, Generative AI

1 INTRODUCTION

Nowadays, an increasing number of learning activities take place on digital learning platforms, which are often equipped with instructional videos and quizzes. Research suggests that providing feedback on assessments after viewing videos can enhance students' retention of the material. In these platforms, multiple-choice tests are the most commonly used assessment format. Three types of feedback are typically applied: (1) Verification Feedback (Kulhavy & Stock, 1989): Offers a simple right or wrong indication without further explanation, providing immediate validation of the answer; (2) Elaboration Feedback (Shute, 2008): Utilizes an "answer-until-correct" approach by offering hints or explanations after incorrect responses, guiding students toward the correct answer through progressive reasoning; and (3) Adaptive Feedback (Narciss, 2008): Adjusts dynamically based on the learner's performance, providing personalized guidance and modifying the difficulty of subsequent questions.

2 METHODOLOGY

A total of 460 5th-grade students participated in this experiment, with 123 in the control group and 111, 95, and 131 in Groups 1, 2, and 3, respectively. After watching a 15-minute instructional video, students completed the same multiple-choice assessment, but each group received different feedback. Group 1 received verification feedback (right or wrong without explanation), Group 2 received elaboration feedback (answer until correct with hints), and Group 3 received adaptive feedback, delivered by S-TALPer, a system integrated into the Taiwan Adaptive Learning Platform (TALP). S-TALPer combines Generative AI (powered by GPT-4o) with TALP's cross-grade diagnosis system, which tracks students' learning weaknesses from previous grades. It provides personalized guidance and adjusts the difficulty of questions based on students' responses and performance. The content focused on 5th-grade math topics, and both pre- and post-tests were administered to assess learning gains.

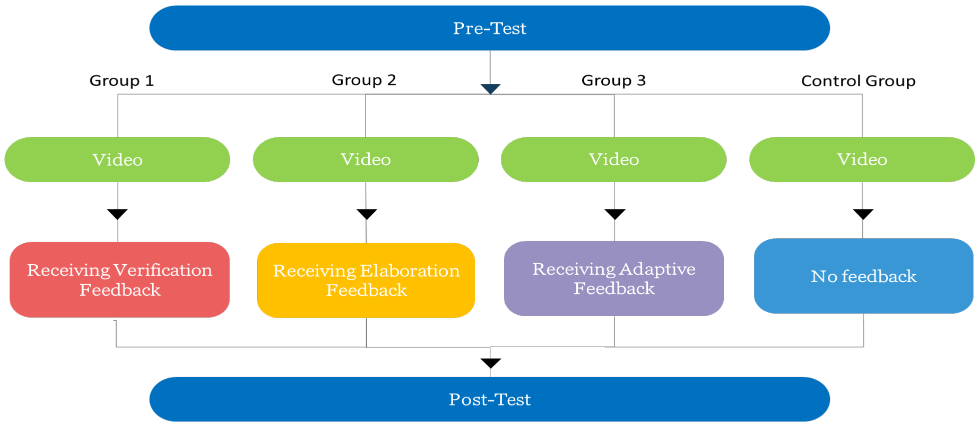


Figure 1: Experiment Design

3 RESULT

The scores of Pre-test and post -Test are shown as Figure 2, there is no significant difference across four groups in Pre-test (F=.654, df=3, p=0.58). As post-test was higher than pre-test in every group, even the control groups (no feedback) improved the least but still with 12.38%.

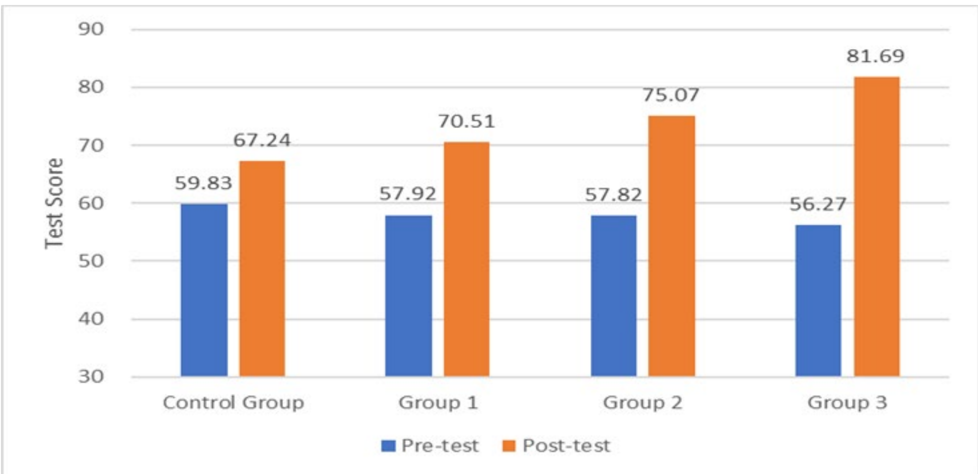


Figure 2: Pre-test and Post-test score across groups

An ANCOVA was conducted to evaluate the effectiveness of the three experimental groups, with no significant interaction found among the slopes (F=1.415, p=.238). The results revealed that the three types of feedback had a significant effect on enhancing learning outcomes. Specifically, Group 1 showed a marginal effect (F=3.99, p=0.47), while both Group 2 (F=17.37, p<.01) and Group 3 (F=61.19, p<.01) demonstrated statistically significant improvements. Figure 3(a) illustrates the effect sizes (Cohen's d) for each group: 0.21 for Group 1, 0.45 for Group 2, and 0.86 for Group 3. Group 3, which received adaptive feedback, significantly outperformed Group 2, which received elaboration feedback (p=0.03, with confidence intervals overlapping by less than half a bar), and Group 1, which received verification feedback (p<.01).

A closer inspection of the scatter plots comparing the performance of Groups 1, 2, and 3 with the control group reveals that the gap between the two lines (indicating the effectiveness of the assigned group) is relatively consistent across all pre-test score levels for both Group 1 (Figure 2(b)) and Group 2 (Figure 2(c)). However, for Group 3 (Figure 2(d)), the gap is notably wider, especially among students with lower pre-test scores, suggesting that the adaptive feedback was particularly beneficial for low achievers.

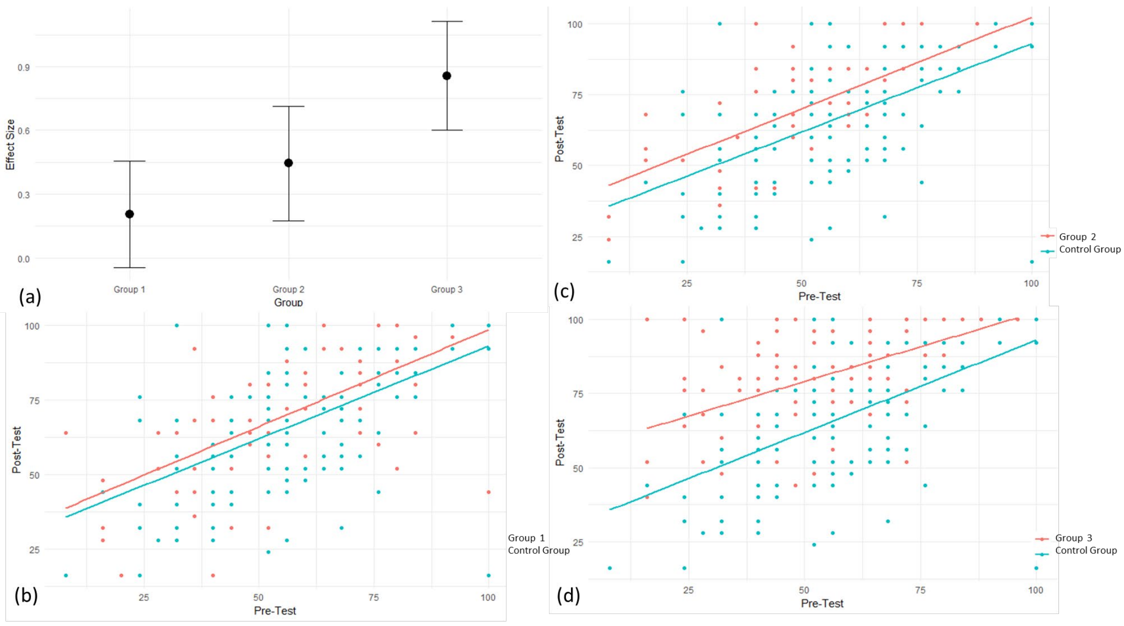


Figure 3: The effectiveness of feedback for experimental groups

REFERENCES

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279–308.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125–143). Mahwah, NJ: Lawrence Erlbaum Associates.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Academic Writing Enhanced by Generative Artificial Intelligence in a Virtual Forum

Diego Cheuquepán-Maldonado
Universidad de Santiago de Chile
diego.cheuquepan@usach.cl

Luz Chourio-Acevedo
Universidad de Santiago de Chile
luz.chourio@usach.cl

ABSTRACT: Posting a message on a virtual forum to present information is a daunting task for many students because academic writing requires a formal structure with a logical flow of ideas. While Generative Artificial Intelligence has emerged as a promising tool for supporting academic writing due to its capacity for idea generation, concept organization, and even text production, a research gap lies in understanding the effect on the quality of textual productions of students who use this technology to support their academic writing process. This study was conducted to explore the effect of Generative Artificial Intelligence on vocational education and training students' academic writing using text mining methods. We focused on two indexes that describe readability in the students' answers in an online course forum activity. Preliminary findings showed no significant difference between groups, so our results suggest that students in the group with the forum powered by Generative Artificial Intelligence did not achieve significantly higher readability than the other traditional forum groups. Future research will focus on other aspects of academic writing, such as coherence and cohesion.

Keywords: Generative artificial intelligence, academic writing, educational technology, online learning, text mining

1 INTRODUCTION

Academic writing is characterized by its formal tone, structured format, and precise language, which together aim to improve clarity and cohesion in the presentation of ideas. Unlike other types of writing, it requires a formal structure that accounts for a logical flow of the ideas being presented. For this reason, the academic writing process is often overwhelming for many students (Shin & Epp, 2023). Generative Artificial Intelligence offers potential benefits by assisting in idea generation, concept organization, and even text production, which could be especially beneficial for students who have difficulties with academic writing skills (Schmohl et al., 2020). Furthermore, recent studies show positive results of the integration of Generative Artificial Intelligence into academic writing skills, highlighting the importance of incorporating this technology in the educational field (Maphoto et al., 2024).

However, in this context, the existing research gap lies in understanding the effect on the quality of textual productions of students who use Generative Artificial Intelligence to support their academic writing process. Therefore, it is necessary to explore the potential of integrating Generative Artificial Intelligence to improve academic writing skills.

The general research question guiding this study was: To what extent does the use of Generative Artificial Intelligence affect the academic writing skills of Spanish-speaking vocational education and training students? We hypothesize that in the context of textual productions in online discussion forums, texts written by students with the support of Generative Artificial Intelligence (GenAI) would show greater textual difficulty than texts written by students without GenIA support since the formers rewrite their responses using AI-generated feedback.

Preliminary findings reveal no significant differences in textual difficulty between the groups. However, these initial results highlight the need to analyze other dimensions, such as coherence and cohesion, to gain a more comprehensive understanding of the impact of GenAI on academic writing. These aspects will be further studied in future phases of our study.

2 CONCEPTUAL FRAMEWORK

We used a conceptual framework that included three variables: academic task, academic writing skill, and the use of Generative Artificial Intelligence in forums.

The term academic task was defined as discussions in virtual forums assigned to students in an online course. Students could complete their academic task by writing their responses in one of the two types of forums enabled in the online course, the traditional forum and the forum supported by Generative Artificial Intelligence. To understand academic writing skill we considered the concept of textual difficulty as readability. Readability is the ease with which a text can be read and understood. To measure it, we used the Fernandez-Huerta Readability and Szigriszt-Pazos Perspicuity metrics (Checa-Moreno et al., 2021) using the Textstat Python package to calculate statistics from the text for the Spanish language. Both traditional metrics are limited to estimating the difficulty of the text by considering only two factors, the average length of words and sentences (McNamara et al., 2014). Finally, the use of Generative Artificial Intelligence refers to using the GenAI forum app in a Learning Management System to perform and accomplish an academic task.

3 METHOD

Students who participated in this study were given a consent form. Then, they were asked to answer a question in an online forum within a week. Participants who answered the academic task in an AI-powered forum received automated feedback while writing their answers so that they could modify them in the process of writing the answer while participants in the traditional forum did not have access to any kind of feedback to enrich their answers. This study followed a quasi-experimental design, comparing the linguistic readability of the text of writing between students using AI-powered forums and those in traditional forums, without random assignment to groups. The type of sampling was by cluster where the unit of analysis was the virtual classroom.

Table 1: Study participants.

Group	Students	Students valid	Num female	Num post
Traditional Forum	94	77	69	91
GenAI Forum	83	73	65	87

4 FINDINGS

To test our hypothesis, we analyzed texts written by students from a communication literacy online course who provided their responses in an online discussion forum. Student posts were analyzed using the Fernandez-Huerta Readability (FHR) and Szigriszt-Pazos Perspicuity (SPP) metrics to assess the linguistic readability of the text of academic writing. The data was analyzed using descriptive statistics and a Mann-Whitney U test was used to compare the distributions of readability indices between groups with traditional forums (Trad-F) and groups with forums powered by GenAI (GenAI-F) because the data did not follow a normal distribution. The test revealed that students in the group with the GenAI forum did not achieve significant difference from the other traditional group.

Table 2: Mann-Whitney U test results.

Measure	Trad-F Median	GenAI-F Median	U	p-value
FHR	81.62	83.66	3637.00	0.2330
SPP	77.81	80.16	3616.00	0.2618

5 DISCUSSION AND CONCLUSION

The study aimed to investigate the readability of the text in academic writing tasks in online forums focusing on detecting potential Generative Artificial Intelligence effects on academic writing. We conducted a text analysis involving methods such as Mann-Whitney U test and descriptive statistics. The outcomes of the study showed that there is no greater readability of the text in forums powered by GenAI. The study showed limitations. To assess academic writing skills only was considered readability, instead of other metrics and the text data analysis involved traditional methods rather than more advanced approaches. In future works to provide a more comprehensive understanding of other academic writing skills such as coherence, we will use some cohesion indices using Coh-Metrix 3.0 tool (McNamara et al., 2014). Additionally, future phases will expand the study by exploring GenAI-supported writing in face-to-face courses and evaluating academic writing in summative activities rather than formative forums to capture a broader range of contexts. These additional analyses will offer further insights into the educational value of GenAI in supporting academic writing skills.

REFERENCES

Bansal, S., & Aggarwal, C. Textstat on PyPI. Available online: https://pypi.org/project/textstat/ (accessed on 1 October 2024).

Checa-Moreno, V., Díaz-Mohedo, E., & Suárez-Serrano, C. (2021). Analysis of the readability of questionnaires on symptoms of pelvic floor dysfunctions adapted to Spanish. *International Journal of Environmental Research and Public Health, 18*(19).

Maphoto, K. B., Sevnarayan, K., Mohale, N. E., Suliman, Z., Ntsopi, T. J., & Mokoena, D. (2024). Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills. *Open Praxis, 16*(2), 142-159.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., y Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.

Schmohl, T., Watanabe, A., Fröhlich, N., & Herzberg, D. (2020, June 18-19). How can artificial intelligence improve the academic writing of students? In Conference proceedings. International conference The future of education, 10th edition. Florence, Italy.

Shin, J., & Epp, C. D. (2023). Understanding the effect of cohesion in academic writing clarity using education data science. In A. Peña-Ayala (Ed.), *Educational Data Science: Essentials, Approaches, and Tendencies: Proactive Education based on Empirical Big Data Evidence* (pp. 193-218). Singapore: Springer Nature Singapore.

Comparing Mirroring and Advising Dashboards: Teachers' Plans for Direct Instruction of Self-regulated Learning Strategies

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

ABSTRACT: Teachers play an important role developing effective self-regulated learning (SRL) strategies in young learners. However, previous research shows that many educators lack adequate SRL knowledge and do not systematically teach these strategies in the classroom. Digital tools, like teacher dashboards, offer valuable support to teachers by visualizing students' SRL processes, thereby aiding pedagogical decision-making. Dashboards can vary in the type and level of support they provide. To utilize these tools effectively, teachers need to interpret the data correctly and turn it into meaningful instructional actions. Thus, this experimental study examined teachers' plans for the instruction of SRL strategies by primary school teachers who used either a mirroring dashboard (n = 25) or an advising dashboard (n = 29). The results indicated no significant difference in monitoring accuracy between the two groups. Importantly, teachers using the advising dashboard showed significantly higher quality in their planned SRL strategy instruction. This finding highlights the impact of dashboard type on teachers' pedagogical choices and underscores the need to incorporate features that support SRL in dashboard design.

Keywords: teacher dashboard, dashboard type, self-regulated learning, primary education, direct strategy instruction

1 INSTRUCTION OF SELF-REGULATED LEARNING STRATEGIES AND TEACHER DASHBOARDS

Teachers play an important role in developing effective self-regulated learning (SRL) skills for young learners. Research indicates that teachers' direct instruction of SRL strategies positively influences primary school students' SRL skills (Dignath & Veenman, 2021). To support students' SRL, teachers need to monitor students' learning process, make informed decisions, and take appropriate pedagogical actions. This is particularly challenging in dynamic and large primary school classrooms. Besides, previous research reveals that teachers' knowledge and implementation of SRL strategies are often inadequate (e.g., Karlen et al., 2020). Digital tools, such as teacher dashboards, may facilitate this process by visualizing different phases of students' learning and providing SRL data (Wiedbusch et al., 2021). While these tools can offer valuable support, if teachers struggle to understand and interpret the data displayed, the dashboards may hinder rather than help their ability to support students (Hoogland et al., 2016). It is, therefore, crucial to examine teachers' use of dashboards and their influence on their pedagogical actions during the design phases. Dashboards can be categorized based on their types and levels of support. Both mirroring and advising dashboards enhance teachers' classroom awareness by visualizing students' learning processes. However, advising dashboards go further by also providing actionable recommendations (van Leeuwen & Rummel, 2019). Only a few studies addressed the impact of different types of dashboards on teachers' use of dashboards (van Leeuwen & Rummel, 2020) and pedagogical actions relating to SRL. Therefore, in this study, we aimed to compare the effect of mirroring and advising versions of our teacher dashboard prototypes on teachers' planned direct strategy instruction. These prototypes were developed for math subject based on Winne and Hadwin's (1998) COPES model of SRL phases. We employed an iterative co-design approach, including interviews with Dutch primary school teachers who teach math in upper-primary grades (ages 10-12) to align the dashboard information with teachers' pedagogical practices.

2 METHODOLOGY

An experimental vignette study using a between-subjects design was conducted to compare two versions of dashboard prototypes. Fifty-four Dutch primary school teachers (41 female, 13 male, M_{age} = 32.78, *SD* = 10.86) completed the study. Participants were randomly assigned to either the mirroring (*n* = 25) or advising (*n* = 29) dashboard condition. Teachers in the advising group were shown four vignettes depicting classroom scenarios in which the class had difficulties in different phases of SRL using medium-fidelity dashboard prototype that included additional suggestions integrated into the system. In contrast, the mirroring group received the same vignettes without any suggestions (see Figure 1). Dashboard information and suggestions were created based on the theoretical model. The dashboard shows information on students' self-reported motivation, prior knowledge, and goals. The number of assignments made, skill scores, and learning paths were also shown. An example suggestion for the goal-setting phase is: "You can explain, demonstrate, ask, and remind students how to set realistic learning goals and plan effectively. For example, students who over- and underestimate themselves can learn to take into account their prior knowledge, performance, and standards when setting goals.". Monitoring accuracy was assessed by evaluating teachers' ability to identify problems within the vignettes correctly. Teachers were asked to describe their planned instructional strategies through open-ended questions. These were later coded by the researchers based on the set quality requirements. The materials can be found here. Their visualization literacy skills were measured using Mini-VLAT (Pandey & Ottley, 2023). As teachers' visualization skills may impact their understanding and interpretation of data, their Mini-VLAT scores were operationalized as a control variable. Finally, teachers completed a questionnaire to assess their perceptions of the dashboard use.

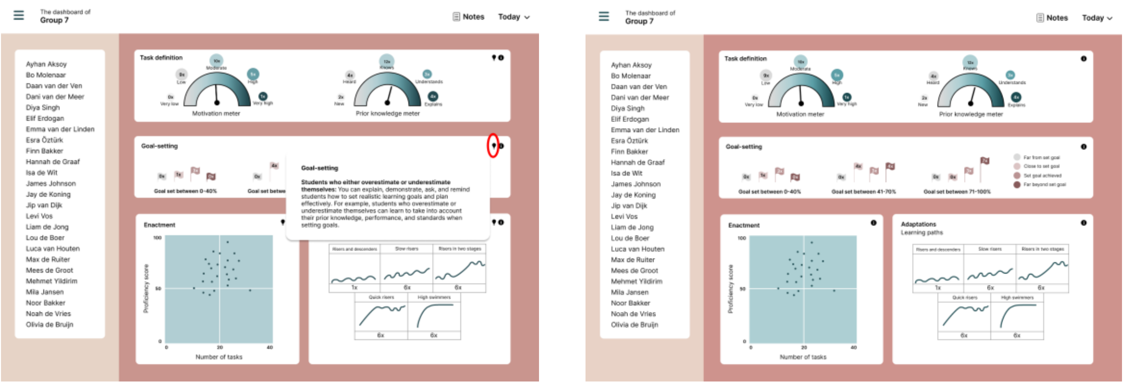


Figure 1: Advising (left) and mirroring (right) dashboard vignettes: The mirroring version omits the light bulb icon for suggestions with identical values. The dashboard was translated into English.

3 FINDINGS AND IMPLICATIONS

Prior to analyses, assumptions were checked to determine the appropriate approach. Preliminary results showed no significant differences in visualization literacy scores between teachers in the mirroring (M = 8.40, SD = 1.44) and advising dashboard conditions (M = 9.00, SD = 1.36), $t(52) = -1.57$, p = .12, 95% CI [-1.37, 0.17]. The effect size, as measured by Cohen's d, was -.43, suggesting a small effect. Thus, we chose to not control for this variable. Similarly, no significant differences were found in monitoring accuracy in mirroring (M = 1.85, SD = 0.98) and advising (M = 2.09, SD = 0.87) dashboard conditions (U = 314.50, p = 0.41). However, there was a significant difference in the quality of planned strategy instruction scores (U = 216.5, p = .01*), with advising dashboard group teachers (M = 4.38, SD = 2.90) scoring higher than those in the mirroring group (M = 2.36, SD = 2.14). The effect size, as measured by Cliff's delta, was -.40, indicating a medium effect. The significant difference in the quality of planned strategy instruction scores favoring the advising dashboard condition suggests that the design of the dashboards may influence teachers' pedagogical decisions. This preliminary finding supports the notion that providing actionable insights regarding SRL has the potential to improve teachers' instructional practices. We will also code and analyze teachers' professional knowledge of SRL as controlling variable, as teachers' prior knowledge regarding SRL and its implementation may influence the results. Teachers' behavior patterns using dashboards will also be investigated using process mining techniques to refine our analysis and interpretations further.

REFERENCES

Dignath, C., & Veenman, M. V. J. (2021). The role of direct strategy instruction and indirect activation of self-regulated learning - Evidence from classroom observation studies. Educational Psychology Review. Springer. https://doi.org/10.1007/s10648-020-09534-0

Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. Teaching and Teacher Education, 60, 377–386. https://doi.org/10.1016/j.tate.2016.07.012

Karlen, Y., Hertel, S., & Hirt, C. N. (2020). Teachers' professional competences in self-regulated learning: An approach to integrate teachers' competences as self-regulated learners and as agents of self-regulated learning in a holistic manner. Frontiers in Education, 5. https://doi.org/10.3389/feduc.2020.00159

Pandey, S., & Ottley, A. (2023). Mini-VLAT: A short and effective measure of visualization literacy. Computer Graphics Forum, 42(3), 1–11. https://doi.org/10.1111/cgf.14809

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 engagement in learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.). Metacognition in Educational Theory and Practice, 277-304. Hillsdale: Erlbaum.

van Leeuwen, A., & Rummel, N. (2019). Orchestration tools to support the teacher during student collaboration: A review. In Unterrichtswissenschaft (Vol. 47, Issue 2, pp. 143–158). Springer VS. https://doi.org/10.1007/s42010-019-00052-9

van Leeuwen, A., & Rummel, N. (2020). Comparing teachers' use of mirroring and advising dashboards. ACM International Conference Proceeding Series, 26-34. https://doi.org/10.1145/3375462.3375471

Revealing Individual Differences in EFL Learners' Reading Engagement from Interaction Logs

Mihwa Lee, Björn Rudzewitz, and Xiaobin Chen
University of Tübingen, Germany
{mihwa.lee, bjoern.rudzewitz, xiaobin.chen}@uni-tuebingen.de

ABSTRACT: Student engagement is a critical component for reading in a second/foreign language (L2). However, traditional methods of assessing engagement often rely on subjective measures (e.g., self-reports) or intrusive techniques (e.g., eye-tracking). This study explores learners' L2 reading engagement in online environments by analyzing interaction logs of English as a Foreign Language (EFL) learners. Using 8,076 data points of click-stream data from an intelligent computer-assisted language learning (ICALL) system, we investigated the reading behavior of 15 students over two weeks. Engagement scores were used to model students' L2 reading engagement, as well as explore the relationship between behavioral variables and L2 reading comprehension performance. Results showed that most students maintained moderate levels of engagement, while a few exhibited sustained high or fluctuating engagement. It further revealed that certain behavioral metrics significantly predicted the performance. These findings highlight the potential of interaction logs to uncover individual differences in L2 reading engagement, providing students and teachers with actionable intelligence.

Keywords: Reading engagement, Learning behavior, Individual differences, EFL, ICALL

1 INTRODUCTION

As English continues to dominate globally, the need for students to develop strong reading skills in English as L2 has become increasingly important. Student reading engagement—the behavioral expression of effort, time, and persistence toward achieving specific reading goals (Guthrie et al., 2012)—significantly impacts L2 literacy. Recent research highlights a positive correlation between reading engagement and achievement in reading comprehension (Zhu et al., 2023). Understanding reading engagement is particularly important in technology-mediated environments, where students often need to self-regulate their learning without direct teacher oversight. In such contexts, identifying factors that foster reading engagement is a key to promoting better learning outcomes. While traditionally, student reading engagement has been assessed using psychological questionnaires, these methods face criticism for their subjectivity and reliance on self-reports. Although eye-tracking has been explored as an alternative (e.g., Child et al., 2020), it poses challenges, including disrupting natural reading behavior and limited applicability in real-life learning contexts. Interaction logs have the potential to provide valuable insights into learners' engagement patterns without the drawbacks of more intrusive methods, yet research on L2 reading engagement using this approach remains scarce. We conducted a pilot study on EFL learners' L2 reading engagement within a computer-mediated environment, focusing on two research questions: *How can interaction logs from an ICALL system provide insights into learner engagement during L2 reading? To what extent do engagement metrics, as extracted from interaction logs, account for L2 reading comprehension performance?*

2 METHODOLOGY

The study utilized 8,076 click-stream data points of 15 university students (F = 4, M = 4, unspecified = 7) from a web-based ICALL system called *ARES* (Lee et al., 2024) that provides interactive support for L2 reading, such as glossing on language means and vocabulary. Among respondents to the background questionnaire, the mean age was 35 years (*SD* = 18.07), with English proficiency ranging from B1 to C1 on the CEFR scale. Over a two-week period, students completed eight reading assignments (mean length = 558 words, *SD* = 26.2). Each assignment accompanied six comprehension questions (three factual, three inferential). Four assignments were due weekly, with feedback provided after submission. In order to answer the first RQ, engagement metrics were defined based on the behavioral features in Table 1, originated from widely used metrics in navigational analysis in reading behavior (e.g., Ma et al., 2024), and transformed via percentile rank per assignment and per learner (ranging 0 to 1) using a formula introduced in Boticki et al. (2019) to account for outliers and to combine diverse data sources. Each assignment's total engagement score was the sum of all variable values (ranging from 0 to 10). L2 reading comprehension performance was measured by the percentage of correct answers to the reading comprehension questions per assignment and the relationship between the performance and engagement was calculated using Spearman's correlation. Concerning the second RQ, we performed multiple linear regression analysis in order to determine which engagement metrics account for L2 reading comprehension performance.

Behavioral variable	**Description**
Access to assignment	Total count of the access to an assignment
Time (min.)	Sum of the total time spent on an assignment
Access frequency	Frequency of the access to an assignment per week
Question open	Total count of reading comprehension questions opened in an assignment
Question completion	Total count of reading comprehension questions completed in an assignment
Feedback open	Total count of feedback opened in an assignment
Own grade open	Total count of the individual grade opened in an assignment
Average grade open	Total count of the class average grade opened in an assignment
Help open	Total count of explanations of vocabulary and language means opened
Finish time (min.)	Difference between the time of the assignment submission and the deadline

Table 1: List of behavioral variables used in calculating students' engagement

3 RESULTS AND FINDINGS

Figure 1 illustrates the transition of total engagement scores and performance scores per student over the two-week period. It reveals that while most students maintained moderate engagement levels, several students exhibited stable low (ID = 28) or decreased engagement (ID = 17, 22, 23), highlighted in red in Figure 1. Notable exceptions include certain students who demonstrated consistently high engagement (ID = 30) or an increase (ID = 31, 21) in engagement, indicating significant individual differences in L2 reading engagement patterns, despite the participants having similar ranges of EFL proficiency levels. The consistent positive correlation between engagement and performance across assignments shows students with higher engagement scores tended to achieve better performance. The results of the correlation analysis between the engagement score and performance score revealed that although there was no correlation observed in the first week, a strong, positive correlation was observed in the second week ($\rho = 0.61$, $p < .001$). The regression analysis revealed that *question completion* ($p < 0.001$), *finish time* ($p < 0.002$), and *access to assignment* ($p < 0.036$) had a significant positive impact on comprehension performance ($R^2 = 0.846$, $F(10, 109) = 60.07$, $p < 0.001$.). These findings highlight the critical role of active task engagement, timely completion of tasks, and active access in contributing to L2 reading performance.

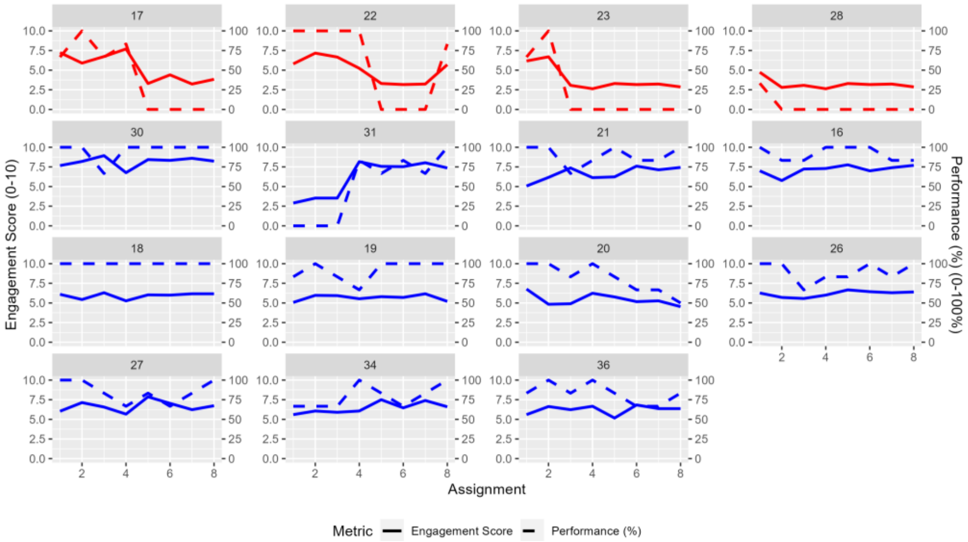


Figure 1: Transition of engagement scores and performance scores per student

4 CONCLUSION AND LIMITATIONS

Our analysis showed that interaction data from the ICALL system can provide an in-depth understanding of individual differences in EFL learners' L2 reading engagement patterns, which is a key behavioral predictor of performance. However, the small sample size and short learning period in this study limits the generalizability of these findings, necessitating future research with larger datasets and longer learning period. Despite these limitations, our approach of unobtrusive and continuous tracking can reveal unique engagement patterns among EFL learners and provides actionable insights to both students and teachers.

REFERENCES

Boticki, I., Akçapınar, G., & Ogata, H. (2019): E-book user modelling through learning analytics: the case of learner engagement and reading styles, *Interactive Learning Environments, 27*(5-6), 754–765. https://doi.org/10.1080/10494820.2019.1610459

Child, S., Oakhill, J., & Garnham, A. (2020). Tracking your emotions: An eye-tracking study on reader's engagement with perspective during text comprehension. *Quarterly Journal of Experimental Psychology*, *73*(6), 929–940. https://doi.org/10.1177/1747021820905561

Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601–634). Springer. https://doi.org/10.1007/978-1-4614-2018-7_29

Lee, M., Rudzewitz, B., & Chen, X. (2024). Developing a Pedagogically Oriented Interactive Reading Tool with Teachers in the Loops. In *Proceedings of the 13th Workshop on NLP for Computer-Assisted Language Learning (NLP4CALL)* (pp. 115-125). LiU Electronic Press. https://doi.org/10.3384/ecp211009

Ma, Y., Cain, K., & Ushakova, A. (2024). Application of cluster analysis to identify different reader groups through their engagement with a digital reading supplement. *Computers & Education, 214,* 105025. https://doi.org/10.1016/j.compedu.2024.105025

Zhu, A., Mofreh, S. A. M., Salem, S., Li, Z., & Yao, M. (2023). A Review of the Effect of Reading Engagement on Reading Achievement. *Encyclopaideia*, *27*(67), 17–28. https://doi.org/10.6092/issn.1825-8670/16180

The human-centered design of a feedback report for pupils and their parents in standardized tests in primary and secondary education

Amber Hoefkens
AP University College Antwerp, Belgium
amber.hoefkens@ap.be

Yasmine Wauthier
AP University College Antwerp, Belgium
yasmine.wauthier@ap.be

Sabrina Govaerts
AP University College Antwerp, Belgium
sabrina.govaerts@ap.be

Elena Van den Broeck
AP University College Antwerp, Belgium
elena.vandenbroeck@ap.be

ABSTRACT: From 2023-2024, standardized tests for reading and mathematics have been introduced in primary and secondary education in Flanders (Belgium) to support school development and enhance educational quality. An inter-university consortium is responsible for test design, implementation, results analysis and feedback distribution to school leaders, teachers, pupils, and parents. As part of the consortium, our aim is to design relevant and user-friendly feedback instruments to disclose the results of these tests to the different user groups. This poster focuses on the human-centered design of a relevant and user-friendly feedback report for pupil and parents. Following an Educational Design Research ('EDR'; Phillips & Dolle, 2006) approach, a first prototype was developed based on government guidelines and evaluated through semi-structured interviews with pupils, parents, and educational professionals. The prototype was optimized and tested in a second cycle. In June 2024, pupils received the feedback report for the first time, followed by a third evaluation cycle using eye-tracking and skin conductance studies, along with cued recall tests and follow-up interviews. This poster will present the research findings and design iterations to date, highlighting how they informed content and interface design and discussing methodological challenges encountered.

Keywords: Educational Design Research, User-centered feedback, Primary and secondary education, Eye-tracking, Skin conductance

1 EXTENDED SUMMARY

From 2023-2024 onwards, standardized reading and mathematics tests were introduced in primary and secondary schools across Flanders (Belgium). These assessments aim to promote school improvement and enhance educational quality. To support their implementation, a Support Centre was established, led by an inter-university consortium responsible for the design, administration, and analysis of these tests. The consortium also provides digital feedback to key stakeholders, including school principals, teachers, pupils, and parents. As part of this team, we are responsible for designing a feedback report that communicates test results to pupils and parents in a clear and user-friendly manner.

Effective feedback, as Hattie and Timperley (2007) describe, provides external information about a person's performance or understanding, which, if applied effectively, can serve as a powerful learning tool (Hattie, 2008). However, this impact requires feedback to meet certain conditions. Based on prior research (Van Gasse et al., 2015), effective feedback systems must (1) be relevant to users, (2) provide actionable insights by offering performance information and clear directions for improvement, (3) facilitate accurate interpretation, and (4) present information in a clear, easy-to-use format. In designing a feedback report for pupils and parents, the author prioritized these principles to make the feedback both valuable and accessible for understanding test performance. This goal involved a focus on both content (e.g., test results, data visualizations) and the user interface (e.g., layout, structure, language).

To achieve this, we adopted a mixed-methods approach using Educational Design Research ('EDR'; Phillips & Dolle, 2006). EDR combines research and design to address complex educational challenges by iteratively developing, testing, and refining tools and strategies. This approach enables the continuous improvement of the feedback report based on user feedback, ensuring that the final product meets users' needs and expectations. As McKinney and Reeves (2012) emphasize, the integration of research and design strengthens both by ensuring that each phase improves the other.

The feedback report's development and evaluation process followed three EDR cycles. In the first cycle, we designed a prototype of the feedback report based on government guidelines. This prototype aimed to present test results clearly while addressing the needs of pupils and their parents. In January and February 2024, semi-structured online interviews were conducted with a variety of stakeholders, including parents, parent associations, and pupil union representatives (N=11). These interviews offered feedback on content, layout, ease of interpretation, and language, as well as key elements needed to make the report meaningful to users. The results informed the first round of prototype optimization, which was tested in a second evaluation cycle.

In the second cycle, the optimized report was evaluated through further semi-structured interviews with pupils (N=4), providing direct insights into how pupils interacted with the report and interpreted the information. Feedback from these interviews guided additional improvements, refining content presentation, user interface layout, language, and data visualizations. Through these first two cycles, we could iteratively adjust the feedback report, ensuring it met user expectations and addressed the needs and challenges identified by both parents and pupils.

In June 2024, the feedback report was officially distributed to pupils, marking the beginning of the third EDR cycle. This phase used a mixed-methods approach to gain deeper insights into users' attention, emotional responses, and interpretation. The cycle combined eye-tracking and skin conductance technology to capture detailed interaction data. Eye-tracking technology provided data on participants' visual attention by recording where users looked, how long they focused on specific areas, and which sections they skipped. Eye-tracking data, including fixation duration, saccades (movements between fixations), and total gaze time on various parts of the report, revealed which sections drew the most attention and which parts were less engaging.

Skin conductance measurements complemented eye-tracking data by recording physiological responses that indicated users' emotional reactions to specific parts of the report. By correlating skin conductance data with eye-tracking moments, researchers could identify which sections triggered stronger emotional responses, possibly indicating confusion, surprise, or relevance.

After completing the eye-tracking and skin conductance sessions, each participant engaged in a cued recall session. Here, participants reviewed their eye-tracking and skin conductance data with the researcher, who asked follow-up questions to clarify specific attention and emotional responses. This process helped determine how participants interpreted specific data and texts in the report, how user-friendly they found the format, and what support they might need to understand the report fully. An interview guide structured these sessions, covering important aspects of the report. Key areas were identified in advance, enabling targeted questions on specific sections that prompted longer fixations or noticeable emotional reactions.

Throughout the cued recall sessions and interviews, the researcher documented key observations, including verbal feedback, participant behaviors, and non-verbal reactions that shed light on user experience. Participants' answers were transcribed into a reporting template to facilitate data synthesis and analysis. Meanwhile, eye-tracking and skin conductance data were analyzed using Tobii Pro Lab software for detailed visual and quantitative insights into attention patterns and emotional responses. The results indicated that while the report was clear and understandable for both pupils and parents, it lacked essential information needed for an accurate interpretation of the findings. Additionally, we gathered extensive feedback on the different specific sections of the report. These results will inform further refinements to the feedback report in preparation for standardized test reporting in 2025.

This poster will present findings from the three EDR cycles and showcase the current feedback report prototype. Additionally, we will discuss methodological challenges encountered during the third cycle, particularly with eye-tracking and skin conductance technology, and implications for future design and research efforts.

REFERENCES

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

Hattie, J. (2008). Visible Learning: A synthesis of Over 800 Meta-Analyses Relating to Achievement. London: Routledge.

McKenney, S. & Reeves, T. (2012). Conducting Educational Design Research. London: Routledge.

Phillips, D. & Dolle, J. (2006). From Plato to Brown and beyond: Theory, practice, and the promise of design experiments. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds), Instructional Psychology: Past, present and future trends. Sixteen essays in honout of Erik De Corte (pp. 277-292). Oxford: Elsevier.

Van Gasse, R., Vanhoof, J., Mahieu, P., & Van Petegem, P. (2015). Informatiegebruik door schoolleiders en leerkrachten. Maklu.

The Stayers, Stragglers, and Slippers: Tracking Student Journeys in MOOC Certification Programs

Ji Hyun Yu
University of North Texas
Jihyun.yu@unt.edu

ABSTRACT: This study explores learner engagement patterns in MOOC-based professional certification programs through longitudinal analysis, logistic regression, predictive modeling, and sequential pattern mining. Analyzing data from 1,539 learners across three sequential courses, three distinct engagement trajectories were identified: Consistently High Engagers ("Stayers", 50.1%), Persistent Low Engagers ("Stragglers", 26.2%), and Initial High Engagers with Later Decline ("Slippers", 23.7%). Contrary to common assumptions, reading completion alone did not significantly predict success; rather, early lecture engagement, consistent quiz participation, and involvement in peer-reviewed assignments emerged as critical predictors. Although these results show strong associations, they do not establish causation. Random Forest modeling achieved high predictive accuracy (0.794), with late-stage quiz completion emerging as a key indicator, and sequential pattern analysis uncovered specific engagement sequences tied to course completion. These findings highlight the potential for targeted interventions and offer practical implications for designing effective MOOC-based certification programs.

Keywords: MOOC Certification Programs, Engagement Pathways, Predictive Modeling, Longitudinal Retention

1 BACKGROUND

MOOC-based certification programs offer flexible learning but often face high dropout rates, akin to single-course MOOCs (Joksimović et al., 2018). Despite Coursera's Specializations and similar initiatives (Eriksson et al., 2017), motivation and burnout remain challenges. Engagement—behavioral, emotional, and cognitive—remains underexplored in multi-course settings (Reich & Ruipérez-Valiente, 2019), and solitary tasks like reading may not sustain motivation (Kizilcec et al., 2013). This study examines a six-month certification program spanning three sequential courses (September 2020–April 2023). Using transition matrices, logistic regression, Random Forest, XGBoost, and sequential pattern mining, I analyzed learner engagement in three states: Not Started (NS), Incomplete (IC), and Complete (CP). Results suggest adaptive and collaborative interventions can bolster long-term engagement, though contextual factors may limit generalizability.

2 ENGAGEMENT PATTERNS ACROSS COURSES

Three distinct engagement patterns emerged: Consistently High Engagers ("Stayers," 50.1%), Persistent Low Engagers ("Stragglers," 26.2%), and Initial High Engagers with Later Decline ("Slippers," 23.7%). These categories were derived from combined activity frequency (lectures, quizzes, peer reviews) and progression metrics (NS, IC, CP) via threshold-based segmentation. Learners with steady engagement were likelier to finish, while those with declining participation often did not. As shown in

Figure 1 (with enlarged axis labels), transition matrices revealed a significant drop between the second and third phases, aligning with findings on engagement decay (Reich & Ruipérez-Valiente, 2019). This drop highlights the challenge of sustaining long-term motivation and suggests context-specific factors (e.g., course difficulty, scheduling) may affect generalizability. It also underscores the need for adaptive pacing and continuous support to maintain engagement throughout multi-course programs.

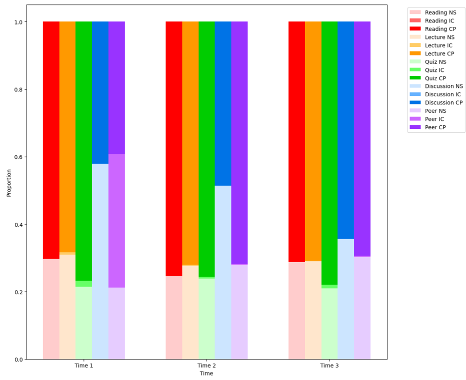


Figure 1: Normalized Stacked Bar Chart of Engagement Proportion Across Time

3 PREDICTORS OF PROGRAM COMPLETION

Using logistic regression and Random Forest (see Table 1), we found that early quiz attempts and timely submissions of peer-reviewed assignments correlated strongly with completion, while discussion forum participation also positively influenced outcomes. However, reading completion was not significant (Table 1), suggesting interactive activities may be more pivotal in extended certification contexts. Although these indicators are highly predictive, we emphasize that correlation does not imply causation. Nonetheless, they offer actionable insights: focusing on early engagement and facilitating interactive tasks can help sustain motivation throughout longer programs.

Table 1: Logistic Regression Results

Variable	VIF	Coefficient	Std. Error	z	P>|z|	95% Conf. Interval
Constant	-	-0.9530	0.178	-5.359	0.000	-1.302 ~ -0.604
Reading	8.731851	-0.2361	0.216	-1.091	0.275	-0.660 ~ 0.188
Lecture	7.912747	0.2188	0.190	1.152	0.249	-0.154 ~ 0.591
Quiz	5.945734	2.3981	0.357	6.719	0.000	1.699 ~ 3.098
Discussion	1.982945	0.1803	0.086	2.104	0.035	0.012 ~ 0.348
Peer-reviewed Assignment	3.380423	0.8951	0.124	7.196	0.000	0.651 ~ 1.139

4 INSIGHTS FROM MACHINE LEARNING MODELS

The Random Forest model slightly outperformed XGBoost, achieving 0.794 accuracy and 0.829 ROC–AUC, versus XGBoost's 0.785 accuracy and 0.822 ROC–AUC. Although both models performed strongly, I acknowledge that they reflect correlations rather than causation. Notably, late-stage quiz completion emerged as a key predictor (see Table 2), challenging the typical focus on early engagement. This underscores the sustained importance of assessments throughout each course phase, indicating that ongoing, well-timed quizzes can help maintain commitment and enhance completion rates.

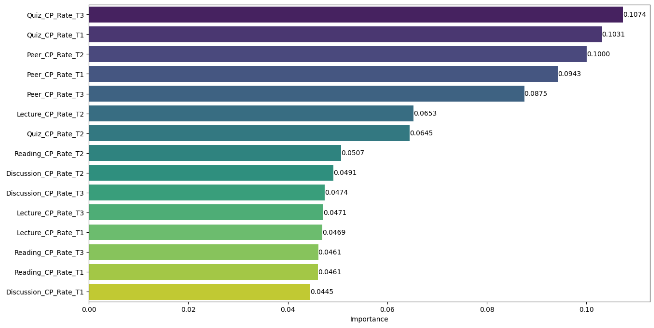


Figure 2: Feature Importance from Random Forest Model

5 SEQUENTIAL ENGAGEMENT PATTERNS

Using sequential pattern mining (see Figure 2), consistent participation across lectures, quizzes, and peer-reviewed assignments emerged as the strongest indicator of program completion. The Complete Engagement Maintenance Pattern was most predictive, underscoring the importance of sustained involvement over the entire program. Although these patterns correlate strongly with success, they do not prove causation. This suggests that course designs should integrate continuous assessments, embed collaborative tasks, and provide adaptive pathways to accommodate diverse learner needs.

Table 2: Sequential Patterns of Engagement

Pattern		N
Complete Engagement Maintenance Pattern	L1.0 → Q1.0 → P1.0 → L1.0 → Q1.0 → P1.0 → L1.0 → Q1.0 → P1.0	469
Interactive and Reflective Activities Emphasis Pattern	Q1.0 → D1.0 → P1.0 → Q1.0 → D1.0 → P1.0 → Q1.0 → D1.0 → P1.0	323
Complete Reading and Evaluation Engagement Pattern	R1.0 → Q1.0 → P1.0 → R1.0 → Q1.0 → P1.0 → R1.0 → Q1.0 → P1.0	300
Partial Engagement Allowance Pattern	R0.96 → Q1.0 → P1.0 → R0.96 → Q1.0 → P1.0 → R0.96 → Q1.0 → P1.0	232

REFERENCES

Eriksson, T., Adawi, T., & Stöhr, C. (2017). "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education, 29*(1), 133-146. https://doi.org/10.1007/s12528-016-9127-8

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research, 88*(1), 43-86. https://doi.org/10.3102/0034654317740335

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science, 363*(6423), 130-131. https://doi.org/10.1126/science.aav7958

The Rise and Fall of Conversations: Tracing Discussion Engagement Across MOOCs with Latent Growth Modeling

Ji Hyun Yu
University of North Texas
jihyun.yu@unt.edu

ABSTRACT: This study examines how discussion engagement (posts, replies, votes) evolves across a six-MOOC professional certification program using Latent Growth Modeling (LGM). Analysis of 8,808 learners reveals a declining trend in posting and replying, with consistently low voting activity. Significant predictors include prior MOOC experience, instructor presence, workload, and demographic variables such as age, race/ethnicity, and education. Findings indicate that structured early interventions, tailored workload management, and strategic instructor involvement can sustain discussion engagement across multiple courses. This work also addresses the use of demographic features in engagement prediction, recognizing potential biases while noting the importance of inclusive course design.

Keywords: MOOC Discussion, Latent Growth Modeling, Engagement Trajectories

1 INTRODUCTION

MOOCs offer scalable learning opportunities, but sustaining engagement, especially in discussion forums, remains challenging. Engagement decay, where participation declines after the initial weeks, is a persistent issue (Evans et al., 2016). While prior studies explore engagement in single courses, little is known about how learner behavior changes over multi-course programs (Ayer et al., 2018). This study addresses this gap by tracking how posts, replies, and votes evolve over six MOOCs using LGM to model engagement trajectories, offering insights into the impact of demographic and participation factors.

2 RESEARCH CONTEXT

This study examined a six-MOOC certification program delivered by a U.S. university over six months. Learners completed two eight-week courses every two months, with activities including lectures, assessments, and discussion forums. Engagement data were collected from 8,808 learners across three phases: Time 1 (months 1–2), Time 2 (months 3–4), and Time 3 (months 5–6). The engagement metrics included the number of posts, replies, and votes. To account for potential biases, missing data were addressed using full-information maximum likelihood estimation. Predictors such as gender, age, education level, enrollment motivation, prior MOOC experience, workload, instructor presence, and course pacing were also analyzed. The data were anonymized, cleaned, and prepared for LGM analysis.

3 RESULTS

3.1 Engagement Patterns Across Time

Engagement followed a declining trend, with the highest activity observed in Time 1 (months 1–2). Posting and replying decreased significantly by Time 2 (months 3–4), while voting activity remained consistently low throughout the program. A subset of learners maintained engagement across all phases, revealing a diversity in participation patterns.

Latent Growth Modeling (LGM) allows for the estimation of both individual-level variation and overall trends in engagement over time. In this study, an unconditional model was first tested, followed by a conditional model incorporating predictors (demographics, motivation, prior experience). The extended model refined these predictions by incorporating interaction effects.

Table 1: Model Fit for Unconditional, Conditional, and Extended Latent Growth Models

Measure	Unconditional Model	Conditional Model	Extended Model
χ^2	9515.728	9252.219	9543.094
Degree of freedom	18	27	42
p-value	0.000	0.000	0.000
CFI	0.872	0.877	0.874
TLI	0.744	0.671	0.648
RMSEA	0.245	0.197	0.160
SRMR	0.151	0.103	0.078
AIC	80584.664	79825.037	79742.451
BIC	80775.916	80207.541	80337.458

The unconditional model shows an initial fit to the data, while the conditional model slightly improves fit by adding predictors. The extended model, with interaction terms, provides a more nuanced view of engagement but has a slightly higher chi-square value. RMSEA and SRMR values indicate moderate fit across models. These fit indices suggest that the conditional model offers a slightly better explanatory power than the unconditional model.

Learners with prior MOOC experience and higher education levels showed stronger growth in discussion engagement over time. Instructor presence played a positive role in sustaining participation, while high workloads suppressed engagement early on. Results also suggest that the interaction between workload and instructor presence moderates engagement trends, particularly in later phases of the courses. Younger learners tended to increase participation in later phases, while older learners started with high initial engagement but declined over time.

4 DISCUSSION

The results confirm a common “rise-and-fall” arc, particularly for posting and replying, supporting prior observations of early engagement decay (Evans et al., 2016). Nevertheless, a “superposter” subgroup contributed substantially to forums across all phases, underscoring the outsized role certain learners play in driving discussions. The LGM results showed that demographic factors and prior MOOC experience explain some variance in engagement over time, with the Extended Model revealing interactions that clarify how instructor presence or workload can mitigate or amplify these effects.

These findings also indicate that learners who post actively at the start of the course tend to maintain or even increase their participation, reinforcing the importance of early engagement. Structured introductory activities can help channel this initial momentum, preventing rapid disengagement. However, learners who focus heavily on posting are less likely to ramp up their replying and voting, suggesting a preference for content creation over interaction. This pattern highlights the need for diverse participation incentives, rather than relying solely on traditional posts.

To encourage a more balanced engagement, MOOC platforms can integrate multi-dimensional discussion tasks that require posting, replying, and voting, supported by gamification elements such as badges. Empirical evidence from similar studies supports the idea that gamification can foster deeper involvement. Linking these engagement forms to peer recognition or feedback could elevate the perceived value of replying and voting. The decline in voting activity over time points to a need for assignments that tie voting to genuine learning experiences, for example through peer assessment tasks.

Although demographic predictors highlight meaningful differences, their use demands care regarding fairness (Baker & Hawn, 2021; Kizilcec & Lee, 2020). Age, race/ethnicity, gender, and education were included in line with prior work in educational data mining, but with an awareness of potential algorithmic bias. Non-White learners showed high initial participation but faced challenges in sustaining it, emphasizing the importance of inclusive course design. Older learners, while active early, experienced sharper declines, suggesting that more structured support may be needed to maintain their initial momentum. By contrast, younger participants might benefit from early encouragement to post. Education level was positively linked to prolonged engagement, indicating that highly educated learners may be more accustomed to online discussion. Workload management also proved critical, as heavy workloads suppressed early engagement, though participation rose once learners adjusted. Phased instructor engagement, involving high visibility and guided activities later in the course, has the potential to re-energize discussions that may otherwise lose steam after the initial surge.

REFERENCES

Ayer, N., Sukhathankar, H., Deshmukh, U., & Sahasrabudhe, S. (2018). Impact of learner-centric discussion forums on learner engagement in skill development MOOC. In 2018 IEEE Tenth International Conference on Technology for Education (T4E) (pp. 69–72). IEEE.

Baker, R. S., & Hawn, A. (2021). Algorithmic fairness in education. Journal of Learning Analytics, 8(2), 1–17.

Evans, B. J., Baker, R. B., & Dee, T. S. (2016). Persistence patterns in massive open online courses (MOOCs). The Journal of Higher Education, 87(2), 206–242.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. Educational Research Review, 12, 45–58.

Kizilcec, R. F., & Lee, H. (2020). Algorithmic fairness in education. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 43–47).

Learning Analytics of Assessors' Pairing Choices and Focus of GenAI-facilitated Feedback in Peer Assessment

Author: Zexuan Chen
Institution: School of Foreign Studies, Southern Medical University (China)
Email: SerlinaChen@163.com

Author: Bart Rienties
Institution: Institute of Educational Technology, The Open University (UK)
Email: bart.rienties@open.ac.uk

Author: Simon Cross
Institution: Institute of Educational Technology, The Open University (UK)
Email: simon.j.cross@open.ac.uk

ABSTRACT: To examine assessors' choices of peer work across different proficiency levels and their feedback focus when supported by GenAI in peer assessment, we invited 179 students to participate in a one-semester experiment during which they completed three peer assessment tasks using a customized system—PeerGrader. Preliminary findings from learning analytics suggest that (1) when given the autonomy to choose proficiency levels, assessors were influenced not only by their own proficiency levels, in line with Homophily Theory, but also engaged with materials that facilitated their learning, consistent with ZPD principles; and (2) when using GenAI, assessors' feedback focus may broaden to place greater emphasis on discourse-level aspects compared to non-GenAI-assisted feedback. Future research will explore the design and effects of larger-scale GenAI-assisted peer assessment.

Keywords: Learning Analytics, peer assessment, pairing choice, GenAI, feedback focus

1 INTRODUCTION

It is widely recognized within the learning analytics community that formative assessment is crucial for enhancing learning processes and outcomes [3]. However, there has been relatively limited focus on the learning analytics of peer assessment [2], where students act as assessors and provide feedback on their peers' work. To examine assessors' learning behaviors and cognitive processing, we have been conducting design-based research on student assessors' engagement in the feedback-giving process. Our ongoing research questions are: 1) What choices do assessors make when given the autonomy to select peer writings across different proficiency levels? 2) What are assessors' key focus areas when collaborating with GenAI in formulating feedback?

2 METHODOLOGY

This Design-based research was conducted in a blended learning course that included peer assessments of three EFL writings (English as a Foreign Language), carried out during the Spring semester of 2024 at monthly intervals. A total of 179 first-year undergraduates participated voluntarily, categorized into high (H), medium (M), and low (L) proficiency groups based on their grade distribution. Among them, six assessors also tested the use of GenAI to assist in formulating and providing feedback.

A customized peer assessment system called "PeerGrader" was used to conduct peer assessments and collect data. PeerGrader provided all H, M, and L assessors with the autonomy to select peer writings across three proficiency levels and recorded their selections. After collation, we obtained 1,347 peer writing retrievals contributed by the 179 assessors and 66 qualitative feedback entries written by the six GenAI-assisted assessors, which were manually coded by two researchers. The data were then imported into SankeyMATIC and Excel for analysis and visualization.

3 RESULT AND DISCUSSION

3.1 Assessors' Pairing Choices

Figure 1 depicts H, M, L assessors' choices of H, M, L writings across three peer assessment tasks. Each vertical node represents the total number of retrievals of writings from the respective proficiency groups (H, M, L) within a single peer assessment, with arrows indicating the sequential progression of the three peer assessment tasks from left to right. The thickness of the arrows provides a visual comparison of the number of writings chosen and retrieved by the assessors in every task.

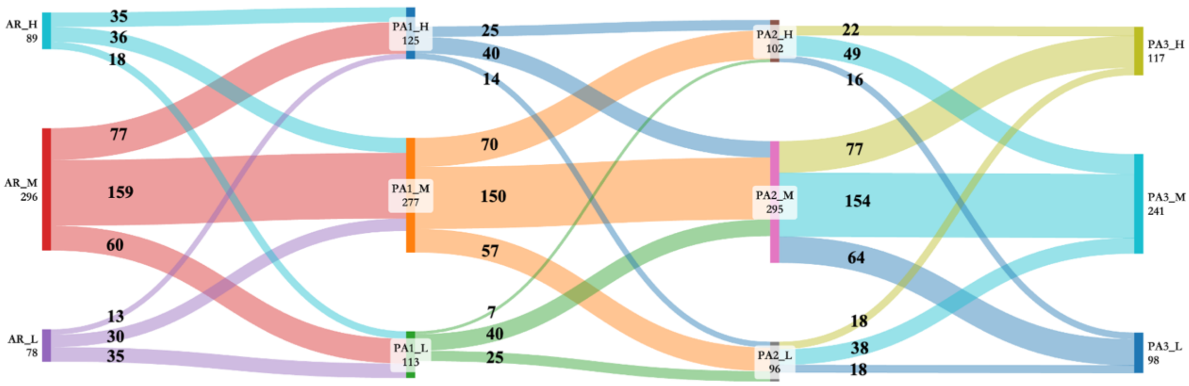


Figure 1: Assessors' Retrieval of Writings from Different Levels Across the Three Tasks

Results from the peer assessment analytics indicate that H, M, and L assessors tend to select peer writings that match their own proficiency levels, which aligns with Homophily Theory [1]. Additionally, the results showed that more assessors consistently chose to review more M writings; meanwhile, H and L assessors tended to review each other's work less frequently. This finding is supported by Vygotsky's ZPD theory [4], which suggests that assessors engage more effectively with peer work that is suitably challenging based on their capabilities and expertise.

3.2 GenAI-assisted Assessors' Feedback Focus

Figure 2 presents the feedback focus identified by the six assessors across discourse, sentence, and lexical aspects in three peer assessment tasks. Note that the six GenAI-assisted assessors had the independence to decide whether or not to incorporate GenAI into any of the three peer assessment tasks. As a result, three types of assessors were identified: the "Self-sufficient Master" (S1, S2) utilized GenAI to assist feedback giving for the first and second tasks but provided feedback without AI intervention for the remaining task; the "Cautious Adopter" (S3, S4) gave GenAI-facilitated feedback exclusively for the third task; while the "Sustained User" (S5, S6) employed GenAI to help them provide feedback throughout the entire semester.

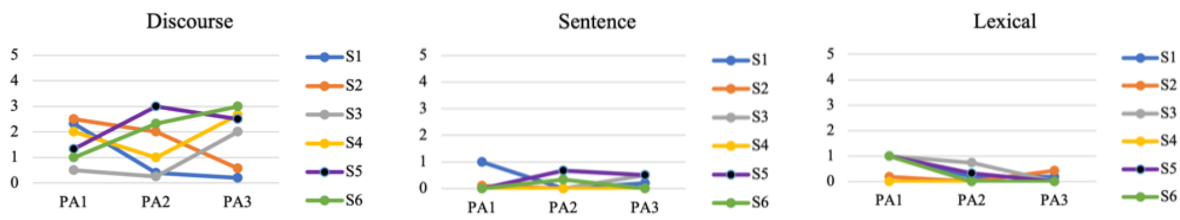


Figure 2: The Six Assessors' Feedback Focus in the Three Peer Assessment Tasks

Figure 2 shows the average number of feedback focuses on discourse, sentence, and lexical aspects for each of the six GenAI-assisted participants across the three peer assessment tasks (total counts of feedback focuses/number of feedback submissions). Further analysis revealed that the Self-sufficient Masters' quantity of feedback across the three feedback focus aspects gradually decreased as they transitioned from a GenAI-facilitated to a non-GenAI-facilitated condition. Meanwhile, the Cautious Adopters demonstrated a tendency towards heightened focus on commenting at the discourse and sentence aspects as they shifted from a non-GenAI-facilitated to a GenAI-facilitated condition. In addition, the Sustained Users of GenAI exhibited a consistent increase in their focus on discourse, accompanied by a decrease in their focus on sentences and lexical aspects when providing feedback.

4 MAIN TAKEAWAYS

- By enabling flexible and self-selected peer writings across different proficiency levels, learning analytics in peer assessment harnesses the strengths of Homophily while capitalizing on the developmental benefits of interactions within the ZPD.
- The GenAI assistance might be beneficial in enabling assessors to broaden their focus, encompassing discourse, sentence, and lexical aspects, particularly at the discourse aspect, which relates to the overall structure, coherence, and logic of the text.

REFERENCE

[1] Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. Freedom and control in modern society, 18(1), 18-66.

[2] Misiejuk, K., & Wasson, B. (2023). Learning Analytics for Peer Assessment: A Scoping Review. The Power of Peer Learning: Fostering Students' Learning Processes and Outcomes, 25-46.

[3] Rienties, B., Tempelaar, D., Nguyen, Q., & Rogaten, J. (2024). The use of data analytics to support the development of assessment practices in higher education. In Research Handbook on Innovations in Assessment and Feedback in Higher Education (pp. 194-209). Edward Elgar Publishing.

[4] Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (Vol. 86). Harvard university press.

An Investigation of the Relationship between Open-Ended Questionnaires and Lectures

Masaki Koike
Chubu University
masa1357@mprg.cs.chubu.ac.jp

Tsubasa Hirakawa
Chubu University
hirakawa@mprg.cs.chubu.ac.jp

Takayoshi Yamashita
Chubu University
takayoshi@isc.chubu.ac.jp

Hironobu Fujiyoshi
Chubu University
fujiyoshi@isc.chubu.ac.jp

ABSTRACT: With the increasing digitalization of educational environments, predicting student comprehension through learning log data has become a prominent area of study. However, such methods often lack the capacity to provide detailed insights, such as pinpointing which lecture parts students found unclear. To address this limitation, we analyze open-ended questionnaires that capture students' subjective understanding. By examining questionnaire responses over time using distance metrics and correlating these with student grades, we aim to predict students' lecture comprehension. We use a decision tree model to improve prediction accuracy and provide explainable insights. Our findings suggest that incorporating temporal changes in questionnaire responses significantly enhances prediction performance.

Keywords: Students' Performance Prediction, Open-Ended Questionnaires, Decision Tree

1 INTRODUCTION

As educational environments continue to digitize, there has been growing interest in using machine learning to predict student comprehension and performance. Early detection of students struggling with lectures can enable timely and targeted learning support (Leelaluk et al., 2024; Namoun et al., 2021). However, methods relying solely on learning log data can only capture students' interactions with digital materials, without providing context for why students may not be grasping certain concepts. In contrast, open-ended lecture questionnaires capture individual perspectives and subjective interpretations of the lecture content. This study aims to combine the subjective insights from open-ended questionnaires with machine learning models. By doing so, we intend to improve the prediction of student comprehension and provide more actionable insights for educators.

2 PROPOSED METHOD

This study introduces a method to analyze temporal changes in open-ended questionnaire responses and predict students' comprehension of lectures using a decision tree model.

2.1 Calculation of Semantic Changes Using Word2vec

To capture the temporal changes in questionnaire responses[冗長?], we first convert the text into embeddings[get embeddings…] using a Word2vec model trained with the Skip-gram algorithm. The semantic changes between responses from different lectures are then quantified by calculating Euclidean and cosine distances between the sentence embeddings. These distance metrics reflect how students' understanding and responses evolve throughout the course.

2.2 Predicting Students' Lecture Comprehension Using a Decision Tree Model

For the prediction task, we employ the Light Gradient Boosting Machine (LightGBM), a decision tree model, to train on the calculated distance metrics. LightGBM is chosen due to its efficiency in learning and its ability to provide interpretable decision rationales[削減可能]. The students' grades are used as a proxy for their comprehension levels, and we use the temporal changes in the questionnaire responses to train and predict comprehension outcomes[重複表現]. By utilizing SHapley Additive exPlanations (SHAP), we can identify the most influential features contributing to the model's predictions, enabling us to provide insights into which factors are most correlated with students' comprehension of lecture content.

3 EXPERIMENTS

We visualize the semantic changes obtained for each lecture and investigate their relationship with grades. Additionally, we analyze the accuracy of grade prediction using these semantic distance measures.

3.1 Experimental overview

The dataset used in this study was collected from the "Information Science" course at a Japanese University, which spans 15 weeks. After each lecture, students were asked a reflective question: "Please explain today's content in your own words." The students were graded on an A-F scale, with grades D and C combined due to the small sample size of D, which accounted for only 5% of the total. The dataset includes responses from 377 students, split into 80% for training and 20% for evaluation.

3.2 Results

Figure 1 displays the median of temporal cosine distances of questionnaire responses categorized by students' grades. In this analysis, a distance of zero signifies identical responses across lectures. From Figure 1, it is evident that students with an F grade exhibited little change in their responses in over 70% of the lectures, while students with higher grades demonstrated larger cosine distances, indicating more variability in their responses. Next, we assess the effectiveness by incorporating the temporal changes in the questionnaire responses into a decision tree-based prediction model. As a baseline, we trained a LightGBM model on the responses for each of the 15 lectures, averaging the

Table 1: Quantitative evaluation results for the questionnaire dataset.
Baseline : 15-model ensemble
Ours : Model using temporal changes

	Baseline	Ours
Accuracy [%]	60.41	**68.99**
F1 score [%]	53.15	**68.84**

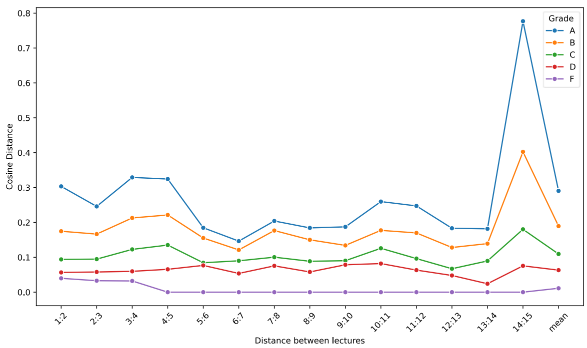
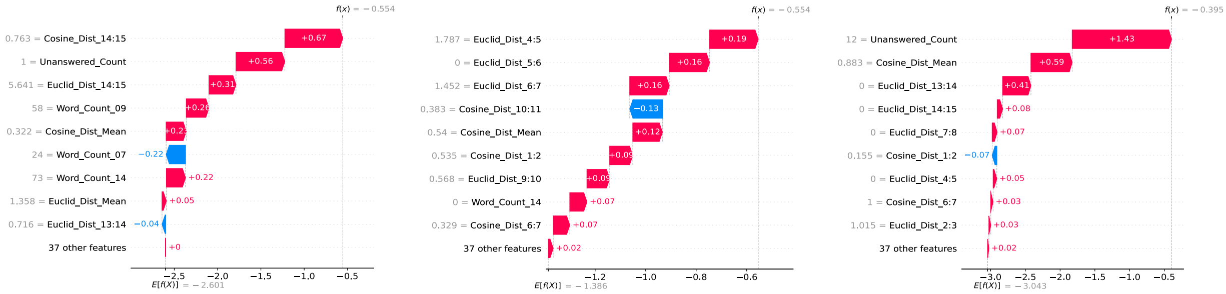


Figure 1: Cosine distance between lectures

logits to obtain final predicted probabilities. We used 5-fold cross-validation to compare the accuracy and F1 score between the baseline and our proposed method. Additionally, we applied class balancing by assigning weights inverse to the number of students in each grade category.



(a) Grade A (b) Grade C (c) Grade F

Figure 2: Model's reasoning of student with each grade

Table 1 presents the grade prediction accuracies for each method. The results show that the accuracy and F1 score of the proposed method improved by 8.58 points and 15.69 points, compared to the baseline. We also analyzed which features were most correlated with grades by examining feature importance in the proposed model. Figure 2 illustrates the interpretation of the model's reasoning for a randomly selected student's questionnaire responses for each grade. The results reveal that, for students with grades A and F, the model places significant weight on the number of unanswered and the average cosine distance, reflecting the semantic changes in responses. Additionally, for students with grade C, the model highlights the response changes between the fourth and fifth lectures. In this case, the student’s responses for these two lectures were identical, suggesting potential issues with the student’s attitude during the fifth lecture. These findings suggest that the model considers students’ attitudes toward lectures, such as reusing previous responses, when making its predictions.

4 CONCLUSION

In this study, we evaluated the effectiveness of a machine learning model that incorporates semantic temporal changes in open-ended questionnaires to predict student comprehension. Our findings reveal a clear correlation between these semantic changes and students’ grades. We achieved significant improvements over the baseline by utilizing a LightGBM model trained on temporal change distances, with a 15.69 point increase in the F1 score. We also demonstrated through visualizations of the model’s decision rationale that it considers various student behaviors, such as response repetition and semantic shifts in answers across lectures. This insight offers valuable interpretability, allowing us to better understand the factors influencing the model’s predictions. Our future work will focus on developing models that integrate deeper text comprehension to enhance prediction performance.

REFERENCES

Leelaluk, S., Tang, C., Minematsu, T., Taniguchi, Y., Okubo, F., Yamashita, T., & Shimada, A. (2024). Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities. IEEE Access, 12

Namoun, A., & Alshanqiti, A. (2021). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. Applied Sciences, 11(1), 237.

What Are the Most Wanted Features in a Student-Facing Learning Analytics Dashboard? Results from a User-Centric Study Combining Needs Assessment and Iterative Design Cycles

Ba Vinh Le, Binh Nguyen, Zijue Fang, Ju Petal, Ville Kivimäki
Aalto University
vinh.v.le@aalto.fi, binh.t.nguyen@aalto.fi, zijue.fang@aalto.fi, ju.petal@aalto.fi, ville.kivimaki@aalto.fi

ABSTRACT: This study aimed to develop a learning dashboard that places students at the center of the design process by exploring their needs for dashboard features and testing the understandability of the dashboard prototype. Employing an iterative and user-centric approach, we consulted students—the end-users of the learning analytics dashboard—in two phases. The first phase included a survey with 123 responses to gather student perceptions of the existing dashboard and identify desired features, categorized into must-have, potential, and least-wanted, to guide subsequent prototyping efforts. In the second phase, three prototypes were developed, and 19 individual interviews were conducted. The final prototype had an average System Usability Score (SUS) of 83.57, and such a positive result underscores the effectiveness of our design process, highlighting the importance of involving students as core stakeholders in creating relevant and understandable digital learning tools.

Keywords: Learning Analytics, Dashboard, user-centric design, learning design

1 INTRODUCTION

Demand for automated personalized feedback is increasing due to the larger number of students and limited university resources (Kivimäki, 2024). Dashboards, visual tools that present learning insights and individualize feedback, can be considered a viable solution to this problem, empowering students to proactively manage academic progress. However, it is unclear how to build the dashboard, especially to comprehensively meet the needs of a broad profile of students in the multidisciplinary environment across the six schools at the case university. Therefore, this study aims to address two gaps. Since the case university does not know which theoretical-based features are seen as relevant by educational stakeholders, following the recommendation of Verbert et al. (2020), the first purpose of the study is to understand students' needs in a learning analytic dashboard. Secondly, as the dashboard is a visual-based learning analytics system, there is a need to evaluate the understanding and usability of data visualizations, involving educational stakeholders in validating the dashboard functionality and effectiveness against its intended outcome. By addressing these gaps, this study aims to give an active voice to educational stakeholders in the design process, placing students at the center of each identified gap to ensure that the learning dashboard can genuinely add value to their learning experience.

2 METHOD

This research employed an iterative, user-centric approach, consulting students—the end-users of the learning analytics dashboard—throughout development. The design, shown in Figure 1, included an initial survey followed by two cycles of prototyping and interviews and a final prototype.

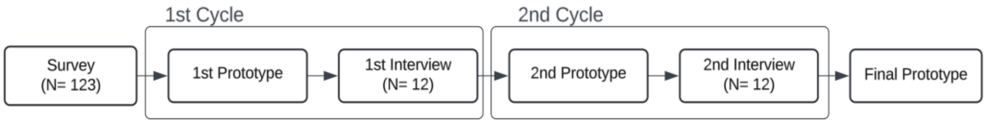


Figure 1: Research Design

To address the first research gap, a survey was conducted at the participating university to understand its students' perceptions of the existing dashboard and identify desirable and helpful features. Students rated the twenty features from the Borter et al. (2024) study, divided into four categories: Performance Prediction, Study Performance, Planning, and Resources, on a three-point scale (1 = must-have, 2 = optional, 3 = unnecessary). To target the research gap, only the feature prioritization section of the original questionnaire was used, thus focusing on first identifying the necessary features rather than refining all features, including those that might be deemed unnecessary.

The second research gap, aligning the final design with student needs, was addressed through iterative prototype testing. Prototypes were created and evaluated based on two key metrics: relevance to student needs and feature understandability. Desired features from survey data informed two development cycles, using paper prototyping for the first version and digital mock-ups for the second and final versions, each including interactive data filtering and exploration. Testing sessions were conducted through individual interviews and had three parts: identifying student challenges to assess relevance, testing the dashboard's understandability through task completion, and evaluating the overall layout and prioritization of features. In the second cycle, a System Usability Scale survey was also used to collect quantitative data, complementing the qualitative feedback from the interviews.

3 RESULTS

3.1 Survey Results

We surveyed 141 students (45.5% bachelor's, 54.5% master's) at the participating university, with 123 fully completing the survey. The findings identified five "must-have" features, each chosen as "1" by over 60% of students, and four "potential" features, each selected as "1" by over 30% and "3" by fewer than 20%, or with an average rating below 2. To validate these classifications, we conducted a T-test comparing each feature's average rating to the overall average rating across all features. A p-value of less than 0.05 indicates a statistically significant difference, suggesting that students overestimated or underestimated the feature compared to the overall average (Figure 2)

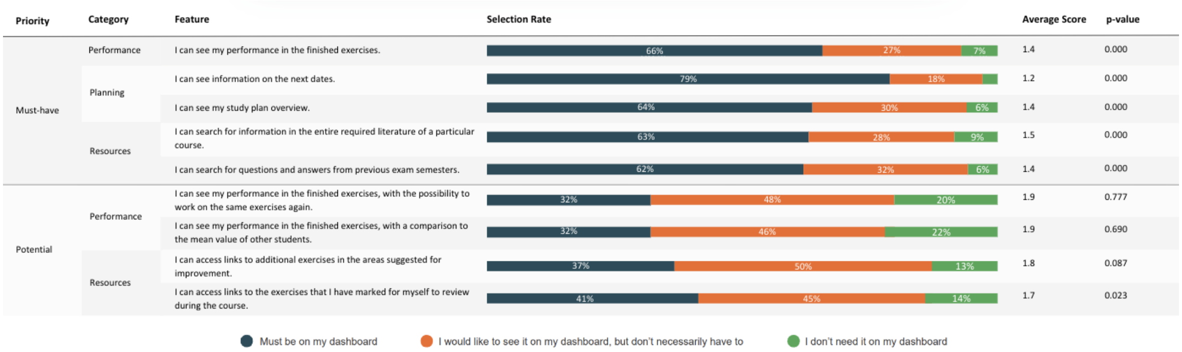
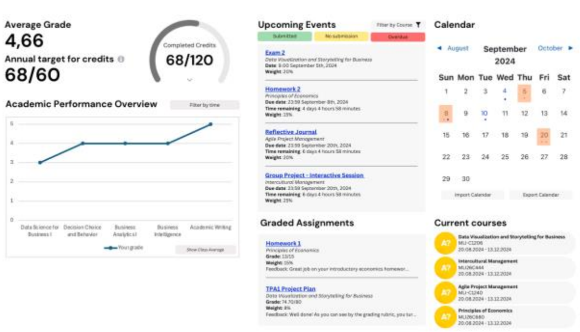


Figure 2: Categories of Features

3.2 Prototype Testing Results

The prototype underwent two testing phases, with 12 participants in the first and 7 in the second. Feedback refined both functionality and design. The final prototype includes an overall dashboard for course overviews (grades, deadlines, and plans) and a course-specific dashboard for tracking progress, assignments, and resources (See Figure 3 and Figure 4 below). Usability was assessed with the System Usability Scale (SUS), yielding an average score of 83.57 ("excellent"). However, scores ranged from 56.66 to 98.34, indicating some areas need refinement for a more consistent user experience.



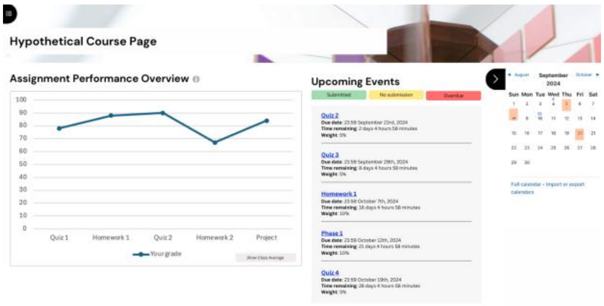


Figure 3: Overall Dashboard **Figure 4: Course-Specific Dashboard**

4 DISCUSSIONS

Extant literature highlights a lack of needs investigation and the failure to center end-users in learning analytics, resulting in dashboards that may not meet students' needs. Our study embraced a human-centric approach by actively involving educational stakeholders to uncover and validate their specific needs through iterative visual design, creating a dashboard for and with students to add genuine value to their learning experiences. This approach is validated by the positive SUS score, demonstrating the effectiveness of our design process for a diverse student profile across six multidisciplinary schools. Nevertheless, this study only represents the early phases of a human-centred design process in understanding, creating, and delivering, rather than an entire co-design process with an additional supporting layer (Prieto-Alvarez et al., 2018), and further studies should be conducted post-deployment to assess the learning dashboard's effectiveness in real-world educational settings, engaging educational stakeholders in a continuous co-design process.

REFERENCES

Borter, N., Bögli, L., & Troche, S. (2024) Students Dashboard Preferences in Blended Learning with Continuous Formative Assessments. In *LAK24 Conference Proceedings*, 222-224.

Kivimäki, V. (2024). Designing Student Agency in Higher Education: The Cases of Individual Study Planning and a Structured Learning Diary. University of Helsinki.

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. D. (2018). Co-designing learning analytics tools with learners. In *Learning analytics in the classroom* (pp. 93-110). Routledge.

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 *Proceedings of the tenth international conference on learning analytics & knowledge,* 35-40.

Supporting Student-Centered Learning with Flexible Learning Trajectories and Open Learner Models

Pankaj Chejara, Kairit Tammets, Mart Laanpere, Annika Volt, Priit Tammets, Pjotr Savitski
Center of Education Technology, Tallinn University, Tallinn, Estonia
pankajch@tlu.ee, kairit@tlu.ee, martl@tlu.ee, annika.volt@tlu.ee, tammets@tlu.ee, gnum@tlu.ee

ABSTRACT: Flexible learning caters to each individual's learning needs, offering a solution to provide personalized learning while preserving the student's agency. However, effective utilization of this approach requires students to have some basic understanding of dependencies among learning contents, and the ability to track and assess their learning (i.e., metacognitive skill). Our research aims to address these challenges with a novel approach built upon an expert-designed domain model, flexible instructional trajectories, and open learner models (OLMs) to support students in flexible learning. We present our ongoing research work on leveraging log data to construct OLMs to provide meaningful insights into students' learning processes with the ultimate goal to foster metacognitive development and enable informed decision-making.

Keywords: Open Learner Model, Bayesian Modeling, Learning Path, Learning Trajectories

1 INTRODUCTION

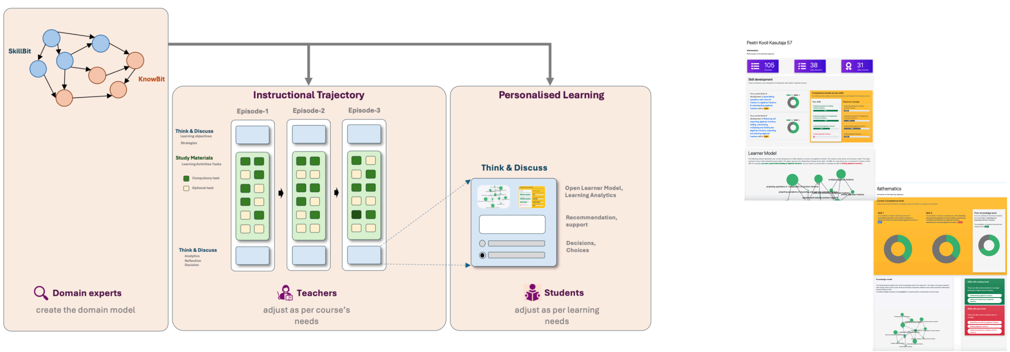
Flexible learning, which enables learners to progress according to their individual needs, is a key priority in the UNESCO 2030 Agenda for Education (UNESCO, 2022). The growing demand for flexibility in education is also emphasized by the European Association for University Trends 2018 report (Geabel et al., 2018). When discussing flexible learning, three core dimensions are often highlighted: time, place, and mode of learning (Hammersley et al., 2013). These dimensions enable students to have a personalized learning experience with greater autonomy. However, despite its importance, implementing flexible learning is not without challenges. Its success depends on students having a foundational understanding of learning content structure, as well as the ability to track and assess their learning progress.

In this paper, we present our research work on designing and developing flexible learning environments through the integration of pedagogical knowledge and learning analytics. Specifically, we developed dashboards for students and teachers that combine expert-designed domain models and Open Learner Models (OLMs) –a machine representation of student's learning– to provide insights into students' learning. This approach empowers learners by delivering data-driven feedback, enabling informed decision-making about their learning, while simultaneously fostering metacognitive development.

2 FLEXIBLE LEARNING WITH OPEN LEARNER MODELS

The research work is part of an Estonian national project with a focus on student-centered learning in primary schools. We have refined the concept of flexible learning by dividing it into three distinct but interconnected components (Figure 1.a): (1) Domain model, which defines the learning outcomes and related knowbits and skillbits; (2) the instructional trajectory, which is part of the instructional design and (3) the learning path, which is related to learning analytics. The instructional trajectory is a forward-looking plan designed by a teacher or instructional designer, grounded in a domain model (typically based on the national curriculum). It includes predefined learning outcomes, instructional episodes, tasks, hints, and additional resources (Volt et al, 2024). Drawing from the instructional design principles of van Merriënboer & Kirschner (2007), these trajectories incorporate flexibility through optional tasks, choices in task type and complexity, and personalized pacing, supported by metacognitive prompts to encourage reflection. In contrast, the learning path represents the learner’s actual journey - a dynamic, digital record of their interactions with content, teachers, and peers. This path captures the learner’s actions, choices, and outcomes in a machine-readable format, enabling advanced learning analytics to provide real-time feedback and support.

We implemented a Drupal-based learning platform[1] for flexible algebra learning for Estonian 9th-grade students, which allows teachers/instructors to create instructional trajectories using H5P elements. A domain model for Algebra was prepared by experts in mathematics didactics following the Estonian national curriculum. This model provided the basis for creating instructional trajectories which consisted of several episodes each covering one specific topic. These episodes were further divided into several tasks of different types: reading, watching videos, assessment, and problem-solving.



a. Flexible Learning b. Student & Teacher dashboard

Figure 1. Flexible learning and generated dashboard[2] with open learner model

Students' interactions with these tasks were recorded using xAPI statements. We extracted various features such as number of attempts, usage of hints, scores, etc. These log features were explored using Bayesian modeling to compute an estimate of students’ mastering specific skills (e.g., dividing common fractions). These estimates were computed using *the Expectation Maximization* algorithm

[1] VARA: https://vara.h5p.ee/

[2] Dashboard prototype: http://eduflex.blog/en/

using extracted features. The estimates were then used to build a *dashboard,* following best practices from the literature (e.g., visualizing students' progress in a way easy to interpret, and along with classroom averages). We visualized OLMs in the form of a Bayesian network where nodes represent skills and knowledge, and edges represent relationships between them according to the domain model. The network shows students' current level of knowledge for each episode in learning trajectories. Figure 1.b shows student and teacher versions of the dashboard built upon developed OLMs.

3 CONCLUSION & FUTURE WORK

In this paper, we presented a novel flexible learning approach integrating three key elements - domain models, flexible learning trajectories, and students' Open Learner Models to create a more adaptive learning experience. Together, these elements form a cohesive system where instructional design, learning analytics, and learner autonomy converge to foster personalized, flexible, and effective learning environments. Our flexible approach creates a dynamic learning environment that is responsive to the unique needs and trajectories of individual learners. In our ongoing research, our next step is to analyze differences in how students navigate through flexible instructional trajectories. Using temporal Learning Analytics (LA), alongside student learning outcomes and self-reported measures, we will investigate how these navigation patterns are linked to learning gains, perceived effectiveness, and metacognitive strategies. Feedback from the LA community will be essential in guiding our research and selecting the most suitable methods and approaches for analyzing these relationships.

ACKNOWLEDGEMENT

This research was supported by the Estonian Ministry of Education and Research under the research project ÕLHAR1: "Flexible Learning Paths for Supporting Learner-Centered Learning in Schools", and by the Estonian Research Council's Personal Research Grant (PRG) under grant number PRG1634.

REFERENCES

1. Gaebel, M., Zhang, T., Bunescu, L., & Stoeber, H. (2018). *Learning and teaching in the European higher education area*. European University Association.
2. Hammersley, A., Tallantyre, F., Le Cornu, A. (2013). Flexible learning: A practical guide for academic staff. The Higher Education Academy.
3. Kirschner, P., & Van Merrienboer, J. J. (2008). Ten steps to complex learning: A new approach to instruction and instructional design. In 21st century education: A reference handbook (pp. 244-253). SAGE Publications Ltd.
4. UNESCO. (2022). Beyond Limits. New Ways to Reinvent Higher Education. In *Working document for the World Higher Education Conference*.
5. van Merriënboer, J. J. G., Kirschner, P. A. (2007). Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design. Lawrence Erlbaum Associates.
6. Volt, A., Laanpere, M., & Kurvits, J. (2024). Supporting flexible learning paths with interactive learning resources in mathematics: lessons learned. Educational Media International, 1-16.

Flexible learning analytics system using APEX on Oracle Cloud facilitates running complex undergraduate medical education

Dong-mi Yoo, Wha Sun Kang, Sun Kim, Chang Jin Choi, Su Young Kim
The Catholic University of Korea, College of Medicine
suyoung.dr@gmail.com

ABSTRACT: This poster illustrates the development of a new learning analytics system using Application Express (APEX) on Oracle Cloud for undergraduate medical education. Due to the complexity of the curriculum, providing a comprehensive overview and monitoring is a formidable task for stakeholders. Because the current learning analytics system is limited, we built a new learning analytics system using APEX on Oracle Cloud. With the new learning analytics system, application building time was greatly reduced. Even one-day distribution of web applications was possible. Because APEX is low-code development software and does not require coding skills, even health professionals without a programming background were able to develop and distribute learning analytics web applications. Because various tutors and learners request personalized learning analytics presentations, one predefined analytics dashboard is not enough to meet their needs. Thanks to the simplicity of development in APEX, tailored learning analytics dashboards for individual learners and tutors were possible.

Keywords: Platform as a Service, Learning analytics system, Low code software, Oracle APEX

1 BACKGROUND

Undergraduate medical education is a curriculum for obtaining a medical degree and license for physicians. Due to diverse stakeholders and an ever-changing education environment, running the curriculum is not easy. To facilitate curriculum management and assist learners and tutors, we are running several systems based on information technology (Shorten, 2024).

In our medical school, the curriculum is 6 years long and more than a thousand tutors and learners participate in regionally separate 8 teaching hospitals. To facilitate learning and teaching, diverse educational pedagogies are adopted. Due to this complexity, providing a comprehensive overview and monitoring is a formidable task for stakeholders.

The current learning analytics system is in a predefined and fixed format. Because it is maintained by IT professionals and developed as database-driven web pages, revising the system takes a lot of time. In addition, customized learning analytics tailored to individual users and new pedagogy is difficult.

2 METHODS

To tackle the limitations of the old system, we built a new learning analytics system.

APEX on Oracle Cloud is adopted. The low code development provides fast development and deployment (Konersmann, 2024). It also provides automatic responsive web applications and database management (Pastierik & Kvet, 2023) . The new system is linked to Google Workspace of our school by OAuth2.0 (Figure 1).

3 RESULTS

The new learning analytics system has the following characteristics:

The time required from user requests to system deployment is greatly reduced. In the old system, system amendments were based on a yearly schedule. At least one year was required to apply system amendments. Thanks to APEX, system amendments were possible in a few days. When the user request was small, even one-day development and deployment was possible in the format of responsive web applications. Low-code software development environment made it possible for even health professionals without programming skills to develop and distribute learning analytics web applications.

Since various tutors and learners participate in undergraduate medical education, their requests are diverse. One predefined analytics dashboard is not enough to meet their needs. Because web application development is simple in low-code PaaS, customized learning analytics dashboards for individual learners and tutors are produced. For example, user-triggered visualizations of real-time data were enabled (Figure 2 and 3). For example, student assessment dashboard are tailored for individual users who are interested in different monitoring, processing and visualization (Yang & Ogata, 2023).

Because the platform is provided as Platform as a Service (PaaS), IT personnels and on-premises hardware is not mandatory. The system is so stable and reliable that system down time is close to 0.

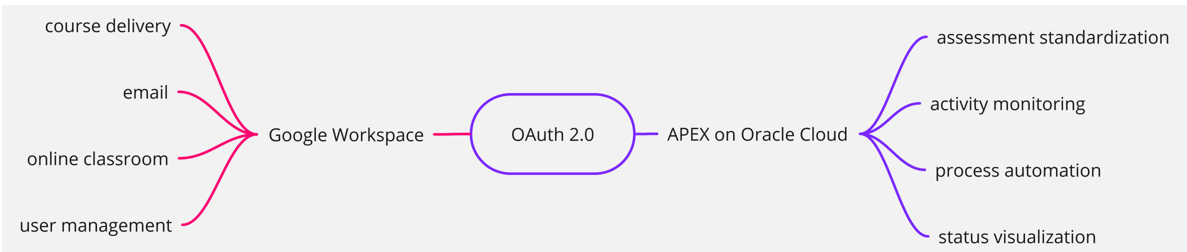


Figure 1: Integration of Google Workspace with APEX on Oracle Cloud.

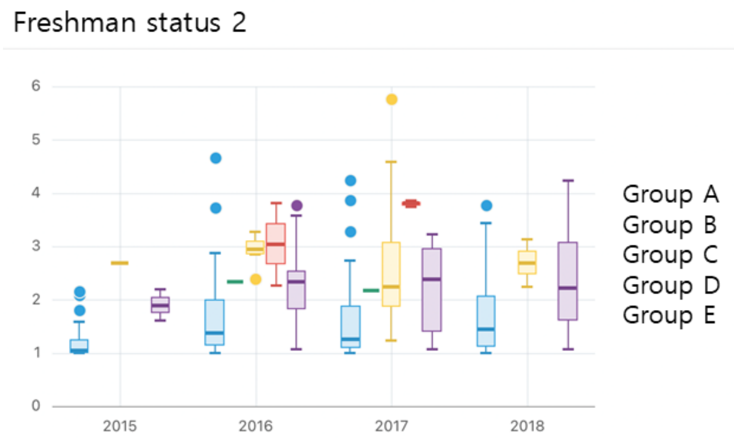


Figure 2: An example of user-triggered visualization using learning analytics web applications; temporal changes of assessment scores among freshman groups.

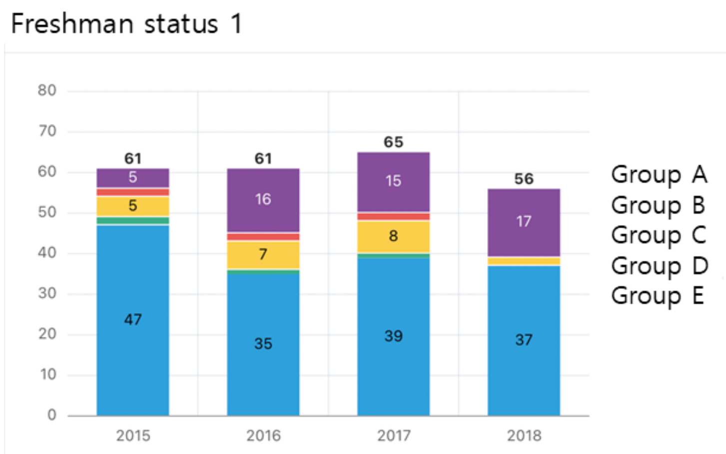


Figure 3: An example of user-triggered visualization using learning analytics web applications; temporal changes of populations among freshman groups.

REFERENCES

Konersmann, M. (2024). Challenges of Low Code PAAS Environments for Future Software Reengineering. In A. Herrmann (Ed.), *Softwaretechnik-Trends Band 44, Heft 2*: Gesellschaft für Informatik e.V.

Pastierik, I., & Kvet, M. (2023, 26-27 Oct. 2023). Exploring Oracle APEX for the University Data Analysis. 2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA),

Shorten, G. (2024). Learning analytics and the future of postgraduate medical training. *Ir J Med Sci*, *193*(5), 2607-2609. https://doi.org/10.1007/s11845-024-03702-9

Yang, C. C. Y., & Ogata, H. (2023). Personalized learning analytics intervention approach for enhancing student learning achievement and behavioral engagement in blended learning. *Education and Information Technologies*, *28*(3), 2509-2528. https://doi.org/10.1007/s10639-022-11291-2

Unlocking Math Learning Through Personalized Programs and Game-Based Progress

Linlin Li, Mingyu Feng, Chun-Wei Huang
WestEd
lli@wested.org, mfeng@wested.org, chuang@wested.org

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

Elizabeth V. Owen
Learning Data Discovery LLC
v.elizabeth.owen@gmail.com

ABSTRACT: With various levels of math knowledge in the classrooms, teachers face the challenge of identifying individual student needs and providing personalized instructions. Well-designed and evidence-based educational technology innovations offer a promising solution, enabling targeted and differentiated instruction to prepare students for success in math. In this poster, we discuss a personalized mastery-based learning program and explore the relationships between students' program usage, game-based performance, and math achievement.

Keywords: Personalized Learning, Adaptive Learning Activities, Game-based Program, Evidence-based Design

1 INTRODUCTION

Research highlights the crucial role of early math competencies - number knowledge, number relations, and number operations - in preparing for later academic success and 21st-century STEM careers (Devlin, et al., 2022). However, with only 37% of U.S. fourth grade students proficient in math (NCES, 2022), there is an urgent need to implement innovative approaches that can improve math performance. The challenge is even more acute for children from low socioeconomic backgrounds who enter school with significantly less math knowledge than their middle-class peers (Nores & Barnett, 2014). With various levels of math knowledge in the classrooms, teachers face the challenge of identifying individual student needs and providing personalized instructions (Goddard, et.al., 2015). Well-designed and evidence-based educational technology innovations offer a promising solution, enabling targeted and differentiated instruction to help teachers better prepare students for success in math. In this study, we examine the relationship among student's usage of such innovations, within-system performance, and math achievement.

2 BACKGROUND AND CONTEXT

My Math Academy (*MMA*) addresses key early math competencies: number sense, number relations, and number operations. As a personalized mastery-based learning ecosystem, *MMA* leverages current approaches in game-based learning (Plass, et al., 2016) and uses evidence-centered design (Mislevy, et al., 2003) to enable learners to master math concepts through playful experiences. Unlike programs that provide fixed learning sequences for all students, *MMA* offers adaptive learning activities and allows students to receive personalized, differentiated support. It supports students' independent practices and connects real-world experiences through game-based learning activities within story contexts and just-in-time feedback. The story context of a math problem makes concepts and operations more meaningful to students and provides students with a framework for understanding what they are expected to do, and why (Sullivan, et al., 2003). Storylines in game-based learning activities help all students, including struggling readers, gain access to the math problems, make sense of math problems, and transfer skills learned in games into the real world. The game-based learning activities also enable integrated, ongoing formative assessments to provide immediate feedback to students during the learning process, which enable ongoing feedback cycles and customized learner difficulty levels (Shute & Kim, 2014). Activities that are under the same math skill are grouped into nodes. Nodes often have a progression of activities from teaching new concepts/skills, through practice, into mastery. Nodes that share a content theme are grouped into modules. Figure 1 shows an example of how a node map is structured. Given these adaptivity and scaffolding mechanisms at multiple levels, each student has a completely personalized experience, tailored precisely to their "ready to learn" math level and learning pace.

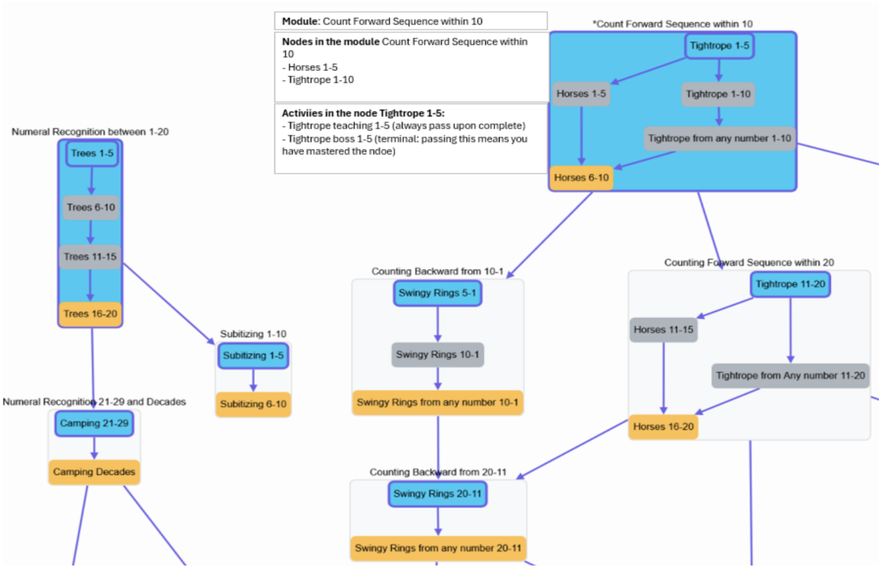


Figure 1 A Sample of the Node Map of Games in *My Math Academy*

The study focused on cohort 1 *MMA* students in a large-scale, multi-site cluster randomized study from the West, Central, and Southwestern of the United States. To explore the relationships between the treatment students' *MMA* usage, game-based performance, and math achievement, we conducted descriptive and correlation analyses of *MMA* system usage data and pre- and post-math achievement data. The system data (422 students in 21 *MMA* classrooms) that captures real-time student game play usage was used to understand students' in-game behaviors and progress on grade-level activities. Pre- and Post- student math achievement (a subsample of 214 students) were evaluated using the Test of Early Mathematics Ability-3 (TEMA-3; Ginsburg & Baroody, 2003).

3 INITIAL RESULTS

Preliminary results indicated that *MMA* could be flexibly implemented with adequate fidelity. Teachers implemented *MMA* for 20 – 26 weeks with students actively engaged in in-game activities for at least 25 minutes each week. Students who improved more on the TEMA-3 assessment appeared to complete and pass more activities, as well as learn and pass more nodes at or below their grade level. Correlational analyses showed strong positive correlations between engagement and in-game learning outcomes (total time spent, r = .77; average weekly minutes, r = .66; active weeks, r = .44). While these relationships are intuitive, exploratory analyses of the usage and progress data using M5Rules regression algorithm identified five distinct learning groups (high-volume, slow-paced, selective-cancel/high-success, fast-paced, and low-duration learner groups). The high-volume learners demonstrated both the highest average weekly minutes and achieved the most learning objectives. Slow-paced learners, despite moderate usage time, completed activities at the lowest rate and achieved the fewest learning objectives. The selective-cancel/high-success group showed strategic behavior, maintaining the highest pass-to-cancellation ratio and demonstrating intentional activity selection. Fast-paced learners achieved efficiency, completing the most activities per hour while maintaining moderate weekly engagement. Low-duration learners showed notable efficiency, achieving more learning objectives than the slow-paced group despite having the lowest total engagement time. These distinct patterns suggest that learning in play-based adaptive systems occurs through multiple viable approaches, with implications for designing systems that can support diverse learning styles.

4 CONCLUSIONS AND FUTURE WORK

The ongoing study provides one example of how a technology- and game-based program can supplement regular math instruction to address individual students' needs. The study results increase the knowledge about students' in-game progress and behaviors and their relationship to learning outcomes. Subsequent analyses are being conducted to optimize our understanding of students' usage profiles and trajectories related to learning outcomes (e.g., cluster analyses of learning behaviors and sequence mining of successful learning patterns, investigation of the relationship between cancellation rates / cancel types and learning outcomes). In addition, future research will include the full sample of the large-scale study to discuss the impact of *MMA* on student math learning, as well as analyzing subgroups by student demographics.

REFERENCES

Devlin, B. L., Jordan, N. C., & Klein, A. (2022). Predicting mathematics achievement from subdomains of early number competence: Differences by grade and achievement level. Journal of experimental child psychology, 217, 105354. https://doi.org/10.1016/j.jecp.2021.105354

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability Third Edition Examiner's Manual*. Austin, TX: Pro-ed.

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.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A Brief Introduction to Evidence-centered Design* (p. 37). Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/RR-03-16.pdf

National Center for Education Statistics (NCES) (2019). The Nation's Report Card. https://www.nationsreportcard.gov/ltt/mathematics/performance/?age=9

Nores, M., & Barnett, S. (2014). *Access to high-quality early care and education: Readiness and opportunity gaps in America*. New Brunswick, NJ: Center for Enhancing Early Learning Outcomes and National Institute for Early Education Research.

Plass, J. L., Homer, B., & Kinzer, C., (2016). *Foundations of Game-Based Learning.* Educational Psychologist. 50. 258-283. 10.1080/00461520.2015.1122533.

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In *Handbook of research on educational communications and technology* (pp. 311-321). Springer, New York, NY.

Sullivan, P., Zevenbergen, R., & Mousley, J. (2003). The contexts of mathematics tasks and the context of the classroom: Are we including all students?. *Mathematics Education Research Journal*, *15*(2), 107-121. https://doi.org/10.1007/BF03217373

Optimizing Learning Analytics Dashboards for Self-Regulated Learning: Insights and Practical Recommendations

Eunsung Park
Tennessee Tech University
epark@tntech.edu

Cory Gleasman
Tennessee Tech University
cgleasman@tntech.edu

ABSTRACT: This study explores how learning analytics dashboards can better support self-regulated learning (SRL) in adaptive environments. Guided by Winne and Hadwin's (1998) SRL model, we examine the disconnect between dashboard design and student usage, particularly around task definition, goal setting, strategy enactment, and monitoring. Findings reveal diverse student interpretations of features such as pre-tests and visualized course structure, along with varied use of granular versus composite indicators. Key recommendations for improving dashboard design include clarifying feature purposes, aligning indicators with grading, and offering flexible content presentation options. These insights provide actionable recommendations for enhancing student engagement and SRL in adaptive learning systems.

Keywords: Learning Analytics, Self-Regulated Learning, Dashboard Design, Adaptive Learning

1 INTRODUCTION

Learning analytics (LA) dashboards are integral to personalized and adaptive learning, offering insights into student engagement and progress (Park et al., 2023; Schumacher & Ifenthaler, 2018). However, these dashboards often fail to fully support self-regulated learning (SRL) due to a disconnect between their design and how students engage with them. SRL, as outlined in Winne and Hadwin's (1998) model, involves task definition, goal setting, strategy enactment, and metacognitive monitoring. While dashboards are intended to support these phases, gaps persist in aligning their functionality with students' needs, particularly in adaptive learning contexts.

2 PROBLEM STATEMENT & PURPOSE OF THE STUDY

The primary challenge in designing effective LA dashboards is supporting SRL in open-ended learning environments, where students shape their understanding of tasks through cognitive and task conditions, impacting how they interpret feedback and set goals (Park et al., 2023). This complexity often creates a misalignment between dashboard design and students' actual learning experiences, limiting their effectiveness in promoting SRL (Matcha et al., 2019). To address this gap, this study adopts a student-centered approach to explore how students use LA dashboards and investigates design improvements to better align dashboard elements with SRL processes.

3 METHODS

This study employed reflexive thematic analysis (RTA) as outlined by Braun and Clarke (2019) to examine how fourteen participants (12 undergraduate students, 1 instructor, and 1 designer) engaged with an adaptive learning system. Student participants were interviewed twice during the semester—at the sixth week and four weeks before its end—to capture evolving experiences and reduce memory demands. A total of 26 one-hour interviews were transcribed and analyzed using MAXQDA software, following the six-phase RTA process: familiarization, generating codes, constructing themes, revising themes, defining themes, and producing the report. The lead researcher independently conducted all phases, emphasizing emergent patterns and context-specific meaning-making aligned with the SRL framework.

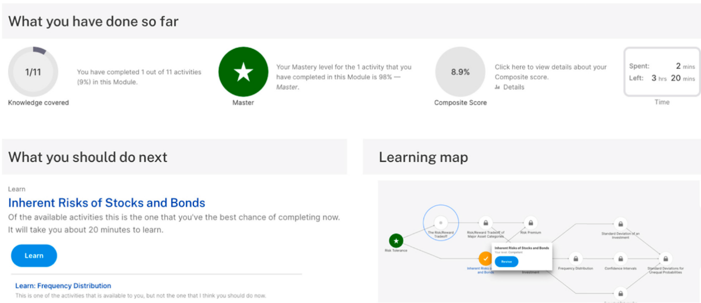


Figure 1: Dashboard Design

4 FINDINGS

Pre-Tests as Learning Tools: The dashboard included pre-tests as an integral design element, offering flexibility in how students approached their learning. While some used pre-tests as intended—to skip mastered content—others incorporated them into study routines to reinforce understanding and achieve mastery. For some, pre-tests became central to their strategies, whereas others engaged with all content sequentially, disregarding the skip option—highlighting diverse interpretations and uses.

Learning Maps and Recommendations for Strategic Planning: Initially, students found real-time recommendations unclear and unhelpful for meaning-making. Over time, many students shifted their reliance to the visualized course structure (learning map) as their primary guide for progress. Presented in a tree-like structure, the map allowed students to plan their next steps and develop adaptive strategies, frequently bypassing the system's real-time recommendations. This shift underscores the map's value as a planning tool aligned with students' performance goals.

Selective Use of Granular and Composite Learning Indicators for Metacognitive Monitoring: Students adapted their monitoring by shifting between composite and granular indicators. They initially focused on high-level, composite indicators to track overall progress, but as they advanced, they shifted to more detailed, concept-specific indicators for targeted monitoring. This adaptive behavior underscores students' prioritization of indicators that directly aligned with grading, enabling focused SRL.

Rehearsal and Review Flexibility: The chunked content provided through the dashboard supported routine study but posed challenges during exams. For regular participation, students preferred repetitive review of smaller chunks to meet ongoing goals. However, for assessments, they shifted to a linear review approach, progressing from simple to complex material. Although not explicitly a dashboard element, the rehearsal process was shaped by how the dashboard organized and presented content, influencing students' navigation and strategies.

Greater flexibility and alignment with evolving goals could enhance dashboard effectiveness, especially in high-stakes assessments.

5 CONCLUSION

This study highlights a disconnect between LA dashboard design and students' SRL processes, particularly in task definition, goal setting, strategy development, and monitoring. Addressing these gaps requires clarifying the purpose of key dashboardelements, such as pre-tests and recommendations, and providing clearer guidance on their use. Aligning learning indicators with grading criteria and progress toward learning goals can reduce confusion and improve students' ability to monitor their progress. Additionally, flexible content presentations that adapt to varying SRL phases, such as learning versus review, can further support engagement and development.

5.1 Practical Recommendations

Based on exploratory findings, educators and designers could enhance dashboards by offering clear guidance on using tools like pre-tests and integrating visualized learning maps with actionable recommendations. These adjustments can help students make sense of their learning process, identify areas for improvement, and enhance strategic planning. Aligning progress indicators with grading metrics and goals may reduce confusion and improve self-monitoring. Flexible content formats, such as chunked and linear options, can help students adapt to diverse learning needs. Additionally, continuous feedback can refine strategies and support more informed decision-making.

REFERENCES

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In P. Liamputton (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer. https://doi.org/10.1007/978-981-10-5251-4_103

Matcha, W., Uzir, N. A., Gasevic, D., & Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226–245. https://doi.org/10.1109/tlt.2019.2916802

Park, E., Ifenthaler, D., & Clariana, R. B. (2023). Adaptive or adapted to: Sequence and reflexive thematic analysis to understand learners' self-regulated learning in an adaptive learning analytics dashboard. *British Journal of Educational Technology, 54(*1), 98-125. https://doi.org/10.1111/bjet.13287

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior, 78*, 397–407. https://doi.org/10.1016/j.chb.2017.06.030

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 291–308). Erlbaum. https://doi.org/10.4324/9781410602350-19

Towards Analytics for Self-regulated Human-AI Collaboration in Writing

Antonette Shibani*, Vishal Raj and Simon Buckingham Shum
University of Technology Sydney
*antonette.shibani@uts.edu.au

ABSTRACT: Learning analytics offers significant potential to examine learner process data in the age of generative AI. This study examines collaboration dynamics in human-AI co-writing systems using keystroke metrics and clustering. Our findings show three distinct types of behaviors when writers co-write with suggestions from a large language model: *Balanced collaborators, AI-reliant writers,* and *Independent creators*, as the first step towards uncovering nuanced dynamics of human-AI interaction. We posit that such analytics on writing behaviours can be used as proxies to provide learners feedback on their use and collaboration with AI in writing for self-regulation.

Keywords: writing, generative AI, coauthor, keystroke log, clustering, writing analytics, LLM

1 INTRODUCTION

Generative Artificial Intelligence (GenAI) tools have blurred the lines between products created by learners and AI, prompting a greater focus on the processes involved in learning and assessment (Lodge, 2023). Studying the process is becoming equally important as the products created through Writing, where Learning Analytics (LA) has significant scope in supporting learners through the analytics it can provide based on observed behaviours.

Analytics on writing processes with AI is particularly important in the generative AI era, where AI use intertwines with the cognitive processes of learners enabled by hybrid human-AI systems. The challenge of understanding human-AI collaboration opens new possibilities for Learning Analytics, in principle to support educators and students, both of whom could benefit from being made aware of good or poor processes. We present a preliminary study to demonstrate how analytics on GenAI usage may be tracked and fed back to the learner for improved self-regulation, with an example derived from human-AI writing process data. We base it on the theoretical lens of the Community of Inquiry framework, previously used to analyze human-AI interactions through clustering to study how learning happens in AI-supported language learning (Wang et al., 2023).

Our study builds on two prior works in human-AI collaborative writing: **1. a publicly available keystroke log dataset called *CoAuthor*** (Lee et al., 2022) - The term 'CoAuthor' is the name of the tool and data set used by authors to refer to the human-AI hybrid writing system where writers could obtain AI suggestions on demand as they wrote argumentative and creative essays (Lee et al., 2022) in response to a given prompt for at least 10 minutes. The system provided initial prompts with the topic and starter sentences for their writing, which were then continued by the writers. Keystroke-level interaction data between writers and a large language model (GPT-3) from 1446 writing sessions in English where 61 writers wrote 830 creative stories, and 615 argumentative pieces released open-source formed the basis of this study. **2. '*CoAuthorViz*' - metrics of human-AI dynamics at a sentence-level**, extended from the CoAuthor data set (Shibani et al., 2023).

While known for complexity in deriving meaningful inferences from fine-grained data, keystroke logging enables data collection in the background without obvious interferences with the writer's performance or their writing process (Leijten & Van Waes, 2013), and can inform writer behaviors. This research investigates the characteristics of writers when engaging with AI co-writing using keystroke log metrics, as the first step towards identifying more nuanced dynamics and feedback.

2 METHOD

K-means clustering, an unsupervised learning algorithm that classifies items into k number of groups was used to cluster writing sessions based on CoAuthorViz writing metrics, including sentence-level (e.g., user-authored sentences, GPT-3-authored sentences), API-level (e.g., GPT-3 calls made, suggestions accepted/rejected), and ratio metrics (e.g., GPT-3 : Total Sentences, User : Total Sentences) - the full list of metrics used and their definitions can be found in the paper: Shibani et al., 2023). Metrics were scaled and standardized to ensure equal contribution during clustering. The optimal number of clusters was determined to be 3 based on an integrated analysis of the Elbow method, Silhouette Score (S = 0.38), Calinski-Harabasz Index (CH = 816.8), and manual evaluation. We used Principal Component Analysis (PCA) to visualise high-dimensional data and reduce its noise and redundancy. The analysis was run using Pandas, matplotlib, seaborn, and sklearn libraries in Python.

3 PRELIMINARY FINDINGS AND DISCUSSION

From 1446 individual writing sessions and each data point here representing a writing session, we found three distinct types of writer profiles below - visualized in Figure 1: Balanced Collaborators (Cluster 0), AI-Reliant Writers (Cluster 1), and Independent Creators (Cluster 2), which aligns with results from prior case studies (Shibani et al., 2023; Wang et al., 2023). The visualization is based on the first two principal components capturing the largest variance in the dataset. While PCA reduces dimensionality for visualization (appearing to contain overlapping data points), this doesn't affect clustering robustness in the original high-dimensional space.

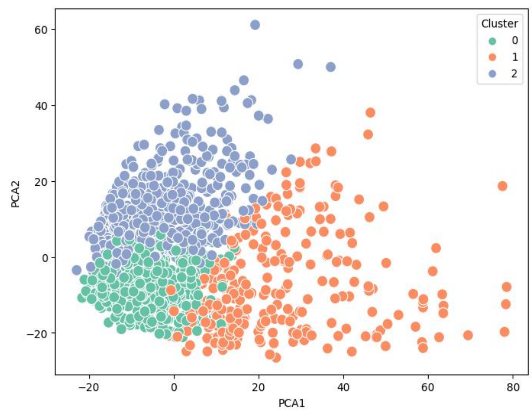


Figure 1: Three clusters of writer profiles visualized using 2D PCA

Most writing sessions (41.7%) consisted of balanced collaborators, 39.6% were created by independent creators, and the least (18.7%) were AI-reliant. The clusters were named based on the behavioral patterns observed in the analysis, reflecting their relative reliance on GPT-3 and the balance between AI and human input – key statistics are as follows:

Balanced Collaborators (Cluster 0): In these sessions, writers showed medium reliance on AI, with an average of 10-11 calls made and a GPT-3 usage amount of 0.28-0.29. They generated content with a mix of their own writing and GPT-3 suggestions, with 46-48% of sentences being user-authored. They produced fewer total sentences (21-22) than other clusters and seemed to have a higher number of sentences in the initial prompt (5.28-5.3).

AI-Reliant Writers (Cluster 1): In these sessions, writers relied more heavily on AI, with an average of 27-28 calls made and a GPT-3 usage amount of 0.56. They generated content with a lower proportion of user-authored sentences (29-32%) and a higher proportion of GPT-3-authored sentences (681-1036%) from other clusters. They produced a moderate number of total sentences (34-36) compared to other clusters and had a higher number of sentences in the initial prompt (4-5).

Independent Creators (Cluster 2): Writers grouped under this cluster showed minimal reliance on GPT-3 during the writing session, with an average of 8-10 calls made and a GPT-3 usage amount of 0.17-0.28. They generated content primarily on their own, with 72% of sentences being fully user authored. These writers seemed to have lower number of sentences in the initial prompt (~3).

It is to be noted that the results are based on writing *sessions*; a writer could have multiple writing sessions where they exhibit different AI-human collaboration dynamics. Grouping the 61 writers into one cluster or another based on *all* their writing sessions proved to be harder, as they tended to fluctuate between genres. Notably, 72% of writers were present in multiple clusters, suggesting genre-dependent collaboration dynamics rather than distinct patterns of writer behavior. This posits the need to account for fluidity in writer characteristics according to writing contexts when profiling writers, for instance, when designing LA support by recognizing the changing nature of writers. Future work could also examine quality indicators for writing that can determine if certain types of collaboration with AI can lead to better writing outputs.

4 IMPLICATIONS

The three prominent clusters of writer profiles emerging from the study highlight the differences in individual writing behaviors when collaborating with AI and can be helpful proxies for learners to reflect on their AI use when returned to them. Extensions of the work can add additional perspectives, explainers, and feedback to inform writers of their AI co-authorship behaviors that best support self-regulated learning. This ensures that learners can reflect on their reliance/ independence on AI for writing to make informed decisions and better equip themselves for an AI-integrated future.

REFERENCES

Lee, M., Liang, P., & Yang, Q. (2022). *Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities* Proceedings of the CHI22 conference on human factors in computing systems, New Orleans, USA.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358-392.

Lodge, J. M., Howard, S., Bearman, M., Dawson, P, & Associates (2023). *Assessment reform for the age of artificial intelligence*. A. G. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

Shibani, A., Rajalakshmi, R., Mattins, F., Selvaraj, S., & Knight, S. (2023, 2023). Visual representation of co-authorship with GPT-3: Studying human-machine interaction for effective writing.*M. Feng, T. K¨aser, and P. Talukdar* 16th International Conference on Educational Data Mining, Bengaluru, India. https://doi.org/10.5281/zenodo.8115695.

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, *194*, 104703.

Learning Analytics Reporting Framework (LARF): Insights into Existing Research on Adoption and Use of Learning Analytics

Barbara Šlibar[1], Diana Šimić[1], Sabina Rako[1,2], Jelena Gusić Munđar[1], Nina Begičević Ređep[1]
[1]University of Zagreb Faculty of Organization and Informatics, [2]University of Zagreb University Computing Centre
bslibar@foi.unizg.hr, diana.simic@foi.unizg.hr, sabina.rako@srce.hr, jelena.gusic@foi.unizg.hr, nina.begicevic@foi.unizg.hr

ABSTRACT: Learning analytics (LA) offers new opportunities for enhancing educational experiences, yet the diversity in its research approaches has led to challenges in synthesizing findings and ensuring reproducibility of results. This study aims to demonstrate a theoretical framework for reporting on LA research (LARF), drawing on the Learning Analytics Reference Model and the Generic Framework for Learning Analytics. Initially structured hierarchically, the proposed framework evolved into a matrix structure with three aspects—Learning, Analytics, and Research—and seven dimensions Pedagogy, Context, Implementation, Stakeholders, Research Methods, Data, and Ethics, based on expert input from a focus group. This framework seeks to standardize LA reporting, supporting quality assessment, reproducibility, and comparative research. Ongoing work focuses on evaluating expert agreement on metadata relevance, restructuring the framework into a checklist aligned with the IMRD (Introduction, Methodology, Results, and Discussion) structure, and conducting expert reviews to refine the checklist for practical use.

Keywords: research repeatability, research reproducibility, research replicability, reporting guidelines, learning analytics application

1 INTRODUCTION

The emergence of learning analytics (LA) in 2010 opened new avenues for enhancing educational experiences. LA's interdisciplinary nature attracts researchers from diverse paradigms, though its novelty has led to varied research content and formats, complicating comparison and synthesis of findings (Baker et al., 2021). Persistent challenges include issues with reproducibility and limited empirical evidence on LA's effectiveness, common in social sciences (Kitto et al., 2023). Lessons from established fields such as medicine can guide robust research practices in LA. Initiatives like the EQUATOR Network[1] and the Cochrane Network[2] provide valuable frameworks for high-quality research protocols and reporting guidelines. The Learning Analytics Community Exchange (LACE) project represents a pioneering effort to compile empirical evidence in LA (Ferguson & Clow, 2016). Additionally, various checklists have been adopted within the LA community, such as the DELICATE checklist for ethical and trustworthy LA (Drachsler & Greller, 2016) and design checklists for

[1] https://www.equator-network.org/

[2] https://www.cochrane.org/

dashboards (Kaliisa et al., 2023). This research is inspired by the LACE project (Ferguson & Clow, 2016), studies on epistemic paradigms in LA (Baker et al., 2021), and literature on research protocols in medicine (Moher et al., 2010). Thus, the objective is to establish a framework for reporting on research related to the adoption and application of learning analytics.

2 METHODOLOGY

To define the framework for reporting on research related to the adoption and use of learning analytics, two research methods were employed: a literature review and a focus group. The literature review structured the initial version of the framework. Following this, the focus group consisted of multiple steps, aimed at eliciting knowledge about the key metadata needed to assess the quality of reported research on the application of learning analytics in practice, ensure research reproducibility, and compare and synthesize research results. However, only the results from the first step of the focus group, which involved discussing the initial version of the framework, are presented here. The focus group was composed of five senior researchers fluent in English, each with experience in editing, reviewing, or publishing in learning analytics or related fields (e.g., educational data mining). Participants were selected to ensure independence from one another (e.g., no hierarchical relationships such as department chair and employee). Participants were asked to provide input from three perspectives: as an editor or reviewer for a journal or conference proceedings, as a reader, and as a writer of a research paper. Participants signed informed consent and agreed to confidentiality. The focus group used Google Jamboard and Mentimeter, with Zoom recording audio and only the Jamboard visually.

3 RESULTS AND DISCUSSION

The Learning Analytics Reference Model (Chatti et al., 2012) and the Generic Framework for Learning Analytics (Greller & Drachsler, 2012) were selected as the basis for developing theoretical framework for learning analytics research. The initial framework was defined hierarchically, with questions nested within dimensions. During the first step of the focus group, experts recommended adding two

Table 1: The learning analytics reference model (Chatti et al., 2012), the generic framework for learning analytics (Greller & Drachsler, 2012), and learning analytics reporting framework (LARF).

<table>
<tr><th rowspan="2">(Chatti et al., 2012)</th><th rowspan="2">(Greller & Drachsler, 2012)</th><th colspan="3">LARF</th></tr>
<tr><th>Learning</th><th>Analytics</th><th>Research</th></tr>
<tr><td>What? (Data, Environments)</td><td>Data</td><td></td><td colspan="2">Data</td></tr>
<tr><td>Who? (Stakeholders)</td><td>Stakeholders</td><td></td><td>Stakeholders</td><td></td></tr>
<tr><td>How? (Techniques)</td><td>Instruments</td><td></td><td rowspan="3">Implementation</td><td rowspan="2">Research methods</td></tr>
<tr><td>Why? (Objectives)</td><td>Objectives</td><td>Pedagogy</td></tr>
<tr><td colspan="2" rowspan="2">Internal Constraints</td><td></td><td rowspan="2"></td></tr>
<tr><td colspan="2" rowspan="2">Context</td></tr>
<tr><td colspan="2" rowspan="2">External Constraints</td><td></td></tr>
<tr><td colspan="3">Ethics</td></tr>
</table>

additional dimensions: Ethics and Context. Ethics reflects e.g., stakeholders' rights, code of ethics, risk of unintended outcomes. Context reflects e.g., educational setting, level of LA implementation, regulatory environment. Although the initial framework was hierarchical, it was noted that some questions could be interpreted from different perspectives. For instance, LA research may have objectives related to generating new knowledge (research objectives) and applied objectives aimed at improving learning (pedagogy). As a result, the final structure of the proposed framework was modified to a matrix format (see Table 1), comprising three aspects (Learning, Analytics, and Research) and seven dimensions (Pedagogy, Context, Implementation, Stakeholders, Research Methods, Data, and Ethics).

4 CONCLUSION AND FURTHER WORK

Ongoing efforts include evaluating expert agreement on the relevance of the proposed metadata, reorganizing the framework's questions into a checklist that aligns with the standard research paper structure (IMRD: Introduction, Methodology, Results, and Discussion), and conducting a post-evaluation in which experts review the checklist to confirm it accurately reflects the focus group conclusions. This process aims to improve the framework's practical usability in LA research reporting.

ACKNOWLEDGMENTS

This work has been fully funded by the Croatian Science Foundation under the grant IP-2020-02-5071.

REFERENCES

Baker, R. S., Gašević, D., & Karumbaiah, S. (2021). Four paradigms in learning analytics: Why paradigm convergence matters. *Computers and Education: Artificial Intelligence*, *2*, 100021.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318.

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 - LAK '16*, 89–98.

Ferguson, R., & Clow, D. (2016). Learning analytics community exchange: Evidence hub. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 520–521.

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, *15*(3), 42–57.

Kaliisa, R., Jivet, I., & Prinsloo, P. (2023). A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards. *International Journal of Educational Technology in Higher Education*, *20*(1), 28.

Kitto, K., Manly, C. A., Ferguson, R., & Poquet, O. (2023). Towards more replicable content analysis for learning analytics. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 303–314.

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for Developers of Health Research Reporting Guidelines. *PLoS Medicine*, *7*(2), e1000217.

Evaluating GPT-4 at Grading Handwritten Solutions in Math Exams

Adriana Caraeni, Alexander Scarlatos, Andrew Lan
University of Massachusetts Amherst
acaraeni@umass.edu, ajscarlatos@cs.umass.edu, andrewlan@cs.umass.edu

ABSTRACT: Recent advances in generative artificial intelligence (AI) have shown promise in accurately grading open-ended student responses. However, few prior works have explored grading handwritten responses due to a lack of data and the challenge of combining visual and textual information. In this work, we leverage state-of-the-art multi-modal AI models, in particular GPT-4o, to automatically grade handwritten responses to college-level math exams. Using real student responses to questions in a probability theory exam, we evaluate GPT-4o's alignment with ground-truth scores from human graders using various prompting techniques. We find that while providing rubrics improves alignment, the model's overall accuracy is still too low for real-world settings, showing there is significant room for growth in this task.

Keywords: Automated Scoring, Handwritten Grading, Large Language Models, Multi-Modal

1 INTRODUCTION

Automated scoring is a key challenge to enable the deployment of open-ended questions at scale. Researchers have studied the problems of automated essay scoring (AES) [Attali & Bernstein, 2006] and automated short-answer grading (ASAG) [Burrow et al., 2015] extensively, often using AI. Fine-tuning language models, such as BERT, has been effective on these tasks [Zhang et al., 2022]; others have recently explored prompting large language models (LLMs) for automated scoring [Stahl et al., 2024]. One important setting in automated scoring is student handwriting: in many practice and assessment settings, students write down their solutions to problems on paper, which are then converted to images, common in science, technology, engineering, and math (STEM) fields [Baral et al., 2023]. This automated scoring task is challenging: compared to scoring textual essays and short answers, images contain rich semantic, visual, and even spatial information on student thought processes, which require significant textual, mathematical, and visual reasoning capabilities from AI. Recent advances in multi-modal foundation models, especially vision-language models, have significantly advanced the textual and visual reasoning abilities of AI. In this work, we perform a preliminary exploration into automated scoring of handwritten student responses to math exams using OpenAI's GPT-4o model. We evaluate several prompts, analyze failure patterns, and find that GPT-4o's scoring ability is significantly lower than that of human graders. One recent study also evaluates AI in handwritten math grading [Liu et al., 2024]. However, they use responses from an optional exam where students may not put in much effort, while we use real final exam responses, yielding a more realistic data source that captures student behavior in actual test-taking settings.

2 EXPERIMENTAL SETUP

Using an IRB-approved process, we collect a dataset of real handwritten final exam responses from a single semester of a probability theory course at a university in the United States. We emailed students from the course asking permission to use their exam responses for research; we use exams from the 18 students who gave consent. We did not collect any demographic information from students. The 120-minute exam contains 5 questions, each with 3 independent sub-parts, covering the topics of probability estimates, game theory, Markov chains, Bayes nets, and parameter estimation. Each question was scored by a single grader using rubrics they wrote to assign credit for partially correct solutions. Student written exam responses contained text, mathematical formulas, and diagrams, all critical to understanding their solutions. Students in our sample scored 89.88% on the exam on average, with many points from partial credit on incorrect answers.

We use OpenAI's recent GPT-4o model to assign scores to student responses. We prompt the model to grade one question at a time, providing a scanned image of the corresponding page from the student's exam and telling the model how many points each part is worth. We experiment with 3 different prompt types: i) no context (**N**), where the model only sees the student response, ii) correct answer (**C**), where the model sees the student response and the correct answer for each question part, and iii) correct answer and rubric (**CR**), where the model sees the student response, the correct answer, and the rubric for each question part. We measure how well GPT-4o can score student responses, which we refer to as *alignment*, by comparing its predicted scores to the ground truth scores assigned by course graders. We examine scores at the question level, resulting in 18 x 5 = 90 samples, and normalize scores between 0 and 1 based on the total points per question. We then compute the mean absolute error (**MAE**), root mean squared error (**RMSE**), accuracy (**Acc.**), and Pearson's correlation coefficient (**Corr.**) between predicted and ground truth question scores. We also show the average score assigned by graders (**Score G.**) and by the model (**Score M.**).

3 RESULTS

Table 1: Average alignment by prompt type. Providing the answer and rubric performs the best.

Prompt Type	MAE ↓	RMSE ↓	Acc. ↑	Corr. ↑	Score G.	Score M.
N	0.0940	0.1533	0.4222	0.2776	0.8988	0.9759
C	0.0989	0.1609	0.4333	0.5502	0.8988	0.8501
CR	0.0766	0.1267	0.4667	0.6174	0.8988	0.8808

To determine how relevant context in the prompt is, we show the alignment metrics and scores averaged over all students and questions partitioned by prompt type in Table 1. We observe that CR performs the best, indicating that a correct answer and rubric is necessary for GPT-4o to grade student responses accurately. We make two observations to explain this result: i) N tends to *overestimate* student scores since it inaccurately judges solution correctness without a reference, and ii) C tends to *underestimate* student scores since it rarely assigns as much partial credit as the human graders. While CR solves these two issues, we see that its predicted scores are still off by 7.66% on average, indicating there is significant room for improvement in the handwritten grading task.

Table 2: Average alignment per question using CR, varying greatly across questions.

Question	MAE ↓	RMSE ↓	Acc. ↑	Corr. ↑	Score G.	Score M.
1	0.0833	0.1302	0.3889	0.4261	0.9028	0.9083
2	0.1235	0.1697	0.3889	0.4353	0.8580	0.7716
3	0.1011	0.1430	0.1667	0.5670	0.8167	0.8211
4	0.0389	0.0850	0.6667	0.3809	0.9694	0.9639
5	0.0361	0.0825	0.7222	0.8512	0.9472	0.9389

To determine which types of questions GPT-4o has difficulty grading, we show the alignment metrics averaged over all students with the CR prompt type for each question in Table 2. We observe a large difference in performance across questions, with questions 4 and 5 generally more accurate than the others. GPT-4o falls short on questions 2 and 3 primarily because students are required to justify their answers in these questions, and the model struggles to identify when these justifications are correct. In question 1, GPT-4o tends to give full marks for faulty solutions in part 2, which is on Chebyshev's inequality. It often cannot identify the incorrect step in these solutions, possibly because they are relatively long, and this type of problem may be infrequent in the model's training data. We also note that model accuracy is roughly correlated with student performance (Score G.), indicating the model has more trouble identifying issues with incorrect solutions than simply identifying correct solutions.

We also perform a qualitative analysis of the model's outputs when using the CR prompt, identify common errors, and propose solutions to explore in future work. First, the model occasionally marks clearly correct answers as incorrect or vice versa. This may be from struggling to read student handwriting, or possibly information overload from the lengthy solutions. It may be possible to ask the model if it can clearly read and understand the student solution and defer grading if it cannot. Second, the model often cannot understand if the reasoning in a student solution is correct. It may be beneficial to provide the model with a full handwritten correct solution as reference. Finally, the model sometimes misinterprets the rubrics. For example, in question 2, the rubric implies that student justifications should reference payoff matrix values; the model often removes points via this item while human graders do not. While the rubrics work for human graders, it may be helpful to write custom rubrics that are more interpretable by the model.

4 DISCUSSION, CONCLUSION, AND FUTURE WORK

In this work, we evaluate the ability of GPT-4o, a powerful multi-modal large vision-language model, to grade real handwritten student responses in college math exams. We find that providing a correct answer and rubric as reference are necessary to improve alignment with human graders, but that GPT-4o still struggles to accurately assign scores for many reasons. In particular, we find that the model struggles to comprehend student solutions, either from i) trouble reading the student's handwriting, ii) not knowing the true correct solution steps, or iii) incorrectly interpreting the reasoning behind a student's response. There are many avenues for future work. First, researchers should assess GPT-4o's performance on sub-tasks, such as transcribing or reasoning over solutions. Second, researchers should investigate if fine-tuning open-source models like Llama 3.2 can improve alignment. Finally, researchers should evaluate visual grading in more domains, such as computer science or visual arts.

REFERENCES

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, *4*(3).

Baral, S., Botelho, A., Santhanam, A., Gurung, A., Cheng, L., & Heffernan, N. (2023). Auto-Scoring Student Responses with Images in Mathematics. *International Educational Data Mining Society*.

Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. *International journal of artificial intelligence in education*, *25*, 60-117.

Liu, T., Chatain, J., Kobel-Keller, L., Kortemeyer, G., Willwacher, T., & Sachan, M. (2024). AI-assisted Automated Short Answer Grading of Handwritten University Level Mathematics Exams. *arXiv*.

Stahl, M., Biermann, L., Nehring, A., & Wachsmuth, H. (2024). Exploring LLM Prompting Strategies for Joint Essay Scoring and Feedback Generation. *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications*.

Zhang, M., Baral, S., Heffernan, N.T., & Lan, A.S. (2022). Automatic Short Math Answer Grading via In-context Meta-learning. *International Educational Data Mining Society*.

Cognition, Metacognition, and Emotions in Self-Regulated Lesson Design: Preliminary Results from EDA Signals and Log Files

Lingyun Huang
The Education University of Hong Kong
lingyunhuang@eduhk.hk

Kuangzhe Xu
Hirosaki University
jokyotetsu@hirosaki-u.ac.jp

ABSTRACT: This study leveraged electrodermal activities (EDA) and computer trace data to capture the cognitive, metacognitive, and emotional processes of self-regulated learning (SRL) when student teachers were performing a technology-enhanced lesson (TEL) design task with the nBrowser, a computer-supported learning environment. We used SCL and SCR to indicate emotional arousal, which can indicate slow and rapid changes in EDA signals. The log data of 44 participants were retrieved from the nBrowser to indicate their cognitive and metacognitive regulatory processes. The results showed the student teachers had diverse emotional experiences in SRL processes while in the TEL task. They had a higher level of emotional arousal in metacognitive regulatory processes but less intensive emotional arousal in some goal-setting processes. These findings provide evidence regarding using EDA and log data to identify the dynamics of SRL processes. It also has implications for the importance of scaffolding metacognitive regulatory processes.

Keywords: SRL, EDA, Log data, CBLE

INTRODUCTION

Student teachers must be self-regulated when acquiring technology-enhanced lesson (TEL) design skills. They should perceive the affordances of digital tools, make connections with content and instructional progress, and tailor their designs to the specific needs of students (Mishra & Koehler, 2006). This emphasizes regulating the cognitive and metacognitive processes involved in lesson designs. Extensive research indicates that self-regulated learning (SRL) describes a multidimensional learning process in which learners monitor and control their cognitive and metacognitive aspects to achieve goals (Azevedo & Gašević, 2019). These cognitive and metacognitive aspects of SRL facilitate the effective allocation of effort toward task completion (Poitras et al., 2017). In the meantime, students will experience different emotional states during an SRL process, and the emotional experiences will impact how students engage in cognitive and metacognitive-regulated learning (Taub et al., 2021). The literature illustrates that cognitive and metacognitive SRL and emotions may occur when student teachers perform a TEL task. For example, Huang et al.'s (2022) study demonstrated a positive relation between the emotional processes and student teachers' cognitive and metacognitive regulation in a TEL task, which lays the foundations for further investigations.

This study sought to reveal the student teachers' emotional arousal in self-regulated TEL tasks. However, it is also noted that assessing cognitive, metacognitive, and emotional regulatory processes is complicated. Many scholars advocate the affordance of behavioral data like computer logs as indicators of cognitive and metacognitive regulatory processes. This is due to the wide adoption of computer-based learning environments (CBLEs) that are designed to support SRL processes and promote increased learning outcomes (Poitras et al., 2017). In terms of the measurement of emotional processes, physiological measures, such as electrodermal activity (EDA), have been considered a more objective indicator than self-reported measures that rely on individuals' subjective perceptions (Strohmaier et al., 2020). Thus, this study aims to address the following research question: How did student teachers' EDA arousals differ in different SRL processes?

METHODS AND ANALYSIS

Forty-four participants (female = 37) were student teachers from an education university in China, with a mean age of 20.86 years (SD = .82). They were asked to design a technology-integrated lesson with nBrowser, a CBLE designed based on and to scaffold SRL. The participants had 45 minutes to complete the task. While doing the task, the participants wore a device that recorded their EDA (4 Hz sampling rate). The system logged participants' events.

For analysis, we first pre-processed the EDA data and removed 11 incomplete datasets. We used Python (NeuroKit2 algorithm) to process EDA data and extracted the skin conductance levels (SCL), and the skin conductance responses (SCR). SCL and SCR reflect the general and rapid changes in the EDA signal, respectively. We conducted two simple regression analyses (EQ1) using SCL and SCR as outcome variables, respectively. Instead of specifying particular predictor variables, the effect of events was evaluated using the estimated results of random effects. In the equations, μ is the distribution of the intercept within the model. re_{event} and $re_{\text{participant}}$ represent the random effect of the events and participants. We examined whether there was a significant difference in the SCL and SCR in different SRL processes by inferring the post hoc distribution of re_{event}. We constructed a hierarchical model based on Bayesian principles using Stan for both the regression and logistic models. The prior distribution of re_{event} and $re_{\text{participant}}$ followed gamma (10,10), the prior distribution of μ follows normal (0,5), and the other parameters follow the uniform distribution. The posterior distribution significance test uses 95%HDI (Highest density interval). If the 95%HDI interval does not contain 0, it can be regarded as a significant difference.

$$SCL\ (or\ SCR) \sim Normal\left(\mu + re_{event} + re_{participant}\ ,\sigma\right)\ (EQ1)$$

RESULTS, DISCUSSION, AND CONCLUSION

Figure 1 (Left) displays the SCL in different events. SCL was high when the participants were doing the Assets_URL activity. Based on the rules mentioned previously, the positive relation was significant because the 95%HDI interval did not contain 0. In contrast, negative relations were detected when the participants were in the activities of Lesson_Details_Grade and Lesson_Details_Focus_Checked. The relations were significant as the 95%HDI interval did not contain 0. Since the nBrowser platform was designed based on SRL models, the event retrieved from the platform can indicate a particular type of SRL process. In this study, Assets_URL means that the participants labeled the online resources and saved them as assets for lesson planning, which can indicate a metacognitive evaluation process. The significant increase in SCL reveals that student teachers might have *a strong* emotional experience when deciding if the collected resources should be labeled. This further suggested that metacognitive evaluation was a bit challenging for student teachers. Moreover, Lesson_Details_Grade and Lesson_Details_Focus_Checked relate to the goal-setting processes in which student teachers define their students' characteristics and instructional content. Figure 1 showed that SCL in these two SRL processes significantly decreased, given the 95%HDI interval did not contain 0. The results could indicate that the participants may have less emotional experience in setting goals. They might feel easier or more relaxed when defining these goals. Figure 1 (Right) demonstrates the SCR in different SRL processes. It showed only one significant positive increase in SCR relating to the activity of Lesson_Details_Grade, a goal-setting process. The result is contradictory to the SCL finding. We explain that SCL and SCR reflect a slow and rapid change in the EDA signals. These participants had less authentic teaching experiences. When defining student characteristics, they had to imagine students in their minds for this task, which may account for the sudden increase in SCR.

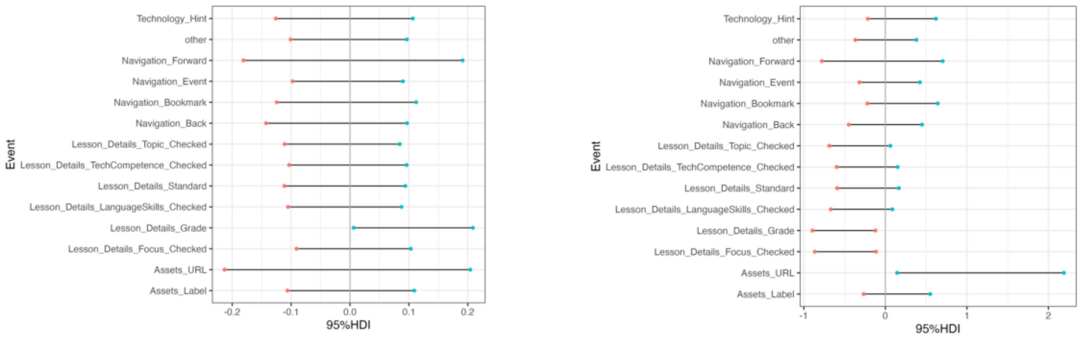


Figure 1: The SCL (Left) and SCR (Right) in different captured SRL events

In conclusion, we investigated student teachers' emotional arousal in a self-regulated TEL task. The emotions were measured by the EDA signals (i.e., SCL and SCR), and log files were used to indicate cognitive and metacognitive SRL processes. The results indicated the student teachers showed diverse EDA arousals in SRL processes while in the TEL task. Specifically, they had a higher level of emotional arousal in metacognitive regulatory processes but less intensive emotional arousal in some goal-setting processes. These findings support the existing research that indicates metacognitive regulation often challenges learners in SRL (Taub et al., 2021). Therefore, this study enriches our comprehension of SRL as a dynamic and interactive process but also underscores the importance of supporting learners' engagement in SRL necessary to optimize their learning experiences (Taub et al., 2021). However, this study has limitations. First, EDA is the only measure of emotions, which limits its accuracy in defining specific emotional experiences. Future studies should consider multimodal datasets (Azevedo & Gašević, 2019). Second, this study did not relate the investigations to learning outcomes. The study presents preliminary findings. We are endeavoring to conduct more analyses to answer more questions, such as how the interactions between the three aspects of SRL will influence student teachers' design performance.

REFERENCES

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, *96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Huang, X., Huang, L., & Lajoie, S. P. (2022). Exploring teachers' emotional experience in a TPACK development task. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-022-10135-7

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

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

Strohmaier, A. R., Schiepe-Tiska, A., & Reiss, K. M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. An application of the control-value theory. *Frontline Learning Research*, *8*(1), 16–32. https://doi.org/10.14786/flr.v8i1.427

Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (2021). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*, *72*, 101200. https://doi.org/10.1016/j.learninstruc.2019.04.001

Relationship between Hand-raising Tendency and Grades for the Design of Teaching Dashboards

Takahiro Yoshino, Kanato Murobayashi, and Hironori Egi
The University of Electro-Communications
y2330128@edu.cc.uec.ac.jp, hiro.egi@uec.ac.jp

ABSTRACT: Multimodal learning analysis is frequently used to support education by analyzing the characteristics of student behavior in the classroom. In particular, actions related to student–teacher interactions are often used. In this study, we focused on the hand-raising action performed by students to attract the teacher or TA's (Teaching Assistant) attention. This action has a clear purpose, i.e., answering a question, and it triggers interaction between the student and the teacher or TA. The number of times the hand-raising action was performed by each student was counted to identify the student's tendency to raise their hands. The results were then divided into three clusters. No marked differences in students' grades were observed within each cluster, which suggests that it may be possible to provide different optimal learning support methods for each cluster of students. In addition, the results of this study could be combined with data that show learning support priority to develop more effective learning support dashboards.

Keywords: Hand-raising tendency, Teaching assistants, Multimodal learning analytics

1 INTRODUCTION

In classroom environments, students with questions raise their hands to attract the attention of the teacher or teacher assistant (TA), and this action is performed at the student's discretion. In this study, we refer to this action as the "hand-raising action." The hand-raising action triggers interaction between the student and the teacher or TA. Students who raise their hands frequently are considered to have many questions or be active in the interaction. In contrast, students who do not raise their hands may have a high ability to solve problems independently, or they may be passive in the interaction. Thus, hand-raising actions may represent the characteristics of the students. In this study, the number of times the hand-raising action was performed was used to cluster and analyze the grades.

Multimodal learning analysis can be effective in the educational space, and it provides good results, especially in interaction-based learning, e.g., cooperative learning (Cukurova et al., 2020). In addition, the data representing the students' characteristics through dashboards must be easily understood by the teachers (Dourado et al., 2021). However, it is difficult to capture a learner's small movements, which can complicate system implementation and movement analysis. In contrast, the hand-raising action is highly visible by teachers and TAs, and it expresses a clear purpose, i.e., having their questions answered. Answering questions is an important task for teachers and TAs, and hand-raising behavior is a common interaction between teachers and students. Clustering the students based on the hand-raising action to determine a hand-raising tendency. It is also easy for teachers and TAs to understands, and it provides data that are easy to handle.

In a previous study, we developed a teaching dashboard for TAs (Ueno et al., 2023). In that study, we estimated the feelings of isolation level based on a questionnaire, and that system suggest increase the effectiveness of the TAs' encouragement. We expect that the findings of this study can be used for this kind of teaching dashboard. In addition, the findings will be useful for the determining most effective learning support method.

2 METHODOLOGY

In this study, clustering was based on the number of times students performed the hand-raising action. First, we counted the number of times each student performed the hand-raising action in each class lesson, as shown in Figure 1 (upper left). To determine the students' hand-raising tendencies, we converted the number of times hand-raising action in each lesson into the number of hand-raising action occurrences in a session for each student, as shown in Figure 1 (lower left). This will allow clustering based on the number of sessions in which students performed the hand-raising action and the number of times the hand-raising action was performed. Here, clustering was performed using Ward's method based on the number of occurrences. Analyze each cluster for differences in grades. In this study, the final exam score (total: 84 points) was used as grade data.

Lesson number	1	2	3	4	5	6
Number of hand-raising action	0	1	0	2	0	2

Number of hand-raising action	0	1	2	3	4	5
Occurrences number	3	1	2	0	0	0

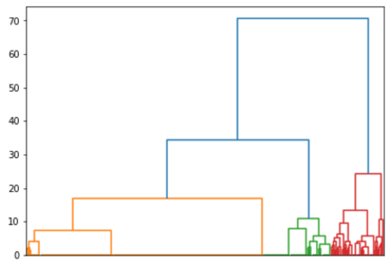


Figure 1: Conversion method for hand-raising tendency (left) and results of the Ward's clustering method (right)

3 RESULTS

Data were collected in a programming exercise class for first-year students at a science and engineering university. The target class lessons included 12 lessons, excluding group work lessons and tests. The number of hand-raising actions for each student was collected by watching a video captured from the back of the classroom. Here, data from 407 students were collected in eight classes. The student with the highest number of hand-raising actions raised their hand 39 times, and the number of students who did not raise their hands at all was 159 (representing 39% of the total number of students). The data were classified into three categories based on the results of Ward's method. The clustering results are shown in Figure 1 (right). As can be seen, in the first split, the distance to the second split is approximately half of the total distance. Then, in the second split, the distance to the third split in the cluster is approximately one quarter of the total distance. Thus, three classifications were made based on these findings. Figure 2 (left) shows the distribution of the total number of times the hand-raising action was performed and the number of sessions with the hand-raising action for each cluster. Here, each cluster is defined as "many," "middle," and "few" in descending order of the number of times the hand-raising action was performed. The number of students in each cluster was 65 for the "many" cluster, 83 for the "middle" cluster, and 259 for the "few" cluster. We found that the “many” cluster had the highest average number of hand-raising actions and the highest number of sessions with the hand-raising action. In addition, the “few” cluster had the lowest average number of hand-raising actions and included students who did not perform the hand-raising action.

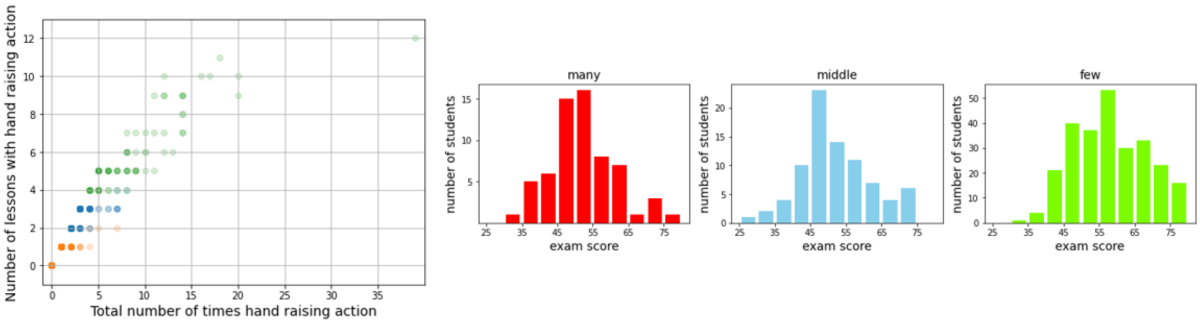


Figure 2: Distribution of hand-raising tendency clusters (left) and grades for each cluster (right)

The distribution of the grades for each cluster is shown in Figure 2 (right). The mean exam score for each cluster was 52 for “many,” 52 for “middle,” and 57 for “few.” Figure 2 also shows that there is very little difference between the grades of the different clusters. Thus, the hand-raising tendency is not related to grade, and each tendency has students who did not receive optimal support. Considering that the hand-raising action has the purpose of solving questions, it is possible that the optimal method for learning support differs for each cluster. The students with lower grades in the “few” cluster did not perform the hand-raising action very often; thus, active encouragement from the teacher or TA will be necessary. However, the students with lower grades in the “many” cluster performed hand-raising action frequently, thus, they may need more time to think on their own. Because the hand-raising action is an explicit action for TAs and teachers, it is possible to implement an automatic counting system. The results of this study may be used to develop an easy-to-understand teaching dashboard for TAs and teachers, combined with data related to grades in the future.

ACKNOWLEDGEMENTS

This work has been partly supported by the Grants-in-Aid for Scientific Research (NOs. 23K25696 and 23H00999) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan

REFERENCES

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, 51(5), 1441-1449. https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.13015

Dourado, R. A., Rodrigues, R. L., Ferreira, N., Mello, R. F., Gomes, A. S., & Verbert, K. (2021, April). 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). https://dl.acm.org/doi/abs/10.1145/3448139.3448187

Ueno, S., Yoshino, T., & Egi, H. (2023). Addressing promotion system based on student data to support desk-to-desk instruction by teaching assistants. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 162-164).

Adaptive Learning System to Suggest Break Timing based on Multimodal Data by Measuring Leg Movement

Masaki Kodaira, Ryohei Shimizu, Hironori Egi
The University of Electro-Communications
k2330044@edu.cc.uec.ac.jp, hiro.egi@uec.ac.jp

ABSTRACT: Adaptive learning systems (ALS) are modify learning content relative to the student's learning level. However, developing an ALS that is dependent on learning content is costly. An adaptive learning system that does not dependent on learning content and dynamically changes the break timing according to the learner's condition and encourages the learner to continue learning can be considered. This study investigates a system that suggests learner break timing based on learner fatigue. In the proposed method, a noncontact leg movement measurement device is used to estimate learner fatigue. Preliminary experiments demonstrated that, due to individual differences, no correlation was observed between the leg movement measurement results and the learner's voluntary break timing. However, a relationship was observed between the leg movement measurement results and time at which the learner's learning performance decreased. Thus, in the future, we plan to investigate the threshold of the system based on individual learner differences and verify the effectiveness of the system.

Keywords: Adaptive learning, Fatigue detection, Multimodal, Leg movement

1 INTRODUCTION

Adaptive learning systems (ALS) change learning content depending on each learner's proficiency level. However, changing the support method depending on the learning content is costly to develop and unrealistic. When designing an ALS that is independent of the learning content, multimodal data must be employed based on the learner's emotions and fatigue. In addition, for a content-independent ALS that promotes sustained learning, we must consider changing the timing of breaks dynamically. Thus, this study investigates a system that suggests break timing based on learner fatigue. In the proposed system, leg movement measurements (Aikawa et al., 2019) are used to estimate learner's mental fatigue because this can be realized in a noncontact manner that does not interfere with the learning activity.

A previous study classified adaptive learning and analyzed its effectiveness (Aleven et al., 2017). Adaptability is said to have three forms, i.e., step-loop adaptability, design-loop adaptability, and task-loop adaptability, and five psychological domains, i.e., student knowledge, the path through an activity, affect/motivation, self-regulation, and student learning styles. Of these 15 combinations, the effectiveness of adaptability has been demonstrated in 13 combinations, excluding design-loop and step-loop adaptability to student learning styles. The proposed method is considered to have step-loop adaptability to affect/motivations. However, to the best of our knowledge, no adaptive learning system that focuses on breaks has been proposed in this study.

A previous study focused on break timing in an adaptive learning system (Aditi et al., 2017). In that study, experiments were conducted to measure learning effectiveness under conditions of fixed intervals and dynamic timing of breaks. The results demonstrated that the learning effect was significantly better when breaks were taken at dynamic timing. However, this previous study used the learning results; thus, it is possible that the system may become dependent on the learning content. Thus, the proposed method suggests break timing based on the learner fatigue level rather than the learning content.

2 PROPOSED METHOD

We consider that learning performance will decline if learners study while fatigued. In addition, learner fatigue may accumulate unconsciously; thus, the learn may, be unaware of their fatigue. Thus, we propose a system that encourages the learner to take a break before their learning performance deteriorates by estimating the learner's fatigue from biological information as an objective indicator.

A previous study attempted to estimate a learner's level of effort using multimodal data (Kshitij et al., 2020), including gaze measurement, heart rate, brain wave, and facial expression data. As a result, the effort level and state of the learner could be estimated and classified successfully. However, having the learner wear multiple devices can be uncomfortable and hinder their learning activities. Thus, the proposed method employs a leg movement measurement technique to estimate the learner's mental fatigue. The design of the proposed system is shown in Figure 1. As can be seen, this method does not hinder the learning activities because it is a noncontact method that is out of the learner's field of vision.

Leg movement measurements vary from person to person; thus, it is necessary to understand the learner's leg movement tendencies before learning begins. After learning begins, the learner's leg movement are measured in real time, and the data are sent to the server. The server then estimates the learner fatigue from the learner's leg movement tendencies and the real-time leg movement data. If the server determines that the learner is becoming fatigued, it suggests that the learner take a break. At this time, the task screen will forcibly change to a screen displaying the remaining break time. The learner will not be able to perform any operations on the task. To evaluate the validity of the break timing determined by the proposed system, the system was designed such that learners cannot refuse breaks.

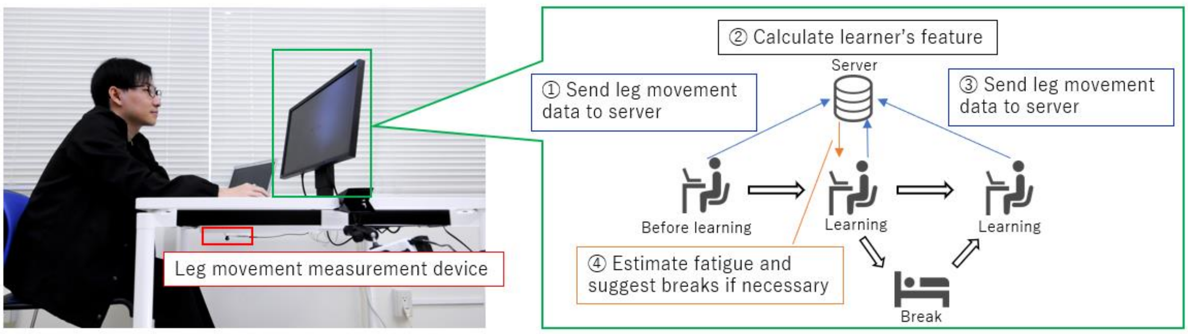


Figure 1: Position of leg movement measurement device (left) and system design (right)

3 PRELIMINARY EXPERIMENT

To make the proposed method operate ideally, it is necessary to investigate the relationship between the learner’s leg movements and the break timing. Therefore, a preliminary experiment was conducted in which mental arithmetic tasks were performed by six science and engineering university students. Here, the learners were able to take a break at any time. The mental arithmetic involved using mouse clicks to determine whether a formula displayed on a screen was correct or incorrect. Using the results, we analyzed the relationship between the break timing and leg movements.

This analysis found no relationship between the leg movements common to all learners and the break timing because there are individual differences in the leg movements. However, when focusing on individual learners, some instances where the timing when a learner’s accumulated fatigue coincided with a reduction in the rate at which they performed the mental arithmetic task correctly were observed. This observation suggests that it may be possible to predict and prevent a decline in performance due to fatigue without using learning results. In addition, by identifying the learner’s leg movements tendencies in advance and setting benchmark values for each individual learner, the system can suggest appropriate break timings for each learner.

Based on the data obtained in the preliminary experiment, we will determine the thresholds required for the system and complete the system. Then, we will conduct further experiments using the completed system to evaluate the appropriateness of the system’s break timings and analyze the learner’s performance. In terms of learner performance, we hope the proposed suggested break condition will be superior to the free break condition (where learners can take breaks at will).

ACKNOWLEDGMENTS

This work was partly supported by Grants-in-Aid for Scientific Research (NO. 21K18496 and 24K21451) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan.

REFERENCES

Aikawa, D., Asai, Y., & Egi, H. (2019). Proposing an estimation method of mental fatigue by measuring learner's leg movement. In The 21st International Conference on Human-Computer Interaction (HCI International 2019). 227-236. https://doi.org/10.1007/978-3-030-21814-0_17

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In Handbook of Research on Learning and Instruction (2nd ed.). London: Routledge, 522-560.

Ramachandran, A., Huang, C., & Scassellati, B. (2017). Give me a break! Personalized timing strategies to promote learning in robot-child tutoring. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17). 146–155. https://doi.org/10.1145/2909824.3020209

Sharma, K., Papamitsiou, Z., Olsen, K. J., & Giannakos, M. (2020). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20). 480–489. https://doi.org/10.1145/3375462.3375498

Enhancing E-Book Learning Dashboards with GPT-Assisted Page Grouping and Adaptive Navigation Link Visualization

Min Lu[1], **Boxuan Ma**[2], **Xuewang Geng**[2], **Masanori Yamada**[2]
Akita University, Japan; Kyushu University, Japan
lu@ie.akita-u.ac.jp, boxuan@artsci.kyushu-u.ac.jp, {geng, mark}@mark-lab.net

ABSTRACT: This research presents an enhanced e-book learning dashboard that supports meta-cognition and self-regulated learning by improving the visualization of navigation patterns in a teaching slide. Key enhancements include a GPT-assisted page grouping technique, which organizes pages into a two-level hierarchical structure, and a page link aggregation method that reduces visual clutter by focusing on significant navigation paths. These improvements enable more meaningful visualizations that allow students to track and compare their learning progress with class norms. Implemented with Python, Flask, and D3.js, the system offers an interactive experience for exploring learning data.

Keywords: learning dashboard, educational log data analytics, GPT-assisted page grouping

1 INTRODUCTION

E-books have become integral to modern educational settings, offering flexibility and accessibility to learners. By extracting insights on engagement and progress from e-book operational log data, learning dashboards support learners in metacognition and self-regulated learning. Our previous research (Lu et al., 2020) developed ReadingPath, which visualizes the navigation links between teaching slide pages as nodes, each annotated with statistical information such as viewing time and number of annotations. The dashboard displays both an overall class view and individual user views side by side, facilitating comparison and aiding students in self-monitoring their learning strategies.

However, the current visualization primarily presents statistical data on student interactions without linking this data to the actual content. While frequent interaction behaviors can be identified, it is challenging to comprehend the underlying reasons. Additionally, when a slide contains many pages, the connections between nodes become cluttered (Figure 2(1)), making it challenging to extract meaningful information. Recent work (Ma et al., 2022) has identified that page jump behaviors depend on the semantic and contextual aspects of e-book pages and has discovered several page categories responsible for these jumps. However, slide pages are categorized by human experts, which is labor-intensive. With the increasing role of AI systems in educational practices, there is potential to use AI tools like ChatGPT to categorize pages and provide more actionable insights.

This research aims to enhance the dashboard to offer users more meaningful visualizations by combining the actual content of pages and student interactions. We propose two methods: (1) grouping pages based on the internal structure of the slide to reduce the number of nodes and (2) aggregating links based on page groups while omitting less meaningful connections. We will introduce the conceptual design and methodology of the enhanced ReadingPath, along with initial visualization demonstrations using actual data.

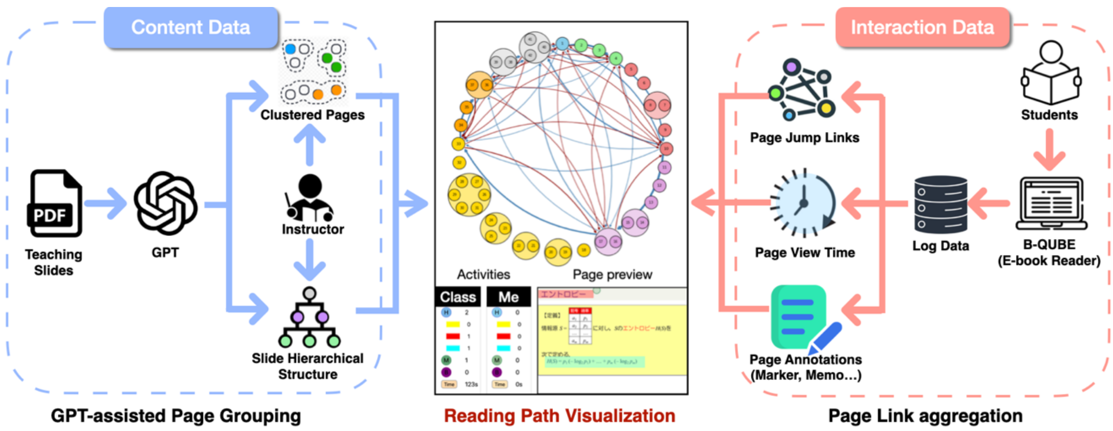


Figure 1: System architecture of the E-book learning dashboard

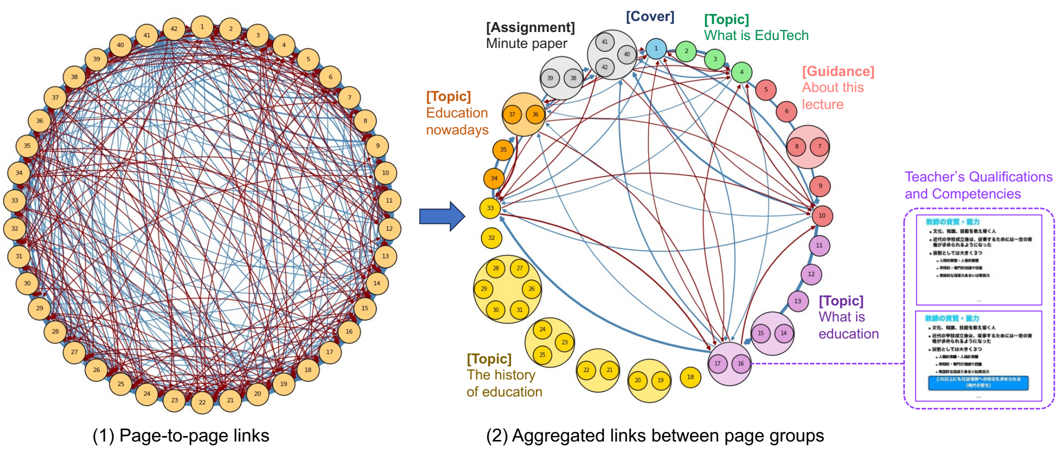


Figure 2: Page jumping links visualized in ReadingPath before and after aggregation

2 METHODOLOGY

The system architecture shown in Figure 1 integrates instructional content and student interaction data to create a comprehensive visualization. It consists of two modules: GPT-assisted Page Grouping and Page Link Aggregation. As students navigate the pages in the e-book system, their operational log data is recorded in the database. The page access logs are synthesized to links between pages, as shown in Figure 2(1). Such links cannot construct an effective cognitive map because of the cluttered view and lack of contextual information cues. To help students figure out important pages with more meaningful links, page categorizing and grouping are performed with the assistance of GPT.

Teaching materials (slides in PDF format) are processed through GPT to cluster pages and categorize them into distinct types. The generated slide types are then inspected and adjusted by the instructor. With the adjusted slide types, the pages are grouped based on topics and content similarity with GPT to form a hierarchical structure. For visualization purposes, we perform a two-level grouping. Ideally, the first level discovers broader sections like guidance, assignments, specific topics of teaching content, and so on. The second level is expected to group similar pages under the same topic. The links between pages are then aggregated into links jumping between groups. As demonstrated in Figure 2(2), the two-level grouping is visualized by coloring (the first level) and node aggregating (the second level). The jumping links between pages are colored arrows (blue: forward, red: backward). The scale of arrow thicknesses indicates the number of the same jump. ReadingPath also displays page viewing times and annotation counts, helping students track their progress and compare to class norms.

3 IMPLEMENTATION AND EVALUATION PLAN

The process of page grouping will be implemented as a web page that allows instructors to adjust the keywords and parameters of prompts to call GPT APIs. The instructors can make a final adjustment to the grouping with a drag-and-drop interface. The log query and preprocessing, including the aggregation of links, are implemented with Python. The visualization is implemented with D3.js, embedded into a web page developed with Flask, creating an interactive environment where users can dynamically explore their learning data. We plan to evaluate the system through user studies involving instructors and learners. The instructor who created the slides will test the effectiveness and usability of page grouping. Students using the enhanced ReadingPath and a control group using the original ReadingPath will engage with the system in actual lectures. The evaluation will focus on metrics such as user engagement, learning outcomes, meta-cognitive awareness, and user satisfaction.

4 DISCUSSION AND CONCLUSION

This poster presents in-progress research of an enhanced learning dashboard that integrates learning content structure and user operation logs of e-books, offering more understandable visualizations and actionable insights. Although the target users are learners, we consider the visualization informative to instructors. The slide's linear and hierarchical structure inherited from the presentation forms can obscure the inherent connections between pages. Page grouping effectively reflects the instructor's intended structuring of the slide, like a concept map (Wang & Walker, 2021). In contrast, the jump-link can reflect how the learners received the structure in actual learning activities. It adds another dimension to understanding the content and provides instructors with insights for future improvements. Future work will focus on improving the workflow of page grouping, conducting larger-scale user studies, and enhancing the system's interoperability with other educational tools. Additionally, we will explore customizable features to allow instructors to adjust the metrics of page grouping and link aggregation, thereby tailoring the dashboard to meet their teaching preferences.

ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Numbers JP21K18134, JP22H00552, JP23K11362, and JP24K20903.

REFERENCES

Lu, M., Chen, L., Goda, Y., Shimada, A., & Yamada, M. (2020, March). Development of a learning dashboard prototype supporting meta-cognition for students. In Companion proceedings of the 10th international conference on learning analytics & knowledge (LAK20) (pp. 104-106).

Ma, B., Lu, M., Taniguchi, Y., & Konomi, S. I. (2022). Exploring jump back behavior patterns and reasons in e-book system. Smart Learning Environments, 9, 1-23.

Wang, S., & Walker, E. (2021, May). Providing adaptive feedback in concept mapping to improve reading comprehension. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-11).

Multimodal collaboration analytics to explore moments of collaborative knowledge-construction

Author: Rachelle Esterhazy
Oslo Metropolitan University
Rachelle.esterhazy@oslomet.no

Author(s): Hugh Alexander von Arnim and Crina Damşa
University of Oslo
hughav@imv.uio.no; crina.damsa@iped.uio.no

ABSTRACT: This poster presents a proof-of-concept study exploring how multimodal collaboration analytics can help identify key moments of collaborative knowledge-construction of student groups in interprofessional education. The empirical setting involves students from various health professions working together on a simulated patient case to practice interprofessional collaboration. Discussing key moments of collaborative knowledge-construction during a debriefing session after the group work can facilitate the students' learning. However, selecting relevant key moments from group work based on memory is a challenging task that is prone to bias. Addressing this challenge, this study uses multimodal collaboration analytics combining video and audio data from three groups to explore which audible and visual behaviours indicate moments of collaborative knowledge-construction. Students' audible utterances are coded based on the TOCK-IP. Video data is coded for visible student behaviour such as facing peers, open position or use of hands. The resulting codes are analysed with Epistemic Network Analysis (ENA) to identify changes in epistemic networks that may indicate intensified knowledge construction during certain episodes of group discussion. The integrated results are used to create qualitative descriptions of key moments of knowledge-construction that can be used as basis for feedback on collaboration during debriefing sessions with the involved students.

Keywords: Multimodal collaboration analytics; Knowledge co-construction; Epistemic Network Analysis; Video analysis; Interprofessional learning

1 INTRODUCTION

This proof-of-concept study explores how multimodal collaboration analytics may help us identify moments of collaborative knowledge-construction in interprofessional group work. The ability to integrate knowledge from different disciplines has become an integral part of higher education (Kidron & Kali, 2015). One field that is particularly dependent on this skill is health care education, where students are confronted with complex problems that can only be solved in collaboration with other professions (O'Keefe et al., 2017). To prepare students for this collaborative challenge, many health care study programs have developed specific interprofessional education (IPE) opportunities. IPE typically entails learning scenarios in which students from two or more professions learn about, from and with each other by working on a patient case. The learning aims of IPE scenarios are often focused on the collaboration skills that students need to develop (Hinyard et al., 2019). To be able to collaborate effectively on a patient case, students need to agree on a common understanding of the health issues at hand and how they may be addressed through an integrated interdisciplinary approach. This collaborative knowledge construction takes mostly place on a verbal plane, as students co-construct meaning by discussing different perspectives from their professions (Floren et al., 2021). At the same time, these verbal exchanges are accompanied by non-verbal behaviour. As is typical during prolonged group discussions, students in IPE scenarios engage in different verbal behaviour with moments of group talk, work in silence or off-task behaviour (Nasir et al., 2022). While collaborative knowledge-construction is a continuous process, we argue that there are certain key moments during discussions in which this learning is driven forward by students articulating their perspectives, making connections, and reaching integrative insights. Previous studies indicate that such moments of collaborative knowledge-construction are often accompanied by certain visible behaviours in the group (Schneider and Bryant, 2022). Based on these assumptions, we define key moments of collaborative knowledge-construction as episodes, in which students share and negotiate knowledge from different disciplines while showing signs of embodied engagement in the discussion. As is common in IPE, recalling and discussing such key moments from previous group work is a powerful pedagogical tool to help students become more aware and advance their collaboration skills. However, selecting relevant key moments from group work based on memory is a challenging task for students and teachers that is prone to bias and fragmentation. This challenge can be alleviated by using data-driven approaches based on multimodal data to identify and describe relevant key moments in debriefing sessions. As a first step to prepare ground for this pedagogical application of multimodal collaboration analytics, this proof-of-concept study aims to explore how an integrated analysis of auditive and visual modalities can be used to study collaborative knowledge-construction during IPE group work. We ask: RQ1: Which audible behaviour indicates knowledge-construction and how does it unfold over time?; RQ2: Which visible behaviour indicates embodied engagement in the discussion and how does it unfold over time?; RQ3: How can differences in audible and visible behaviour be used to identify and describe key moments of interdisciplinary knowledge-construction during group work?

2 METHODOLOGY

Our data set includes both audio and video data from meetings of three IPE groups of 4-5 students from different health profession programs assessing a patient to develop a treatment plan. All meetings are video recorded and transcribed. Transcripts are coded with an adapted version of TOCK-IP (Tool to Observe the Construction of Knowledge in Interprofessional teams) (Floren et al., 2021) (e.g. SHARING=share professional assessment). For each utterance, the group's visual behaviour is coded, identifying whether students face their peers, use their hands or have an open posture (e.g. OPEN POSTURE=sitting back arms uncrossed). Interrater reliability is currently analysed and will be completed by the time of presentation. We then segmented the transcript into differently sized episodes based on the starting time stamps, e.g. segments every 10 min resulting in 5 episodes for a 56h 56s meeting. We did the same respectively, for 5 min, 1 min, 30sec episodes. The constellation of different audible and visible behaviour are then visualised as an epistemic network for each episode (Schaffer et al., 2016). We hypothesize that the shape of epistemic networks during episodes, in which students contribute frequently to the groups' knowledge-construction, will look different than networks during episodes of reduced or absent knowledge-construction talk.

3 PRELIMINARY RESULTS AND SIGNIFICANCE

Initial analyses indicate that there are significant changes in epistemic networks between episodes in IPE group meetings. Figure 1 shows an example of ENA from one meeting divided into five 10-minute episodes. The differences between centrality and density of the different networks indicate that the group's behaviours changed especially from episodes 11-20 min to 31-40min. Analyses are ongoing and will be expanded to all groups and tested with different time segmentations. We will also provide qualitative descriptions of the groups' behaviour in those episodes before and after a significant shift in ENA to understand what causes the change in epistemic networks and to identify which aspects of these networks may be interpreted as reflecting key moments of intensified knowledge-construction. Our study contributes to the empirical understanding of the multimodal collaboration behaviour of students that engage in IPE settings. Further, this proof-of-concept study outlines a methodological approach that allows us to identify key episodes during long group meetings that are relevant for research and pedagogical purposes.

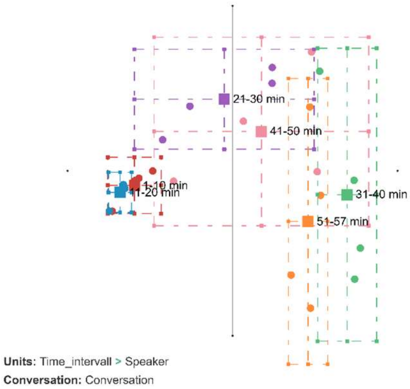


Figure 1: Epistemic network based on one group meeting

4 REFERENCES

Floren, L. C., ten Cate, O., Irby, D. M., & O'Brien, B. C. (2021). An interaction analysis model to study knowledge construction in interprofessional education: Proof of concept. Journal of Interprofessional Care, 35(5), 736–743.

Hinyard, L., Toomey, E., Eliot, K., & Breitbach, A. (2019). Student Perceptions of Collaboration Skills in an Interprofessional Context: Development and Initial Validation of the Self-Assessed Collaboration Skills Instrument. Evaluation & the Health Professions, 42(4), 450–472. https://doi.org/10.1177/0163278717752438

Kidron, A., & Kali, Y. (2015). Boundary breaking for interdisciplinary learning. Research in Learning Technology, 23. https://doi.org/10.3402/rlt.v23.26496

Nasir, J., Abderrahim, M., Kothiyal, A., & Dillenbourg, P. (2022). Temporal pathways to learning: How learning emerges in an open-ended collaborative activity. Computers and Education: Artificial Intelligence, 3. Scopus. https://doi.org/10.1016/j.caeai.2022.100093

O'Keefe, M., Henderson, A., & Chick, R. (2017). Defining a set of common interprofessional learning competencies for health profession students. Medical Teacher, 39(5), 463–468. https://doi.org/10.1080/ 0142159X.2017.1300246

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

Schneider, B., & Bryant, T. (2024). Using Mobile Dual Eye-Tracking to Capture Cycles of Collaboration and Cooperation in Co-located Dyads. Cognition and Instruction, 42(1), 26–55. https://doi.org/10.1080/07370008.2022.2157418

Educator - AI Partnership: Paving the Way to Sound Learning Design

Blaženka Divjak, Josipa Bađari, Darko Grabar, Barbi Svetec and Petra Vondra
University of Zagreb Faculty of Organization and Informatics
blazenka.divjak@foi.hr

ABSTRACT: The poster presents the preliminary research results related to the piloting of an AI assistant functionality in a well-established, collaborative online learning design (LD) tool - BDP. The research included three phases of the design cycle. In the first phase, we conducted problem investigation within several international projects, revealing the potential benefit of introducing an AI assistant as an innovative functionality of the LD tool. In the second phase, we upgraded the LD tool with an AI assistant, with an incremental approach to agile software design. Finally, we validated the upgrade with more than 30 courses, using the AI assistant to support higher education teachers to rapidly generate, customize, and optimize their LDs and interpret learning (design) analytics.

Keywords: learning design, learning analytics, design analytics, artificial intelligence

1 INTRODUCTION AND BACKGROUND

The advent of generative AI (GenAI) poses a substantial challenge but also provides opportunities for educators (van den Berg, 2024). GenAI has a transformative potential for education (Cooper, 2023), including support for teachers to make more informed decisions during instructional design (Muljana & Luo, 2021). Many studies propose how teachers and AI could collaborate to offer more personalized instruction (Holmes et al., 2023). Also (Choi et al., 2024) demonstrate the potential of ChatGPT to decrease instructors' workload and address challenges in course design, while stressing the importance of the educators' domain knowledge. The poster will present an approach to constructive educator-AI partnership in learning design (LD), enabling educators to exploit advantages and mitigate limitations of AI, through a meaningful delegation of tasks, roles, responsibilities, and autonomy (Kim, 2024). Considering the recognized benefits of LD in enhancing teaching and learning and assisting educators (Divjak et al., 2022), a concept and a web-based tool supporting the development of sound LD were launched in 2021 – Balanced Design Planning (BDP) tool. After a few years of intensive development of the tool and its use (> 2.000 users worldwide) it became evident that many educators would benefit from quality AI assistance, especially regarding the design of learning outcomes, ensuring constructive alignment, and interpreting design analytics (Divjak et al., 2023). In response, the latest developments of the tool have been steered towards enabling constructive educator-AI partnership and providing educators with real-time AI assistance in the LD process.

2 RESEARCH DESIGN AND RESULTS

The development of the BDP LD concept and tool (Divjak et al, 2022) has been done in line with the design science cycle, with an incremental approach to the development of the tool's functionalities. When it comes to enriching the tool with an AI assistant using GenAI, the research and development process has also included three phases of the design cycle: problem investigation, treatment design, treatment validation. The poster will present the process of design of the AI assistant and include preliminary results of its validation, which will contribute to answering the broader research questions (RQ): RQ1: How do educators accept and use AI assistance in LD development?; RQ2: What aspects of LD can be improved with AI assistance, and what is the human-only territory?

Problem investigation: determining the requirements. The development of the BDP tool has included several cycles of identifying user requirements and validating upgrades. Within the *Innovating Learning Design in HE* (iLed) project a survey was conducted in 2023, with 53 educators from European HEIs, eliciting the most important needs of educators in LD (Divjak et al., 2023). Some educators would appreciate a chatbot supporting the formulation of learning outcomes (LOs), others needed support in ensuring constructive alignment and interpreting design analytics. Some asked for real-time guidance and examples. These findings inspired the introduction of the beta version of the AI assistant, which was subsequently presented to three focus groups in 2024, within the *Trustworthy Learning Analytics and AI for Sound LD* project. The focus groups included 18 international experts from Europe and Africa, who appreciated the beta version of the AI assistant and gave suggestions. This primarily included AI interpretation of design analytics, implementation of innovative pedagogies, content creation, and collection of research data on the use of an AI assistant in the development of LD.

Treatment design: technical upgrade of the LD tool powered by AI (beta version). The upgrade was guided by the following principles: **A. Context-based prompts** were pre-prepared for different levels of the LD process: LOs, topics, units, TLAs. They were designed in a way to enable educators' preferences and edits. Educators can specify the number and characteristics of LOs/topics/units/TLAs to be generated, as well as the intended pedagogical approach, learning types, sequence and volume of AI-assisted LD-elements. **B. Educators' interventions** are encouraged by providing multi-layered AI-responses. Following the given prompt, the AI-response includes: i/Rationale (reasoning behind the provided output to the given prompt), ii/Output (AI-generated content related to the given prompt) and iii/Disclaimer (reminder to educators of the importance of their role as critical thinkers and responsible designers).

Treatment validation: testing on 30 courses. After pre-testing on several selected courses, the AI assistant was validated on the sample of 30 courses. Course teachers prepared LDs without the use of AI. After that, they used the AI assistant to re-create the LDs in partnership with AI, taking a critical look at AI's suggestions. The teachers' interactions with the AI assistant are being used to evaluate the functionality in two ways. **A. Feedback** is automatically collected after each interaction with AI (numerical rating and comments by users). This refers, first, to the quality of the AI-generated output, as the basis for further improvements of prompts. Second, it serves as a testing playground, enabling developers to check the technical readiness of the upgrade, meaningfulness of prompts and outputs, and usability of outputs. The preliminary findings show that most of the available pre-prepared prompts were meaningful and useful in LD, saving time and enabling creativity, customization, and optimization of LDs. Some complex prompts including simultaneous generation of topics, units and TLAs resulted in extensive and mostly irrelevant content, indicating that prompts should be more specific, shorter and follow the logic of the design process. In conclusion, AI-assistant can be used to co-create LD content when pre-prepared prompts are complemented by educators' domain and pedagogical knowledge. The preliminary findings led to modifications of prompts, contributing to better quality of AI generated outputs. Insights from this research are valuable for guiding the tool designers in further testing and upgrades. Currently, teachers' feedback is being collected via survey including Unified Theory of Acceptance and Use of Technology (UTAUT) constructs. **B. Comparison** of the two versions of LDs (developed without AI vs. with AI) of the 30 courses will be done in the next phase. Quantitatively, data and metadata stored in the tool will be used to analyze the essential features of LD (e.g., learning types, student workload) and its consistency (e.g., constructive alignment, coverage of LOs). Qualitatively, LDs will be inspected to establish differences in terms of pedagogical innovations, and feedback will be collected from the teachers.

Finally, **real-time learning analytics** are provided to educators in the tool, with analyses on the use of AI to generate LD. The aggregated analyses are used as evidence for further tool improvements.

3 CONCLUSIONS AND FUTURE WORK

So far, the AI assistant has been piloted in the LD of 30 HE courses in different disciplines. Feedback has been continuously collected, including real-time feedback provided in the design process and the follow-up feedback from educators. A comparison of the quality of LDs prepared by educators and LD prepared through educator-AI partnership has been done. The preliminary findings based on these data point that constructive educator-AI partnership in LD can be fruitful when educators pave the way and consider AI's abilities and limitations in terms of accuracy, biases and reliability. The AI assistant can successfully generate LD content based on meaningful prompts; however, educators should not over-rely on AI, but act as critical thinkers and domain experts who validate AI-generated outputs. Educators' experience with social aspects of learning, and their understanding of the learning context, are crucial for a constructive human-machine partnership. In further work, to enable better analysis of LDs, we will upgrade the BDP tool with an option for AI-generated interpretation of design analytics. Such interpretations should help educators to better understand analyses and possible implications as bases for data-informed decision-making, and "close the loop" by improving their courses accordingly. The idea is to keep educators in the loop in all stages of LD. While educators and AI can compose together, the role of the educator as the orchestra conductor remains central.

ACKNOWLEDGEMENT

This work has been supported by projects iLed (Erasmus+, 2022-1-HR01-KA220-HED-000085818) and TRUELA (Croatian Science Foundation, IP-2022-10-2854).

REFERENCES

Choi, G. W., Kim, S. H., Lee, D., & Moon, J. (2024). Utilizing Generative AI for Instructional Design: Exploring Strengths, Weaknesses, Opportunities, and Threats. TechTrends, 68(4), 832–844.

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. Journal of Science Education and Technology, 32(3), 444–452.

Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2022). Balanced Learning Design Planning. Journal of Information and Organizational Sciences, 46(2), 361–375.

Divjak, B., Rienties, B., Bađari, J., Grabar, D., Horvat, D., & Vondra, P. (2023). Enhancing Learning Design through User Experience Research: Insights from a Survey in Four European Countries. Proceedings of the CECIIS conference, 213–221.

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.

Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In Data ethics : building trust : how digital technologies can serve humanity (pp. 621–653). Globethics Publications.

Kim, J. (2024). Leading teachers' perspective on teacher-AI collaboration in education. Education and Information Technologies, 29(7), 8693–8724.

Muljana, P. S., & Luo, T. (2021). Utilizing learning analytics in course design: voices from instructional designers in higher education. Journal of Computing in Higher Education, 33(1), 206–234.

van den Berg, G. (2024). Generative AI and Educators: Partnering in Using Open Digital Content for Transforming Education. Open Praxis, 16(2), 130–141.

Developing a Feedback Analytics tool with Educators to Support Dialogic Feedback Processes

Yi-Shan Tsai[1], Hiruni Palihena[2], Bhagya Maheshi[3], Flora Jin[4], and Roberto Martinez-Maldonado[5]
Monash University
yi-shan.tsai@monash.edu[1]; hiruni.palihena@monash.edu[2];
bhagya.gangodagamachchige@monash.edu[3]; flora.Jin@monash.edu[4];
roberto.MartinezMaldonado@monash.edu[5]

ABSTRACT: Feedback is integral to learning as it helps learners self-regulate learning and move towards desired goals. Effective feedback requires both learners and educators to engage in a two-way, dialogic process where ideas are exchanged and meanings are negotiated. However, this has been a missing link in the feedback process due to the difficulty in tracking how learners interact with feedback. We adopted a design thinking approach to develop a learning-analytics-based tool with educators to address this gap. This poster presents a workshop-based study that identified five pain points educators experienced with dialogic feedback and eight design decisions to address these.

Keywords: feedback analytics, learning analytics, dialogic feedback, design thinking, PolyFeed

1 INTRODUCTION

Feedback is critical for improved learning and teaching. Without timely, high-quality feedback, students cannot effectively judge their progress or decide on steps to improve. Effective feedback involves a dialogic process where educators and learners engage in meaningful dialogue to exchange meanings and reconcile different perceptions, a process known as *dialogic feedback* (Yang & Carless, 2013). A key challenge to facilitate a dialogic feedback process is the lack of understanding about how learners use feedback, which hinders effective support for learners and continuous improvement on teaching quality. We propose a learning-analytics-based solution to enhance dialogic feedback by improving feedback quality and provide data-based insights into learner interactions with feedback. This poster presents a study that followed a design thinking process to develop a teacher-facing tool - PolyFeed, and showcases key functionalities that aim to enhance dialogic feedback processes. We aim to answer: *How can we design a learning-analytics-based tool to support dialogic feedback?*

2 BACKGROUND

Dialogic feedback is a two-way process where teachers and learners engage in synchronous or asynchronous conversations and social interactions where ideas are exchanged and meanings are negotiated. Yang and Carless (2013) identified three dimensions of effective dialogic feedback: cognitive, social-affective, and structural. The cognitive dimension considers the design of feedback to help students self-regulate learning and bridge the gap between current and desired performance. The social-affective dimension addresses social interactions, power dynamics, and trust. The structural dimension involves the organisation of feedback, such as timing and modes. Given limited resources, leveraging technologies to mitigate challenges in this dimension has been proposed (*ibid*.).

Learning analytics (LA) has demonstrated the promise of technologies to support feedback processes by providing (near) real-time feedback on learning. When used to track learners' interactions with feedback (also known as feedback analytics (Jin et al., 2024), LA can close the feedback loop by showing how learners make use of feedback, thus informing teaching.

3 METHOD

We employed a Design Thinking (DT) approach proposed by Prieto-Alvarez et al. (2018) to ensure PolyFeed addresses the needs of teachers in facilitating dialogic feedback. The approach includes four distinct stages made of seven components: 1) Understand – Empathise & Define, 2) Create – Ideate, 3) Deliver – Prototype & Test, and 4) Support – Involve & Sustain. We conducted the first three stages iteratively, involving interviews, surveys, user journey maps, and workshops over three years. This poster focuses on presenting the results of a workshop and a validation activity conducted in 2024 that allowed us to iterate one cycle of the first three stages, thus moving from a low-fidelity prototype to a high-fidelity prototype and finally creating a functional prototype.

Before the workshop, eight participants filled in a survey with three open-ended questions asking about their experience with dialogic feedback, encountered challenges, and perceived enablers (*DT – Empathise*). Six of these participants joined a subsequent one-hour-long workshop. All of them came from the Information Technology discipline and have at least one year of teaching experience. During the workshop, we shared the analysed results of the survey with the participants (*DT – Define*), and then invited them to brainstorm features and functionalities they expect in a LA tool that can support dialogic feedback processes (*DT – Ideate*). Both the survey and workshop discussions were analysed thematically in an inductive manner to identify user requirements. Based on the results, a high-fidelity prototype was developed. It was subsequently tested with seven experienced educators who were invited to share suggestions to improve the tool design before we developed a functional prototype.

4 RESULTS

Based on the responses from the workshop participants, we identified five pain points (PP) and eight design decisions (DD) to address these.

PP1: Difficulty in ensuring feedback consistency. This challenge was particularly prominent when multiple educators of varying experience in feedback teach a large class together. **PP2: Difficulty in balancing time constraints with feedback quality.** With the increase in high student-to-tutor ratios, the educators struggled to personalise feedback.
DD1: A Chrome browser extension that activates alongside the learning management platform used by the educators when marking and providing feedback to ease the feedback process. **DD2:** A template repository that supports educators in developing feedback following effective feedback principles, and allows educators to create and save their own feedback templates.
PP3: Difficulty in knowing how feedback is received and used by students. This PP reveals a gap in understanding between students and teachers that impedes dialogic feedback processes.

DD3: A mini dashboard visualising the number of reflective notes and action plans students created.
PP4: Challenges with the effective adoption of new technologies to assist feedback processes. The educators expressed pedagogical and ethical concerns regarding feedback automation.
DD4: A machine-learning model building on learner-centred feedback principles (Ryan et al., 2023) was incorporated to help educators identify the alignment of their feedback with these principles. **DD5:** Providing suggestions to include the missing learner-centred feedback elements in feedback. **DD6:** Using GenAI to assist in constructing personalised feedback based on the suggestions above.
PP5: Difficulty in tracking the effectiveness of feedback. Different from PP1, which concerns students' learning, PP5 is about tracking feedback effectiveness for professional development.
DD7: A web-based dashboard visualising 1) students' ratings on the feedback from educators and 2) patterns of feedback alignment with learner-centred feedback principles (*ibid.*). **DD8:** Feedback analytics allowing educators to monitor students' feedback interactions, including identified strengths, weaknesses, and confusions, at class and group levels

Based on the eight design decisions, we developed a high-fidelity prototype and validated it with seven educators, resulting in further improvements to be included in a functional prototype (see snippets of the tool and the associated design decisions here, and a demo video here), such as allowing the extension to be active on Google Documents, simplifying visualisations, and allowing teachers' inputs into GenAI to improve feedback alignment with learner-centred feedback principles.

5 CONCLUSION AND FUTURE WORK

In this poster, we present one iteration of the "understand", "create", and "deliver" stages of design thinking when designing a feedback analytics tool with educators. Through a design workshop, we obtained a better understanding of the pain points educators experienced when facilitating dialogic feedback, thus defining the user requirements, identifying key features and functionalities, thereby informing prototype development. Although this poster focuses on presenting the teacher-facing tool, it is important to note that the tool needs to work in tandem with the student-facing PolyFeed (see demo) to effectively facilitate dialogic feedback. For example, DD8 relies on the data capture of the learner's interactions with PolyFeed to facilitate two-way feedback. Our future work seeks to pilot the functional tool and explore how feedback analytics may effectively enhance dialogic feedback and support educators in developing feedback literacy.

REFERENCES

Jin, F., Maheshi, B., Martinez-Maldonado, R., Gašević, D., & Tsai, Y.-S. (2024). Scaffolding Feedback Literacy: Designing a Feedback Analytics Tool with Students. *Journal of Learning Analytics*,

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. (2018). Co-designing learning analytics tools with learners. In J. Lodge, J. C. Horvath, & L. Corrin (Eds.), *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers*. Routledge.

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2023). Identifying the components of effective learner-centred feedback information. *Teaching in Higher Education*, *28*(7), 1565–1582.

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, *18*(3), 285–297.

Eval-QUEST: Analyzing Real-time Student Questions and Peer Evaluations to Enhance Their Question Behaviors

Saki Inoue, and Kazutoshi Sumiya
Kwansei Gakuin University
huk50102@kwansei.ac.jp, sumiya@kwansei.ac.jp
Yuanyuan Wang
Yamaguchi University, y.wang@yamaguchi-u.ac.jp
Shizuka Shirai
Osaka University, shirai@ime.cmc.osaka-u.ac.jp
Yukiko Kawai
Kyoto Sangyo University, Osaka University, kawai@cc.kyoto-su.ac.jp

ABSTRACT: This study aims to enhance students' questioning behavior during university lectures by introducing "Eval-QUEST," a system that facilitates real-time question sharing and peer evaluation through a bot. In Japan, students often hesitate to ask questions in class. Our approach utilizes real-time question posting, sharing, and peer evaluation to create a supportive atmosphere that encourages students to engage more actively, thereby improving both the quality and quantity of their questions. This paper discusses implementation issues and analyzes the data collected on questions and student reactions. The results indicated that questions tended to cluster within specific time intervals, with those posted later in the lecture receiving lower evaluations. Additionally, questions that were more specific and relevant to the lecture topics were rated more highly.

Keywords: Blended learning, real-time questions, peer evaluation, question behavior

1 INTRODUCTION

Encouraging active participation in university classes is a critical issue across disciplines. In Japanese educational settings, students are often hesitant to ask questions during class, opting instead to raise their concerns after class or during designated times. This practice limits real-time interaction and reduces opportunities for immediate question-asking. To address this issue, this study developed a question bot aimed at streamlining the process of posting questions. Automated tools, such as bots, have been demonstrated to provide excellent real-time learning support with high interactivity (Forden et al., 2023). Since the coronavirus pandemic, blended-learning approaches that combine online and face-to-face elements have become increasingly popular, leveraging the advantages of both methods (Harrak et al., 2019). To effectively foster deeper understanding through interactive learning, we provide an online question bot to support students attending face-to-face classes. This study presents "Eval-QUEST," a bot-based system integrated into the Slack instant-messaging tool to facilitate real-time question posting, sharing, and peer evaluation among students. The system leverages constructive peer pressure to foster solidarity, encouraging students to ask more and better-quality questions. Eval-QUEST was implemented in a first-year lecture at a Japanese university. This paper details the Bot's implementation using the Slack API, analyzes real-time question and peer-evaluation data, and highlights insights from the analytical results and student questionnaires.

2 METHODOLOGY AND SYSTEM IMPLEMENTATION

2.1 Question-Generation Support

This study introduces a system that allows students to record their questions as they arise during class. In traditional classroom settings, questions are typically addressed at designated times, which limits students' opportunities to express their doubts in real time. With this new system, students

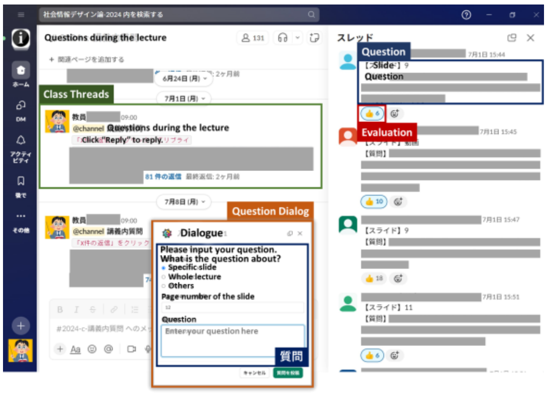


Figure 1: Real-time questions and question dialog

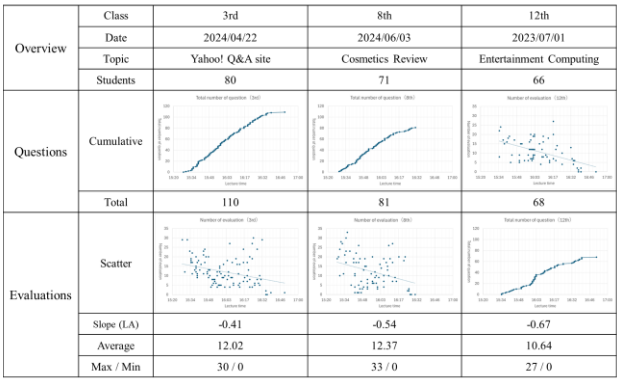


Figure 2: Analysis of real-time questions and peer evaluation

can easily articulate their questions as they come to mind, preventing them from being forgotten. This approach encourages continuous engagement in questioning, helping students develop a habit of active participation in class. To assist instructors in managing the question data and to make it easier for students to input and view questions, a dialog box was created using Slack's API. Further details about this dialog are provided in Section 2.2. Instructors can review real-time questions and evaluations after classes to prepare lecture materials for subsequent classes.

Students can share their questions in real time with their peers and then evaluate those questions. The evaluation process is simplified through a single "Good" reaction button, making it easy to assess questions and manage a high volume of inquiries. This evaluation feature introduces a healthy level of peer accountability, making students aware of how their peers respond. This is expected to enhance both the quality and quantity of questions, leading to more meaningful discussions. Additionally, evaluating their peers' questions helps students develop critical-thinking skills, allowing them to objectively reflect on their own questions, thereby contributing to overall question-quality improvement.

2.2 Implementing the Question Bot Using the Slack API

We implemented a Question Bot using Slack's API, owing to its ease of handling question input and reactions. The Question Bot operates with Slack's History and Event APIs and is supported by a dual-server configuration: Slack's standard server and a Reaction Server managed by the instructor. The Reaction Server facilitates a *question dialog* for students, allowing them to input questions and provide reactions. The *Question Bot* allows instructors to control the timing of question sessions as needed through the *Reaction Server*. Figure 1 shows the Slack interface displayed to students. When students access the *Question Bot*, the *question dialog* is initially displayed, containing instructions for use, as well as fields for "Question Target" and "Question Content." The "Question Target" can be selected from options such as "Slide," "Video," "Entire Lecture," or "Topic." For slides, students can specify a page number, while for topics, they can freely describe the specific content. Upon submission, each question is automatically shared with other students; the *Question Bot* adds the first reaction to encourage further interaction. Additionally, a confirmation message is sent to students, allowing them to immediately verify or delete their questions. Students were instructed to submit at least one question per class and to react to at least ten of their peers' questions, fostering a culture of active participation and peer evaluation.

3 ANALYSIS OF QUESTIONS AND PEER EVALUATION

This study analyzed changes in real-time questioning behavior by examining variations in the number of questions and evaluations submitted. Data were collected from the "Social Information Design" course, comprising a 60-minute lecture and a 40-minute Q&A session, during which students could submit questions and evaluations. Participants included 78 consenting second-year and higher students from the School of Policy Studies at Kwansei Gakuin University, with all data anonymized to ensure privacy. The analysis focused on sessions three, eight, and twelve of the twelve-session course.

Figure 2 provides an overview of each class session and summarizes our findings. We noted a slight decrease in both the number of questions and the number of participants as the course progressed. An analysis of the cumulative number of questions in each class revealed that questions tended to concentrate during specific time intervals. Additionally, as each class progressed, the number of questions generally decreased, with evaluations for questions posted later in the session receiving fewer ratings. However, a comparison of the maximum, minimum, and average evaluations across the three lectures showed no significant differences in question evaluations among the sessions.

To explore the criteria students used to evaluate their peers' questions, we analyzed the top 10 questions (5%) with the highest reaction scores and the bottom 10 questions (5%) with the lowest reaction scores from the twelfth lecture. For this analysis, we included only the questions recorded during the lecture portion, excluding those posted during the Q&A session. This resulted in a total of 52 questions analyzed. Sixteen questions were excluded because they were more likely to receive lower evaluations because of limited reaction time and a focus on the Q&A segment. Each question was assessed using two metrics: specificity and expandability, rated or not rated. Additionally, questions were classified as related to the lecture topic or directed to the speaker. The lead author conducted the initial evaluations and classifications, followed by verification and discussion with co-authors to ensure validity. After one discussion, we concluded that no revisions were necessary.

Of the 52 analyzed questions, 31 (60%) were specific, 15 (28%) expanded, 41 (78%) lecture-related, and 11 (21%) speaker-directed. In the top 10 questions, eight (80%) were specific, three (30%) expanded, and all (100%) were lecture-related. By contrast, in the bottom 10, five (50%) were specific, three (30%) expanded, five (50%) lecture-related, and five (50%) speaker-directed. These findings suggest that students favor questions with higher specificity and lecture relevance when evaluating their peers' questions.

4 CONCLUDING REMARKS

The analysis indicated that questions tended to cluster within specific time intervals, with those posted later in the lecture receiving lower evaluations. Additionally, questions that were more specific and closely related to the lecture topics generally received higher ratings. To measure the impact of the evaluation, based on the intervention of the tool, we conducted a questionnaire survey of the students. Although not detailed here because of space limitations, surveys conducted at the beginning and end of the course (n = 17) revealed that many students found real-time question-posting engaging. It was also observed that posting questions increased students' motivation to question, showing an improvement rate of 11.8%, while reactions to questions enhanced students' motivation to attend lectures, with an improvement rate of 8.8%.

REFERENCES

Forden, J., Gebhard, A., & Brylow, D. (2023). Experiences with TA-Bot in CS1. In *Proceedings of the ACM Conference on Global Computing Education (CompEd '23)* (pp. 57–63). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3576882.3617930

Harrak, F., Bouchet, F., & Luengo, V. (2019). From student questions to student profiles in a blended learning environment. *Journal of Learning Analytics, 6*(1), 54–84. https://doi.org/10.18608/jla.2019.61.4.

This work was supported by JSPS KAKENHI Grant Number JP24K21464 and MEXT as "Developing a Research Data Ecosystem for the Promotion of Data-Driven Science".

Exploring Students' Usage and Perception of a Goal Driven Dashboard

Aditya Joshi, Sergey Sosnovsky
Utrecht University, The Netherlands
{a.joshi, s.a.sosnovsky}@uu.nl

ABSTRACT: Learning Analytics Dashboards (LADs) aggregate educational data and display visualizations that inform students about their learning progress and promote Self-Regulated Learning. Students response to such guidance can be varied due to individual differences. Achievement Goal Orientation theory provides one of the perspectives that explain the impact these visualizations have on the engagement and learning outcome of students. We have developed a dashboard that can display students' progress differently depending on the selected course goal. This study explores how students' achievement goal orientation, employed prominent goal visualization, and self-reported preferences across dashboard designs relate to each other.

Keywords: Achievement Goal Orientation, Social Comparison, Self Regulated Learning

1 INTRODUCTION

Student-facing Learning Analytics Dashboards (LADs) have been found to support Self-Regulated Learning (SRL) and improve students' motivation, engagement and overall learning outcomes; however, studies have also found that these effects are not consistent for all students due to various factors such as personality traits, interests, prior knowledge, learning goals etc. (Matcha, 2019). Achievement Goal Orientation (AGO) Theory (Elliot & McGregor, 2001) explains that students' goals may significantly affect their response to the information about their progress, grades and the performance of their peers. AGO places the goals of an individual across the dimensions of Mastery vs. Performance and Approach vs. Avoidance. Mastery Approach (MAp) goals are related to striving to master the subject and learning as much as possible. Performance Approach (PAp) goals are more related to achieving end results in *comparison* to an explicit standard or other individuals. Mastery Avoidance (MAv) goals drive students to keep away from situations when feel incompetent. Performance Avoidance (PAv) goals make them try hard enough to get a passing grade and/or not be left behind compared to their peers. Thus, the same information on the learning progress provided by a LAD can trigger very different motivational response from students with different goal orientations.

We developed a dashboard that employs several variants of progress indicators that match students' goal orientation in order to investigate students' choices, preferences, and the engagement patterns across the four achievement goal orientations.

2 STUDY SETUP

StudyLens was introduced as a non-compulsory learning tool in an introductory Python programming course. It provided students with access to a large collection of practice material. The interface of

StudyLens is an LAD displaying student progress across all the topics. On clicking a topic tile, it expands to show access a collection of related interactive learning material adopted from (Brusilovsky, 2018). All student interactions with individual learning items are logged and used to compute the progress score for the relevant topic using the basic Elo rating system (Pelánek 2016). The progress indicators take one of the four different forms depending on the currently selected learning goal orientation (see Figure 1). Students could switch their learning goal at any time. Additionally, *StudyLens* reset their learning goals every week to facilitate the conscious choice of a learning goal and avoid a situation when a goal orientation is selected once and used without much thought from there on. The indicators for mastery-oriented goals display individual student's progress; while the performance-oriented indicators supplement student's progress with an additional progress bar showing average progress of all other students in the course for this topic. Indicators for avoidance goals show a warning sign below a passing score of 60% (and below average progress for "performance"). The indicators for approach goals show a checkmark upon reaching 90% (and performing better than the average for "performance)". At the end of the course, students were asked to fill-in a survey that combined the Achievement Goal Orientation Questionnaire (Elliot, 2001) and short set of questions asking them to rank the four dashboard interfaces based on their preferences.

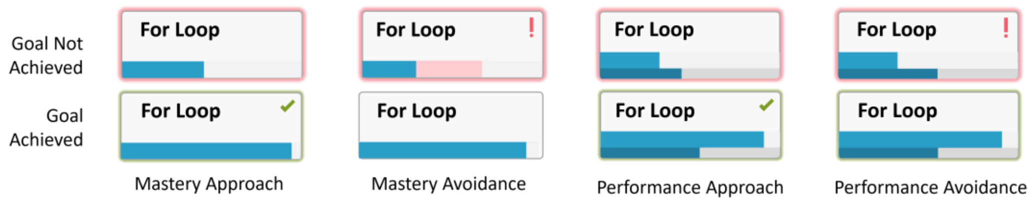


Figure 1: Progress Indicator Designs for AGO selection

3 DASHBOARD USAGE PATTERNS

Students averaged 10 sessions, opened 153 learning activities, and had at least 60% progress on the 15 course topics. Out of the 64 students who used the system regularly, 42 students completed the post-survey. Although students could change their goal at any time, most of them had a prominent learning goal, i.e. a goal selection that guided them to at least 50% of their interactions with learning material throughout the semester. Table 1 shows that the MAp goal was the most popular goal selection, followed by PAp, PAv, and MAv. All the students except one had a clear prominent goal.

Table 1: Distribution of Prominent Goal Selection

N-users	Goal Statement	Prominent Goal	MAp	MAv	PAp	PAv
36	Master the Subject	MAp	**4.40**	3.79	3.44	3.17
12	Perform Better than Others	PAp	**3.96**	3.54	3.79	3.75
10	Perform Just Like Others	PAv	3.76	3.57	3.38	**3.86**
6	Learn Just Enough	MAv	**4.22**	3.78	3.11	3.00

3.1 Self Reported AGO and Prominent Goal Selection

Students' self-reported scores on the four AGO subscales are mildly correlated with their prominent goal selections. Table 1 shows the mean subscale scores of the students with the prominent goal selection. Multinomial logistic regression was performed to create a model of the relationship between the AGO subscales, pre-knowledge, and the prominent goal selection (Pseudo-$R2$ = .28, p = 0.05). Coefficients indicate a strong positive effect of each AGO subscale on corresponding prominent goal selection and a negative effect from the opposite subscale, showing that students' AGO influences their goal choices and dashboard usage.

3.2 Prominent Goal Selection and Preference Ranking

The students ranked the four dashboard variants that correspond to each AGO. The students with the prominent goal = MAp ranked the respective dashboard variant the highest (n = 24, Mean Rank = 1.63). The students with the prominent goal = PAp ranked the PAp dashboard the highest (n = 8, Mean Rank = 1.75). The students with the prominent goal = PAv gave the highest ranks to Performance dashboard, but the one focused on approach, rather than avoidance (n = 7, Mean Rank = 1.86); the PAv dashboard was their second choice. Only two students with the MAv prominent goal have completed the survey; thus their responses were not considered.

3.3 Self Reported AGO and Preference Ranking

To assess whether students' own self-reported AGO correlates with their preferred dashboard variant, we conducted a correlation analysis between AGO subscale scores and rank scores, with a score of 3 to the top-rated variant, and so on. The AGO scores and the rank score of Mastery Approach, Mastery Avoidance, and Performance Avoidance interfaces were found to be moderately positively correlated with r = .31, p = .05; r = .15, p = 0.36; r = .22, p = 0.16 respectively. Performance Approach subscale was negatively correlated with the rank of the respective dashboard variant.

4 CONCLUSION AND FUTURE WORK

This study analyses students' interactions with a dashboard designed to match students' goals in a course. The results indicate that there is a relation between students' self-reported goal orientation, course goals, and the dashboard design corresponding to these goal orientations. This means that the students are more likely to engage with the LAD interface that reflects their learning goals. In the future, we plan to further investigate how LAD interface can be adapted to students' individual factors.

REFERENCES

Brusilovsky, P., Malmi, L., Hosseini, R., Guerra, J., Sirkiä, T., Pollari-Malmi, K.: An integrated practice system for learning programming in python: design and evaluation. Research and practice in technology enhanced learning 13, 1–40 (2018)

Elliot, A. J., & McGregor, H. A. (2001). A 2× 2 achievement goal framework. Journal of personality and social psychology, 80(3), 501.

Festinger L (1954). "A theory of social comparison processes". *Human Relations*. **7** (2): 117–140.

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.

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, *98*, 169-179.

False Feedback or no Feedback? Effects of Inaccurate AI-Feedback on Students' Writing

Author(s): 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: The development of automated scoring algorithms enables the immediate provision of performance-contingent feedback even on complex tasks such as writing. However, feedback based on Artificial Intelligence (AI) is fallible. To better understand the potential effects of implementing AI-based feedback, the present study investigates students' reactions to inaccurate AI-based feedback in the context of writing in secondary education: We compare the outcomes of students after receiving inaccurate feedback to accurate feedback and no feedback. A sample of N = 286 students was randomly assigned to receive either a) AI-based performance feedback regarding the fulfilment on set writing criteria or b) no feedback. Feedback accuracy (inaccurate or accurate) was determined post-hoc by trained human raters. We found differential effects of inaccurate feedback depending on initial performance. Students who initially met a criterion revised their text to the worse after receiving inaccurate (i.e., negative) feedback. Students who did not initially met a criterion benefit from inaccurate (i.e., confirmatory) feedback in a subsequent transfer task compared to accurate or no feedback, suggesting that even inaccurate AI-based feedback can promote learning. This underscores the educational value of integrating feedback and suggests areas for further research in instructional design and feedback strategies in learning contexts.

Keywords: Artificial Intelligence, Feedback, Formative Assessment, Writing, Secondary Education

1 OBJECTIVES AND BACKGROUND

Feedback based on Artificial Intelligence (AI) offers many advantages to supplement teaching and learning by providing students with real-time feedback even on complex tasks such as writing (Fleckenstein et al., 2023). Because the quality and credibility of feedback can influence learning and empirical studies demonstrate that uncertainty of automated feedback's accuracy may lead learners to reject the feedback (Bai & Hu, 2017; Lavolette et al., 2015), a number of studies aim to maximize the accuracy of automated writing feedback (Brand et al., 2020). While inaccurate feedback messages (i.e., feedback that does not mirror the previous performance correctly) may be less effective than accurate feedback (Lavolette et al., 2015), studies have not yet examined how students learn from inaccurate feedback on writing compared to accurate or no feedback. However, this research is crucial because AI enables automated feedback to be integrated into classrooms, but it is still fallible. Understanding the impact of inaccurate feedback is therefore essential for its effective use in education. This study addresses this gap by investigating the research question:

> How do students perform after receiving inaccurate feedback compared to accurate feedback or no feedback, considering their initial performance?

2 METHODS

A sample of $N = 286$ German 7th to 9th grade ($M_{age} = 13.52$, $SD_{age} = .98$) foreign language learners was asked to compose an email in English. The AI-based algorithm was trained in prior studies and used to provide performance feedback according to five criteria (*Content completeness, salutation and farewell, subject, introduction and closing sentence, language style*; Horbach et al., 2022). Students were randomly assigned to one of two experimental conditions. The feedback condition ($N = 190$) provided (potentially inaccurate) performance feedback for each criterion, presented as a salient red cross or green tick. In the control condition ($N = 96$) assessment criteria were presented without feedback (Figure 1).

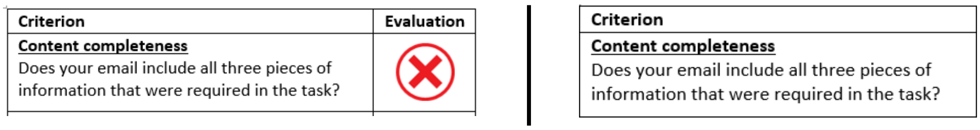


Figure 1: Automated feedback condition (left) and control (right) on one example criterion

All students were asked to revise their texts according to the criteria. Feedback accuracy was determined post-hoc by two independent raters who scored each text on five criteria, with an inter-annotator agreement (Cohen's kappa) ranging from $\kappa = .87$ to $\kappa = .97$. A third rater acted as adjudicator in cases where the first two raters disagreed. To assess writing task performance, we used the human scored performance, that is, the fulfilment of criteria in (1) text revision and (2) a second writing task (posttest). Performance in revision and posttest were each compared at criterion level between inaccurate feedback and a) accurate feedback and b) no feedback using linear mixed models (LMMs) with criterion and student as random effects. All LMMs include the interaction term feedback × initial fulfilment.

3 RESULTS

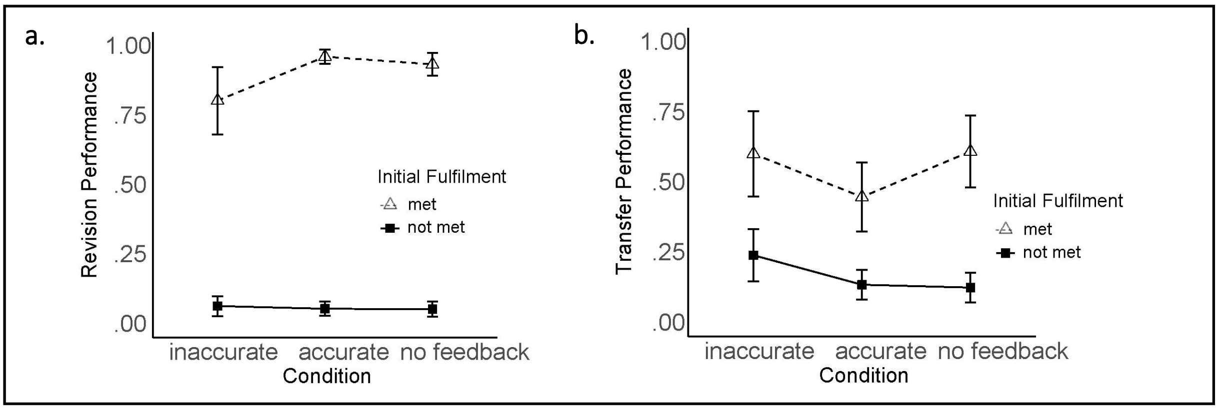


Figure 2: Estimates and standard errors of revision (2a) and posttest (2b) performance after receiving inaccurate, accurate and no feedback for initially met and not met criteria

The algorithm's 60% true positive and 95% true negative rates caused an imbalance of initial performance between the feedback conditions, with more fulfilled criteria in the inaccurate compared to accurate feedback condition ($t_{162} = -2.09$, $p = .038$). However, all models are controlling for prior knowledge and further, LMMs are known to be relatively robust towards unequal distributions (Schielzeth et al., 2020). Effects of inaccurate feedback were moderated by initial performance (i.e., whether the criterion was initially met or not). In text revision, only when the criterion was initially met, students performed significantly better after receiving accurate (i.e., confirmatory) feedback compared to receiving inaccurate (i.e., corrective) feedback (β = 1.81, p = .03; Figure 2a). In the transfer task, only when the criterion was initially not met, students performed significantly better after receiving inaccurate (i.e., confirmatory) feedback than after receiving accurate (i.e., corrective) feedback (β = -.01, p = .02) or no feedback (β = -.80, p = .02; Figure 2b).

4 SIGNIFICANCE

This study investigated the impact of inaccurate feedback in automated writing feedback. Overall, the effect of inaccurate feedback depended on a) whether students initially met the criterion or not (and therefore, whether the feedback was confirmatory or corrective), b) the performance outcome (revision or transfer), and c) what feedback condition it was compared to (i.e., accurate or no feedback). For text revision, inaccurately corrective feedback on initially met criteria may have led students to edit text passages that were already correct, resulting in a deterioration. In the transfer task, students performed better after receiving inaccurate feedback on criteria that were initially not met (i.e., confirmatory instead of corrective feedback). One attempt to explain this unexpected finding is the possibility that inaccurate feedback may be more salient, evoking affective-motivational processes that may lead to deeper engagement with the criterion and more effort in the transfer task. Our results suggest that even flawed AI-based feedback can positively affect young learners' implementation of the feedback compared to no feedback and highlight the educational value of enriching classroom settings through the implementation of AI-based feedback. Future research is needed to explore how learners internalize inaccurate feedback and the underlying mechanisms behind the effects of more or less accurate feedback, as well as the potential long-term effects of inaccurate feedback.

REFERENCES

Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, *37*(1), 67–81. https://doi.org/10.1080/01443410.2016.1223275

Brand, D., Novak, M. D., DiGennaro Reed, F. D., & Tortolero, S. A. (2020). Examining the Effects of Feedback Accuracy and Timing on Skill Acquisition. *Journal of Organizational Behavior Management*, *40*(1–2), 3–18. https://doi.org/10.1080/01608061.2020.1715319

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, *6*, 1162454. https://doi.org/10.3389/frai.2023.1162454

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. *Swedish Language Technology Conference and NLP4CALL*, 72–83.

Lavolette, E., Polio, C., & Kahng, J. (2015). The Accuracy of Computer-Assisted Feedback and Students' Responses to It. *Language Learning & Technology*, *19*(2), 50–68.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Analytics and Educational Data Visualization to Support Learning: A Case Study from Mexico

José Gerardo Moreno-Salinas
National Autonomous University of Mexico (UNAM)
gerardo_moreno@cuaed.unam.mx

Erik Carbajal-Degante
National Autonomous University of Mexico (UNAM)
erik_carbajal@cuaed.unam.mx

ABSTRACT: In today's educational landscape, the integration of digital technology and the abundance of data have revolutionized the fields of learning analytics and data visualization. These advancements have transformed our understanding and support of the teaching-learning process by harnessing complex educational data. Mexico ranks fourth globally in terms of registered Moodle sites. The Visual Learning Analytics (VLA) project aims to support distance learning professors at the National Autonomous University of Mexico (UNAM) in monitoring and evaluating student performance, as well as early identification of potential dropouts. The VLA project incorporates the five key elements of advanced visualizations to provide an effective and useful tool.

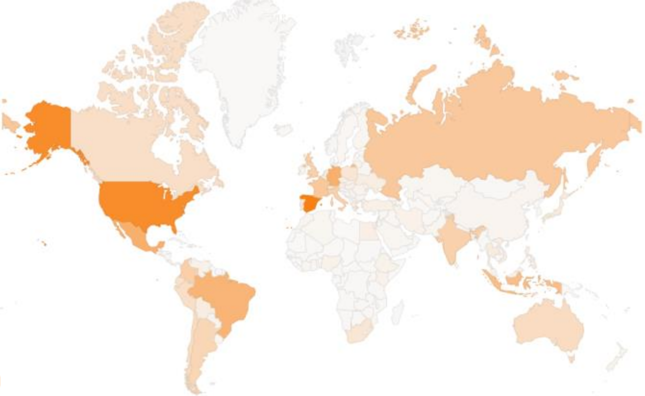
Keywords: Learning Analytics, Educational Data, Visualization, Moodle, Distance Education.

1 INTRODUCTION

The data generated in educational contexts is often large, complex, and heterogeneous, making it difficult to understand—even with advanced data analysis capabilities (Vieira et al., 2018). The Visual Learning Analytics (VLA) project leverages data from the learning management system (LMS) to create interactive dashboards. These dashboards provide professors, who may not have the capability to set up or install learning analytics and data visualization tools in Moodle, with a clear view of key metrics such as student participation, collaboration, progress, resource usage, and grades. The implementation of VLA aims not only to monitor and assess student performance but also to enable early identification of those at risk of dropping out, addressing a critical issue in distance education (Shaikh & Asif, 2022).

1.1 Moodle in Mexico

Moodle is an attractive LMS choice for public universities, as it is open-source, flexible, customizable, and free. Mexico is the fourth-largest user of Moodle worldwide (Figure 1), with 8,195 registered sites (Moodle Statistics, 2024).



Country	Registered sites
Spain	13,628
United States	12,165
Germany	8,201
Mexico	8,195
Brazil	7,599
Indonesia	6,515
Russian Federation	5,397
France	5,393
Colombia	4,558
United Kingdom	4,435

Figure 1: Moodle map and top 10 countries by registration

According to the first census of course management tools and digital repositories for learning at UNAM, 86% of schools and faculties use Moodle (González-Videgaray et al., 2017). Despite the various plugins available for learning analytics and data visualization in Moodle, different versions and access restrictions can hinder their effective implementation. This directly impacts university professors, as tracking and identifying students at risk of dropping out adds to their daily advisory responsibilities, creating an additional burden.

1.2 Learning Analytics and Educational Data Visualization

Learning analytics and data visualization are rapidly evolving fields that are transforming education by providing tools and methods to analyze and visualize educational data effectively. This emerging field, known as Visual Learning Analytics (VLA), uses computational tools and methods for understanding educational phenomena through interactive visualization techniques. Using graphs, tables, and maps, professors can explore trends, identify anomalies, and make informed decisions. Also, Vieira et al., (2018) suggest that, although various VLA tools have been developed, their implementation in classroom settings is limited. Furthermore, few studies consider multiple sources of student information, and few use performance data to support their VLA approaches. They also found that the most commonly used statistical visualization techniques are simple and traditional, limiting the potential to enhance learning.

2 VLA PROJECT

2.1 The Design

Using a sample of educational data (Time, Username, Affected user, Event context, Component, Event name, Description, Origin, and IP address) recorded in UNAM's Moodle, various dashboards were designed and programmed (Figure 2). These dashboards identify and display student data and time spent on the platform, platform activity (schedule: minutes per day and hour), event context (number of actions), and the collaboration network between students and professors. Although the dashboards are primarily focused on providing insights for future opportunities, they can be adapted to specific goals such as monitoring dropout rates. The dashboards incorporate the five key elements of advanced visualizations as outlined by Vieira et al., (2018): (a) use of multiple visualizations; (b) connection between visualizations; (c) representation of data at multiple levels; (d) interactive visualizations; and (e) novel visualizations. According to the authors, these visualizations are rarely used.

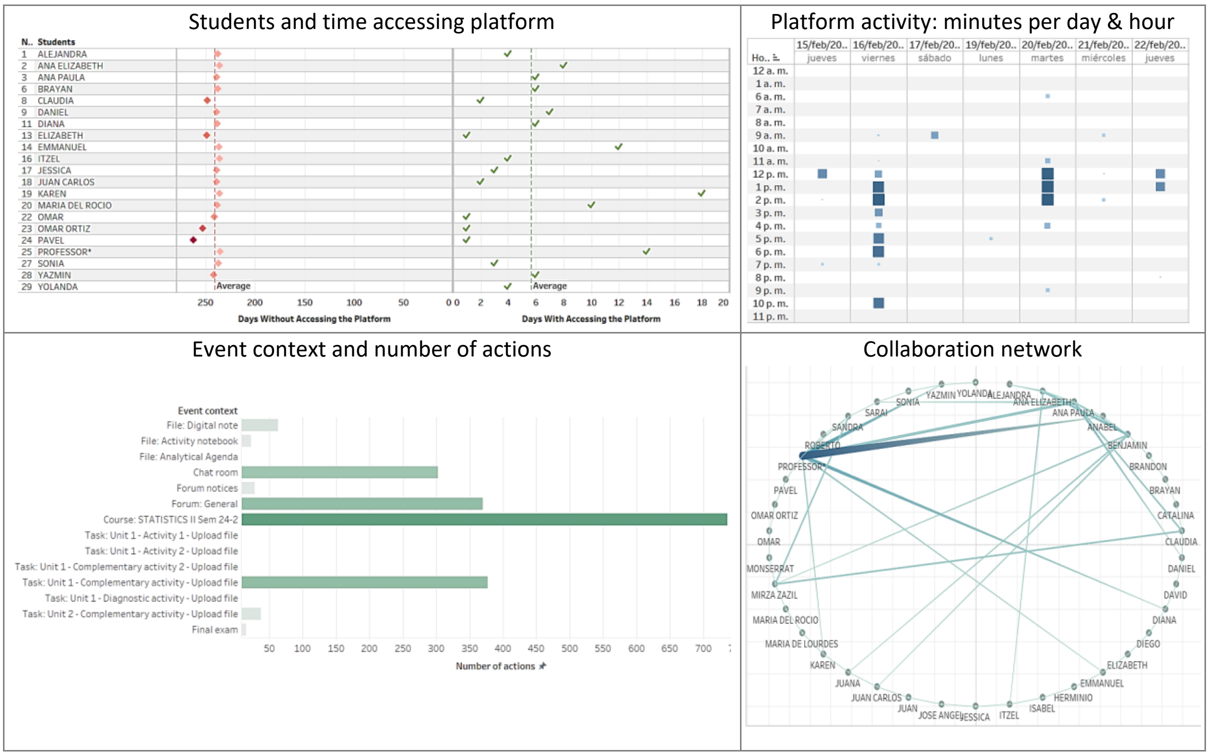


Figure 2. VLA Project Dashboard

3 CONCLUSION

Learning analytics and educational data visualization have become essential tools for improving the quality of the educational process by allowing continuous monitoring of student performance. Professors can detect early signs of difficulties, such as low participation or lack of engagement. By intervening promptly, dropouts can be prevented. Ongoing improvement and updating of the VLA project, along with training professors in the use of learning analytics and data visualization, will be crucial to advancing educational transformation and ensuring the success of students in distance education. In addition, these dashboards are designed to be effective in a big data environment.

REFERENCES

González-Videgaray, M., Valenzuela Argüelles, R., & Romero Ruíz, R. (2017). Primicias del primer censo de herramientas de gestión de cursos y repositorios digitales de aprendizaje en la UNAM. *Revista Digital Universitaria*, *18*(2). http://www.revista.unam.mx/vol.18/num2/art12/

Moodle Statistics. (2024). [Moodle registration map]. https://stats.moodle.org/

Shaikh, U. U., & Asif, Z. (2022). Persistence and Dropout in Higher Online Education: Review and Categorization of Factors. *Frontiers in Psychology*, *13*, 902070. https://doi.org/10.3389/fpsyg.2022.902070

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119–135. https://doi.org/10.1016/j.compedu.2018.03.018

Sensor-based Learning Analytics for Mentoring in Higher Education: An Example in the Context of the AI Act

David Buschhüter
German Research Center for Artificial Intelligence
david.buschhueter@dfki.de

Norbert Pengel
Leipzig University
norbert.pengel@uni-leipzig.de

ABSTRACT: Given the European AI Act's current state regarding affective computing technologies in educational institutions, there is a need for good practice examples of applications that could be exempt from prohibition. We aim to provide such an example for the context of AI-supported mentoring that uses the strength of sensor-based analytics to notify students of candidate incidents. These can serve as a starting point for a dialogue-based self-reflection on their emotional state and subsequent mentoring support. Thereby, the mentee is given the agency to interpret his/her emotions. By decoupling the incident detection system from the emotion interpretation, we reduce invasiveness and increase validity. We discuss the application in the context of the AI Act and argue for research focusing on sensor-based analytics that empower learners.

Keywords: affective computing, mentoring, AI Act, ethics

1 SENSOR-BASED ANALYTICS AS A STARTING POINT FOR MENTORING SUPPORT

Even if the reasons for dropout in higher education are generally known, providing interventions at the right time is challenging. Automatic emotion recognition technologies could solve this problem by providing sensor-based analytics to identify critical moments, suggest support, and enable scalable mentoring (Kadar et al., 2016). However, the AI Act prohibits emotion recognition systems in educational institutions (Artificial Intelligence Act: European Parliament Legislative Resolution of 13 March, 2024). The main concerns are intrusiveness (1), the lack of validity (2), and the possibility of abuse of such systems given the power imbalances (3) in educational institutions (Häuselmann et al., 2023). Even if the AI Act is already fixed and binding, the AI Office is in need of good practice examples for legal line drawing. Any exceptions need to address the concerns listed above (1-3). In line with research suggesting adverse effects of "[…] being told how we feel" (Hollis et al., 2018), we suggest decoupling the identification of physiological incidents based on physiological signals from inference to emotions. This approach does not only reduce intrusiveness (1) but also holds rigorous scrutiny from the perspective of validity (2), as it is in line with the constructionist framework of emotions, where physiological reactions are individually interpreted as emotions by the subject itself (Barrett, 2006). Applications such as "How We Feel" (https://howwefeel.org/) build on emotional intelligence (Nathanson et al., 2016) and can help users identify more granular emotions to respond constructively. Consistent with this approach, Martin and Pengel (2024) describe requirements of a chatbot to provide advisory mentoring support in higher education, e.g., via employing techniques such as mirroring. Such a service should also enable a low threshold of contact initiation for users

(Martin & Pengel, 2024). Including physiological signals could lower this threshold and help the students seek mentoring support at the right time without claiming authority concerning the mentee's emotional state.

2 DETECTING CANDIDATE SITUATIONS FOR MENTORING SUPPORT

In the following, we present a concept for a mobile application that notifies students of candidate incidents, which can serve as a starting point for a dialogue-based reflection on their emotional state and subsequent mentoring support.

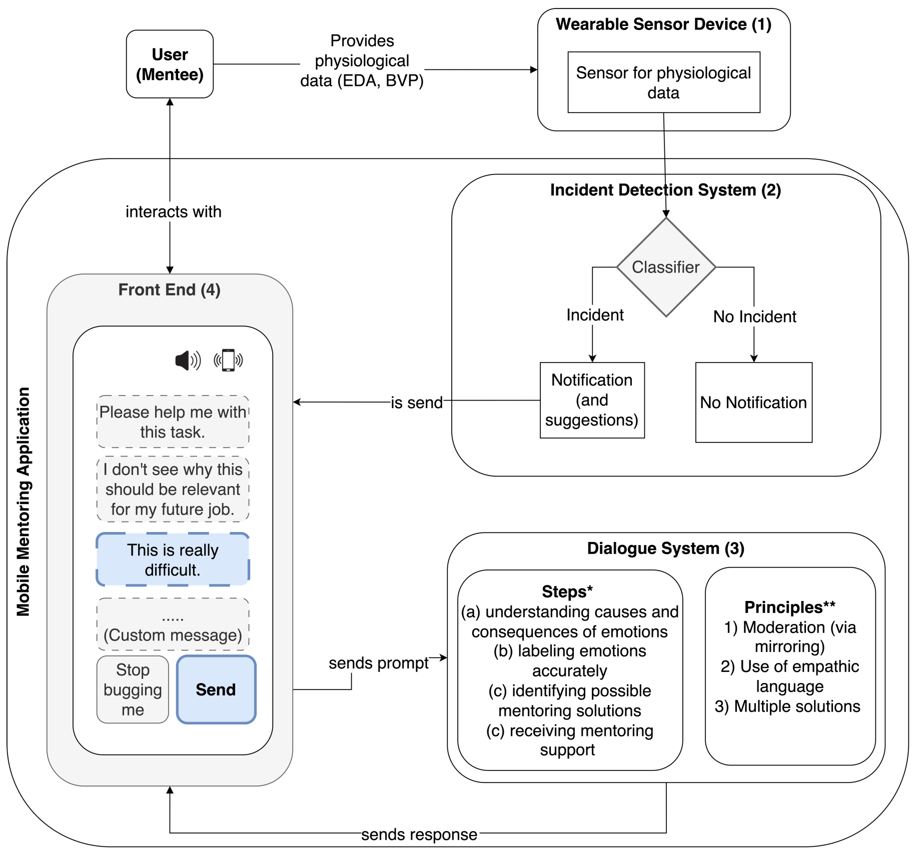


Fig 1: Mentoring application decoupling incident detection from the inference of emotions (adapted from *Nathanson et al., (2016), **from Martin and Pengel, (2024)).

The proposed system (see Fig 1) is based on the following modules: (1) "Sensor Device" for physiological data (e.g., a wristband with Electrodermal Activity and Blood Volume Pulse), private "Incident Detection System" (2) which receives sensor data, applies a classifier on the data (1 = incident, 0 = no incident). To help maintain regulatory compliance, the current version of our application under development relies on thresholds based on the raw physiological data (e.g., mean EDA values), while we also integrate a stress classifier reserved for research purposes. If an incident is detected, the user receives a notification where s/he has options to "write a custom message" or "choose standard suggestions". This message can be forwarded to an independent "Dialogue System" (3). This dialogue workflow builds on the RULER-approach (Nathanson et al., 2016) adapted to the needs of the mentoring situation and consists of the following steps: Recognizing emotions via understanding causes and consequences of emotions (a), labeling emotions accurately (b), identifying possible mentoring solutions (c) receiving mentoring support (d). One application scenario relates to detecting affective states of students who perceive that a text assignment lacks relevance.

3 DISCUSSION FROM THE PERSPECTIVES OF THE AI ACT

The application presented reduces intrusiveness (1) by providing real-time notifications to the user of a candidate's physiological incident. In line with Barret (2006), the validity (2) is increased by handing the interpretation of emotions to the subject. The power imbalance (3) aspect implies that no one else, being part of the power structure (e.g., teachers), can abuse the analytics in any way, so it is proposed that such systems must run locally on a user device. Furthermore, such applications must not be institutionalized (e.g., teachers prompting or expecting the use of such devices). The authors of this paper do not and cannot guarantee that the system as described above would be legal. However, as the AI Act's definition of emotion recognition excludes the recognition of obvious facial movements (e.g., smiling) from prohibition, we suggest that legal line-drawing could be well performed on the continuum from physiological to emotional state (e.g., elevated heart rate, apposed to anger). Applying the idea of decoupling physiology from emotion interpretation (Hollis et al., 2018) to the field of education, as suggested here, might not only be a thoughtful response to the AI Act but also hold the potential to achieve fruitful outcomes in the contexts of mentoring and education generally.

ACKNOWLEDGMENTS

This research was funded as part of the project tech4compKI by the Federal Ministry of Education and Research (BMBF) in Germany [Grant No 16DHB2206, 16DHB2208]. Views and opinions expressed are those of the authors and do not necessarily reflect those of the BMBF.

REFERENCES

Artificial Intelligence Act: European Parliament Legislative Resolution of 13 March 2024 (2024). https://www.europarl.europa.eu/RegData/seance_pleniere/textes_adoptes/definitif/2024/03-13/0138/P9_TA(2024)0138_EN.pdf

Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Häuselmann, A., Sears, A. M., Zard, L., & Fosch-Villaronga, E. (2023). EU law and emotion data. 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII), 1–8. https://doi.org/10.1109/ACII59096.2023.10388181

Hollis, V., Pekurovsky, A., Wu, E., & Whittaker, S. (2018). On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(3), 114:1-114:31. https://doi.org/10.1145/3264924

Kadar, M., Gutierrez y Restrepo, E., Luis-Ferreira, F., Calado, J., Artifice, A., Sarraipa, J., & Jardim-Goncalves, R. (2016). *Affective Computing to Enhance Emotional Sustainability of Students in Dropout Prevention*. 85–91. https://doi.org/10.1145/3019943.3019956

Martin, A. & Pengel, N. (2024). Beratung via Chatbot? Möglichkeiten und Anforderungen an den Einsatz generativer KI in einem bildungswissenschaftlichen Modul. *e-beratungsjournal, 20(1).* https://doi.org/10.48341/rfbq-t940

Nathanson, L., Rivers, S. E., Flynn, L. M., & Brackett, M. A. (2016). Creating Emotionally Intelligent Schools With RULER. *Emotion Review*, *8*(4), 305–310. https://doi.org/10.1177/1754073916650495

Elevating E-Learning Engagement through Real-Time Gaze Tracking and Adaptive Feedback

Bineeth Kuriakose[1,4], Pedro Lind[1,2,3] and Pedro Lencastre[1,2]
[1]Department of Computer Science, Oslo Metropolitan University, Oslo, Norway
[2]NordSTAR - Nordic Center for Sustainable and Trustworthy AI Research, Oslo, Norway
[3]Numerical Analysis and Scientific Computing, Simula Research Laboratory, Oslo, Norway
[4]Department of Education, University of Oslo, Norway
{bineethk,pedrolin,pedroreg}@oslomet.no

ABSTRACT: As digital learning environments expand, maintaining student engagement and focus is a major challenge. This study presents a prototype e-learning framework that uses real-time gaze tracking and adaptive feedback to boost engagement and comprehension. Using webcams to track gaze patterns, the framework identifies Areas of Interest (AOIs) and provides visual and auditory cues when attention shifts. This study outlines the framework's design—covering gaze detection, AOI monitoring, and adaptive feedback—and explores its potential for personalized, interactive e-learning. Preliminary testing suggests that gaze-based feedback may help refocus attention, but further analysis is needed.

Keywords: Gaze Tracking, Adaptive Feedback, E-Learning, Personalized Learning

1 INTRODUCTION

The rapid proliferation of digital learning platforms has transformed education, offering unprecedented flexibility and access to resources. However, maintaining student engagement and focus on these environments remains a significant challenge. Unlike traditional classrooms, where instructors can respond to non-verbal cues in real time, digital learning platforms often lack mechanisms to dynamically monitor and address student disengagement, leading to reduced comprehension and learning outcomes. Research consistently highlights that sustained attention is critical for effective learning, yet many e-learning systems deliver passive, non-interactive content that fails to adapt to students' engagement levels (Benabbes et al., 2023).

Gaze-tracking technology offers a promising way to monitor student focus during learning (Wang et al., 2021). This study presents a prototype framework that detects attention shifts from critical content and provides adaptive feedback through predefined Areas of Interest (AOIs). Educators can tailor AOIs to learning objectives, with visual and auditory cues redirecting focus without disrupting the flow. By addressing a critical gap in current e-learning platforms, this framework balances engagement with attention dynamics, acknowledging that brief focus lapses can enhance cognitive processing and support deeper learning.

The primary objective of this research is to evaluate how real-time gaze tracking paired with adaptive feedback can improve student engagement and comprehension in e-learning environments. Preliminary proof-of-concept results assess the feasibility and effectiveness of this approach, highlighting its potential for interactive and personalized e-learning.

2 SYSTEM DESIGN

The proposed framework is built on three core modules as illustrated in Figure 1.

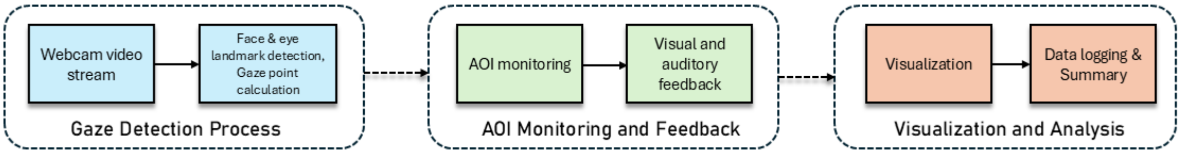


Figure 1: Process flow of the real-time gaze tracking and adaptive feedback framework.

2.1 Gaze Detection Process

The framework initiates engagement tracking by capturing real-time video through a standard webcam. Facial recognition algorithms identify critical facial landmarks, particularly around the eyes, enabling the system to detect eye position and determine the user's gaze point on the screen. By mapping this gaze point, the prototype monitors where the user's focus is directed, laying the foundation for responsive engagement tracking in real time.

2.2 AOI Monitoring and Adaptive Feedback

Educators define key Areas of Interest (AOIs) on the screen (e.g., instructional text or images) to represent core content. The prototype continuously checks if the user's gaze remains within these AOIs or shifts to less relevant areas. If a shift away from an AOI occurs, the feedback module responds with real-time cues to refocus attention. Visual and auditory feedback gently direct the user back to essential information, supporting continuous engagement without interrupting the learning flow.

2.3 Visualization and Analysis

To assess user engagement, the prototype logs detailed gaze data, including metrics like time spent within each AOI, gaze duration, and frequency of attention shifts. After each session, the data is analyzed to generate summary statistics, such as fixation duration on specific AOIs and time spent on distractors. This analysis offers insights into individual engagement patterns, enabling educators to adjust learning content as needed to enhance attention and focus.

3 PRELIMINARY RESULTS

A proof-of-concept validation was conducted with a reading comprehension task using the proposed framework. The working of the prototype is illustrated in Figure 2. Early observations showed an encouraging trend, with participants spending increased time on key AOIs compared to baseline metrics. Feedback from participants indicated that the adaptive cues helped maintain focus on essential content without disrupting the learning flow. However, quantitative improvements in comprehension are not yet established, and further qualitative analysis is needed to assess the framework's effectiveness comprehensively. Participants appreciated the adaptive feedback, noting that it reinforced focus on critical content without disrupting their reading flow.

Figure 2: The first image shows eye gaze tracking as the user focuses on the AOI (red box). The second image displays a text prompt, "Focus on the screen," with audio feedback when the user's gaze shifts away from the AOI.

4 DISCUSSION

The proof-of-concept results suggest that gaze-based adaptive feedback shows promise for enhancing learning engagement and focus. By guiding users to predefined AOIs, the prototype helps sustain attention on critical content, aligning with cognitive psychology principles related to attention-based learning. The visual and auditory feedback mechanisms appear to support users in maintaining focus, especially during cognitively demanding tasks. Nonetheless, managing the balance between sustained attention and natural attention lapses is critical, as brief lapses may contribute to cognitive processing and overall learning outcomes. Optimizing feedback delivery is essential to ensure that it supports rather than overwhelms learners. Early observations indicate a need for adjustable feedback thresholds that dynamically respond to individual engagement levels and content type.

Although these initial findings are promising, broader testing is required to fully understand its effectiveness across varied age groups and educational backgrounds. Moreover, integrating privacy-preserving mechanisms, such as data anonymization and user consent protocols, is vital to address ethical concerns related to gaze data collection and usage.

5 CONCLUSION

This research introduces a novel approach to adaptive e-learning by leveraging gaze tracking and real-time feedback to enhance student engagement and comprehension. The proof-of-concept demonstrates a scalable method for directing attention to critical content, with preliminary results indicating potential improvements in focus. Future work will address current limitations by integrating a wider range of engagement metrics, optimizing feedback delivery, and ensuring ethical handling of gaze data through anonymization and user control.

REFERENCES

Benabbes, K., Housni, K., Hmedna, B., Zellou, A., & El Mezouary, A. (2023). A new hybrid approach to detect and track learner's engagement in e-learning. IEEE Access.

Wang, Y., Lu, S., & Harter, D. (2021). Multi-sensor eye-tracking systems and tools for capturing Student attention and understanding engagement in learning: A review. IEEE Sensors Journal, 21(20), 22402-22413.

Digital Blind Spots: Why Students Trust But Don't See Learning Analytics in Their Courseware

Martha Hubertz
University of Central Florida
Martha.Hubertz@ucf.edu

Patsy Moskal
University of Central Florida
Patsy.Moskal@ucf.edu

ABSTRACT: This study investigates student perceptions of learning analytics and courseware systems at a large Research-1 Hispanic Serving Institution (N = 1,422). Through survey data, we examined the relationship between student trust, awareness, and comfort with learning analytics tools integrated within institutional courseware. Results revealed a significant disparity between students' trust in educational technology and their awareness of its analytical capabilities. While students demonstrated high trust in courseware data security (M = 3.33, SD = 1.09) and educational data collection (M = 3.34, SD = 1.11) compared to general technology systems (M = 2.24, SD = 1.06), their awareness of learning analytics features was remarkably low, with only 1.44% recognizing these capabilities in their learning management system. Comfort levels varied across different analytics applications, ranging from 35.07% for assessment integrity tools to 56.1% for general analytics tools. These findings highlight a critical trust-awareness gap in higher education's digital landscape and suggest the need for enhanced institutional communication about learning analytics implementation. The study contributes to our understanding of how student perceptions may influence the effective deployment of learning analytics in higher education settings.

Keywords: learning analytics, courseware, educational technology, student perceptions, data trust

1 INTRODUCTION

Research on student awareness and trust in learning analytics (LA) within higher education online courseware has emerged as a critical area of study as institutions increasingly adopt these technologies. Studies have consistently shown that while students generally express positive attitudes toward LA, their awareness of how their data is collected and used remains limited (Roberts et al., 2016; Slade & Prinsloo, 2015). A comprehensive survey by Whitelock-Wainwright et al. (2020) found that only 32% of students reported understanding how their learning data was being collected and analyzed, despite 78% indicating they believed LA could positively impact their academic performance. This gap between perceived utility and understanding highlights a significant challenge in LA implementation.

Trust in LA systems has been closely examined, particularly regarding data privacy and institutional transparency. Research by Wong (2017) demonstrated that students' willingness to share their data was significantly influenced by their trust in their institution's data handling practices and the perceived benefits of LA interventions. Building on this work, Chen and Ferguson (2019) found that students were more likely to trust and engage with LA systems when they received clear communications about data collection purposes, storage methods, and potential uses. These findings align with earlier work by Pardo and Siemens (2014), who emphasized the importance of establishing ethical frameworks and transparency in LA implementations to build student trust.

The intersection of awareness and trust has important implications for LA effectiveness. Studies indicate that students who better understand LA systems are more likely to trust and actively engage with them (Harrison & Greenfield, 2018; Thompson et al., 2021). This relationship appears to be particularly strong in online learning environments, where students rely more heavily on digital tools and analytics for feedback and progress monitoring. However, research also suggests that increased awareness can sometimes lead to decreased trust if students perceive the data collection as too invasive or the analysis methods as potentially biased (Edwards & Smith, 2020).

2 METHOD

Participants were 1,422 undergraduate and graduate students recruited through the university's SONA research participation system at a large R1 Hispanic-Serving Institution (HSI) in the southeastern United States during Fall 2022. Participants completed an online survey administered through Qualtrics. After providing informed consent, participants completed demographic questions followed by several standardized measures assessing their attitudes, beliefs, and experiences related to artificial intelligence and technology use. The survey took approximately 30-45 minutes to complete, and participants received course credit for their participation.

After excluding participants who self-reported not being truthful in their responses ($n = 185$), the final sample consisted of 1,183 students. The sample was predominantly undergraduate students (99.57%) with a small number of graduate students (0.43%). Most participants were first-year students (52.78%) or sophomores (21.87%). The sample was majority female (57.22%) with an age range of 18-54 years ($M = 19.48$ years, $SD = 3.44$). Participants identified as White (71.89%), Asian (11.79%), Black or African American (11.19%), and other racial categories (4.26%). Additionally, 30.08% identified as Hispanic/Latino. Most participants were native English speakers (81.90%).

3. RESULTS

Preliminary analysis focused on student awareness and trust in learning analytics within institutional courseware. Students (Figure 1) demonstrated high trust in courseware data security ($M = 3.33$, $SD = 1.09$) and educational data collection ($M = 3.34$, $SD = 1.11$), significantly higher than trust in general technology systems ($M = 2.24$, $SD = 1.06$), $t(2737) = 24.86$, $p < .001$. However, awareness of learning analytics features was notably low, with only 1.44% recognizing such capabilities within their learning management system. Students showed moderate comfort with analytics-based assessment tools (56.1% expressing comfort), particularly when integrated within familiar courseware environments. Notably, student comfort with learning analytics applications varied by purpose: assessment integrity (35.07% somewhat comfortable), adaptive learning materials (2.79% recognition rate), and educational data analysis (49.12% expressing trust). These findings suggest a disconnect between students' trust in institutional courseware and their awareness of embedded learning analytics, indicating opportunities for institutions to better communicate the presence and benefits of these analytical tools within educational technology systems.

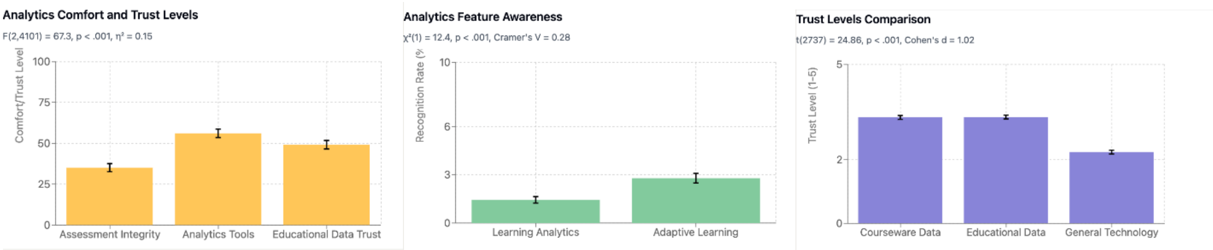


Figure 1. Student comfort, trust, and awareness in learning analytics and educational technology.

4. DISCUSSION

While previous studies have identified gaps between trust and awareness in learning analytics, our findings reveal an unprecedented disparity in an HSI context (1.44% awareness despite high trust scores of M = 3.33). This suggests that institutional type may significantly influence how students perceive and interact with learning analytics systems. This study advances our understanding of institutional data practices by demonstrating that HSIs' approaches align with those of other institution types. This consistency provides a crucial foundation for future research directions, particularly in developing disclosure strategies for analytics, evaluating how awareness influences trust, and establishing industry best practices.

To address implementation barriers, we propose specific strategies: implementing mandatory disclosure statements in course syllabi, developing interactive student orientations to analytics features, and piloting opt-in analytics programs where students actively participate in data usage decisions. These approaches must account for resource constraints, varying levels of digital literacy, and institutional privacy policies. Moving forward, our research agenda will focus on developing evidence-based strategies for analytics disclosure, measuring the impact of awareness on trust levels, and establishing best practices for analytics communication in higher education. This work aims to help institutions maintain their trusted status while fostering student understanding of the educational technologies that increasingly shape their learning experiences.

REFERENCES

Chen, B., & Ferguson, J. (2019). Student perceptions of learning analytics in higher education: An exploratory study. Journal of Computing in Higher Education, 31(2), 292-308.

Edwards, M., & Smith, K. (2020). The double-edged sword of learning analytics awareness: Impact on student trust and engagement. The Internet and Higher Education, 45, 100722.

Harrison, T., & Greenfield, D. (2018). Understanding student engagement with learning analytics: A case study of undergraduate online learners. Assessment & Evaluation in Higher Education, 43(7), 1069-1084.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438-450.

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: "The Fitbit version of the learning world." Frontiers in Psychology, 7, 1959.

Slade, S., & Prinsloo, P. (2015). Student perspectives on the use of their data: Between intrusion, surveillance and care. European Journal of Open, Distance and E-learning, 18(1), 291-305.

Thompson, K., Chen, X., & Anderson, R. (2021). Trust and transparency in learning analytics: A mixed-methods investigation of student perspectives. Learning Analytics and Knowledge Review, 13(2), 156-173.

Whitelock-Wainwright, A., Gašević, D., & Tejeiro, R. (2020). What do students want? Making sense of student preferences in technology-enhanced learning. IEEE Transactions on Learning Technologies, 13(1), 142-153.

Wong, B. T. M. (2017). Learning analytics in higher education: An analysis of case studies. Asian Association of Open Universities Journal, 12(1), 21-40.

What are the Possible Futures of Learning Analytics?

Oleksandra Poquet
Technical University of Munich
sasha.poquet@tum.de

Kirsty Kitto
University of Technology Sydney
kirsty.kitto@uts.edu.au

Catherine A. Manly
Fairleigh Dickinson University
c.manly@fdu.edu

Rebecca Ferguson
The Open University
rebecca.ferguson@open.ac.uk

ABSTRACT: Trends in the emerging technologies can easily reshape the direction of fields such as learning analytics (LA). To help the LA community reflect on its possible futures, our poster presents insights from the past and present visions of LA. Data for the poster will be collected during a pre-conference workshop. The workshop's aim is to facilitate a discussion with experts about the grand challenges and possible futures of LA. This poster, together with the workshop, will feed into the interactive panel offered as a third keynote at the LAK'25 conference. This will help the LA community to point to well-established "blue skies" requiring more work when applying for funding. It will also support new entrants to LA in seeing the bigger picture when plotting out their research trajectory.

Keywords: Grand Challenges; Theory; Evidence; Synthesis

1 THE PROBLEM

How do the grand challenges of Learning Analytics (LA) connect with possible future directions for the field? When we consider our theoretical contributions, how does our work specifically add to the field of education as we envision our futures? Some attempts have been made to highlight how LA might address large-scale challenges (Buckingham Shum, 2023). Further, efforts in related fields such as artificial intelligence in education (AIED) can be informative (Kay, 2012) and indeed, a list of grand challenges for the field has been put forward (Baker, 2019), but we are yet to coalesce around a community-defined set of research priorities to guide future work. Against this backdrop, the aims of large LA research groups are not always aligned, and there have even been recent bandwagon effects where a 'hot topic' emerges and distracts attention from areas with potential for benefitting the field. Most crucially, new entrants to the field and early career researchers could find it difficult to understand the rich background landscape of the field and why certain problems have been identified as important. Without a clear unifying set of challenges, it is likely that LA will make only incremental contributions, if any. We are at risk of becoming feudalistic, with various teams staying within their safe, identified subfields.

While education itself is often touted as a field that will help us create a more equal and just society, LA is sometimes accused of supporting agendas that will track people, violate their privacy, and manipulate them towards acts that they might not have undertaken on their own. How can we work towards ensuring that the field is solving big issues that help to ensure the next generation of people are more mindful and accepting of each other and the differences between us, respond less to the abundance of false information, and are able to adjust in ways that are well reasoned rather than simply reactive to societal shifts and new challenges? An important component for doing this involves having a clear understanding of our past.

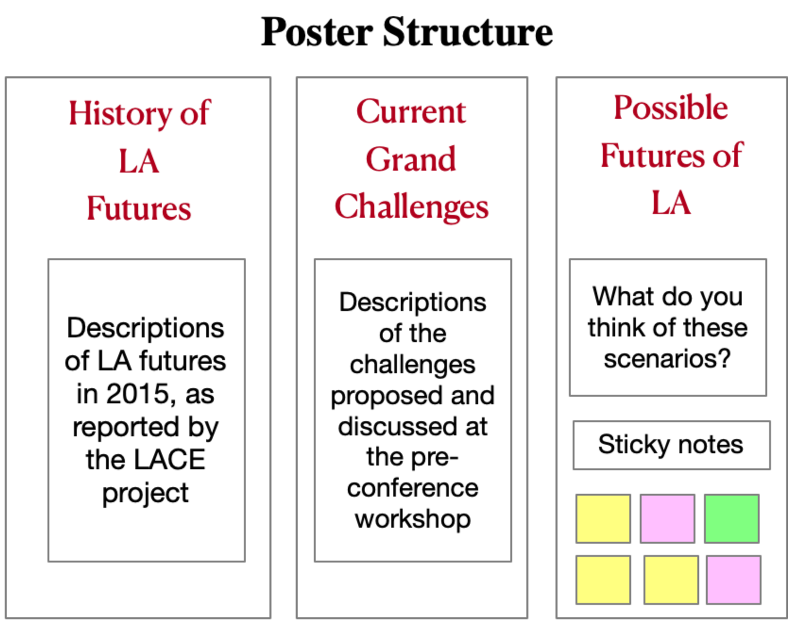
This poster will help link the Grand Challenges of Learning Analytics pre-conference workshop with the third LAK keynote run as an interactive panel focused on the possible futures of Learning Analytics during the Friday session. As such, this poster will help the LAK community engage with the history of work in this area, to connect with the current grand challenges, and to explore its possible futures.

2 THE HISTORY AND PRESENT OF LA

This poster will build on an important past contribution on the possible futures of LA from work completed by Ferguson et al. (2019). They reported on a Delphi process conducted in 2015, where over 100 LA experts were invited to evaluate eight LA future visions that described the reality of learning analytics in 2025. Examples of those future scenarios included: "In 2025, classrooms monitor the physical environment to support teaching and learning" and "In 2025, most teaching is delegated to computers". The experts were invited to rank the desirability of these possible outcomes, as well as their likelihood. Importantly, while some of the futures were deemed quite likely at the time, none have yet come to pass. As such, it is time to revisit these possible futures and to think about whether the work being completed in the field is the work we ought to be completing.

The second section of the poster will present the grand challenges that emerged during the pre-conference workshop (Kitto et al., 2025). These will be an important mechanism for helping the field to define and understand its present. The third section will focus on the future by eliciting responses from poster attendees during the session itself.

The general structure that the poster will take is illustrated in Figure 1.



3 THE POSTER'S CONTRIBUTION TO THE BROADER SCOPE OF WORK

To tackle this issue of planning communal effort as well as critically reflecting on the past and current work, we propose a three-step process that will take place at the LAK Conference:

1) The first step of this process is the pre-conference workshop devoted to identifying and refining a series of grand challenges supported by the broader LAK community. In addition to the grand challenges, a set of enabling problems will also be identified, and various LA subfields will be mapped into the different programs of work.

2) The outcomes of the workshop will form the input for the second component of this poster. The poster will not only present the results of the workshop but have an interactive component, inviting viewer to contribute input around their perceptions and feasibility of the newly defined grand challenges, as presented on the poster.

3) The third part of this process will be an interactive panel taking place in the keynote slot. In this interactive format, the panelists will provoke the audience with potential grand LAK challenges and seek the audience's feedback. In a structured format, the audience will interact with the provocations to negotiate the final set of challenges in a larger community.

This poster, therefore, is a part of a scaffolded structured process to engage various parts of the LA community (experts, conference attendees, keynote attendees) into a distributed conversation about the values and directions of research in the LA community. This conversation is perhaps more valuable than the outcomes of it, as the learning analytics community is facing stalling in its innovation, challenges in scaling, and questions around its boundaries with the AI hype in its full development.

REFERENCES

Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *Journal of Educational Data Mining*, *11*(1), 1-17.

Buckingham Shum, S. (2023). Trust, sustainability and Learning@Scale. Keynote Address, *Proceedings of the Tenth ACM Conference on Learning@Scale (L@S '23).* Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3573051.3593375

Ferguson, R., Clow, D., Griffiths, D., & Brasher, A. (2019). Moving forward with learning analytics: Expert views. *Journal of Learning Analytics, 6*(3) 43–59. https://doi.org/10.18608/jla.2019.63.8

Kay, J. (2012). AI and education: Grand challenges. *IEEE Intelligent Systems*, *27*(5), 66-69. https://doi.org/10.1109/MIS.2012.92

Kitto, K., Poquet, O., Manly, C. A., and Ferguson, R. (2025). What are the grand challenges of learning analytics? *Companion Proceedings of the 15th International Conference on Learning Analytics & Knowledge, Dublin, Ireland.*

Strategic Networking for Venture Creation in University-Based Entrepreneurship Education

Yuya Takeda*, Yuta Tomokiyo*, Yukihiro Murata*, Lui Yoshida*
*The University of Tokyo, Tokyo, Japan
{takeda-yuya190, tamokiyo-yuta721, yukihiromurata, luiyoshida}@g.ecc.u-tokyo.ac.jp

ABSTRACT: This study explores how entrepreneurship education (EE) fosters effective networking for venture creation among pre-career students. Prior research has focused primarily on professional entrepreneurs and MBA students, leaving a gap in understanding how students with limited initial networks can build essential entrepreneurial connections. Grounded in the "network success hypothesis," which posits that access to key resources impacts venture creation more than network size or diversity, this study tracks the networking patterns and venture outcomes of students from a Japanese university's pre-career EE program. Using network analysis and multinomial logistic regression on survey data from 135 alumni, we examined the roles of network structure and strategic networking in students' venture progress. Network analysis results indicate that broad acquaintance networks did not correlate with venture advancement; targeted connections to influential individuals significantly contributed to venture creation. Regression analysis further highlights continuous engagement with individuals who serve as connectors to key resources and opportunities proved critical for advancing students' business. These results underscore the value of incorporating well-connected mentors in EE, suggesting a design in EE towards strategic mentorship and resource-accessible networks for students. In addition, we contribute to EE evaluation by providing insights into the longitudinal development of pre-career students' entrepreneurial networks.

Keywords: Entrepreneurship Education, Network Analysis, Higher Education

1 INTRODUCTION

Entrepreneurship plays a vital role in driving industrial competitiveness and fostering innovation [1], leading to a growing emphasis on entrepreneurship education (EE) at universities. However, much research has focused narrowly on pre-post measures of entrepreneurial intention [2, 3]. While longitudinal studies on entrepreneurs underscore the importance of evolving networks, such research is scarce in EE, despite increasing recognition of its relevance [4]. As a result, understanding how initial networking efforts contribute to venture progress over time remains limited [5].

Networking is crucial for entrepreneurial success. The "network success hypothesis" suggests that access to valuable resources—more than network size or diversity—influences venture creation [6]. Existing studies have primarily examined working adults [7] or MBA students [8] who already possess professional networks. Despite validations across various contexts, how pre-career students establish and leverage networks remains insufficiently studied. This lack of research leaves uncertain whether existing insights apply to this group, hindering the effective design and evaluation of EE programs tailored to their needs.

These gaps highlight the growing need to study how students without established networks develop connections and how these networks influence their entrepreneurial behaviors, thereby informing the design of more effective, action-oriented EE programs. To address this, we examine how EE influences students' network development and identify the strategies that effectively support venture creation by tracking the progress of alumni from venture-focused EE programs. Furthermore, by analyzing the networking practices of students demonstrating measurable progress in their projects, we investigate what types of networking are effective for company creation.

2 METHODS

2.1 Sample

This study surveyed 166 alumni who had completed a Japanese university’s pre-career EE program within the past six months to three years; 135 persons replied, and the response rate was 81.3%. The program includes a 4-month for-credit course offering lectures and business plan development, followed by ongoing support through feedback and networking within and beyond the alumni community. Distinctive features of the program include (1) an action-oriented approach that engages experienced entrepreneurs and mentors to support venture creation; (2) a curriculum focused on science-based startups, integrated with university education for practical insights; and (3) extensive networking facilitated by a 1:3 mentor-student ratio, supported by corporate partnerships and alumni Bridging Tutors (BTs) who assist in networking and foster community engagement.

2.2 Measure

A structured questionnaire was administered with questions about alumni connections, venture progress, networking, number of entrepreneurs around, activities before entering university, and aspects of the EE program that they found helpful. The network was constructed based on alumni connections, and the index of network centrality was used in the analysis. We examined three centrality metrics: eigenvector centrality, capturing influence through connections to highly connected nodes; betweenness centrality, indicating the node's role in connecting disparate parts of the network; and degree centrality, measuring the node's direct connectivity. In addition, clustering was performed using the Louvain method to see connections at the sub-community level. We also conducted multinomial logistic regression analysis of 135 alumni. The dependent variable is "venture progress" categorized into three stages—"not yet (n=84)", "planning (n=16)" and "creation (n=35)"—following the phase in the establishment process used in [9]. "Planning" is a situation in which they have participated in business model competition but have not actually taken action, while "creation" is in which they have already started a business, developing based on grants, or starting a business and taking on projects on consignment. The independent variables are the following six: "degree centrality" adopted as the abundance of connections because of multicollinearity among the centrality indices; "communication" with pivotal introducers—In this analysis, defined as a person with a betweenness centrality of 0.05 or greater in another network created by the response items of networking that participants reported whom they referred or asked for referral and communicating about once a week; The number of "friends" and "relatives" within two degrees of kinship to explore the influence of surrounding entrepreneurs on entrepreneurial behavior; "awards" before university entrance, such as science Olympiads; "BTs" experience. Additionally, participants identified "which aspects of the course and alumni activities contributed to their venture progress" through multiple-choice questions.

3 RESULTS

T-tests and U-tests revealed no statistically significant differences in eigenvector, betweenness, and degree centrality in the acquaintance network across three groups of phases. This indicates that there are no significant correlations between these metrics and venture progress. Figure 1 presents a visualization of the network, where node size represents eigenvector centrality and node color intensity (logarithmic scale) represents betweenness centrality. Additionally, nodes belonging to the "creation" group and the "planning" group are highlighted in yellow and lime green, respectively.

Clustering analysis using the Louvain method revealed a concentration of "creation" group members within two of the four clusters. In Figure 2, these clusters are represented by red and blue nodes, with nodes from the "creation" group highlighted in yellow, consistent with the marking in Figure 1. In addition, regression analysis (Table1) shows that frequent and focused "communication" significantly increased "creation" likelihood while higher "degree centrality" reduced the odds of it. However, it is worth noting that although

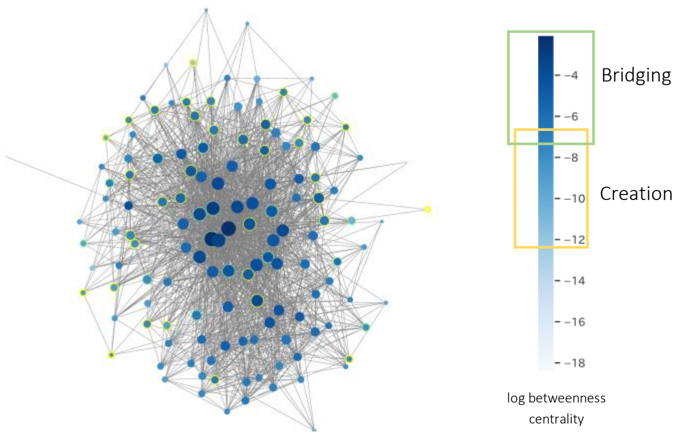


Figure 1: Visualization of the acquaintance network with eigenvector centrality and betweenness centrality

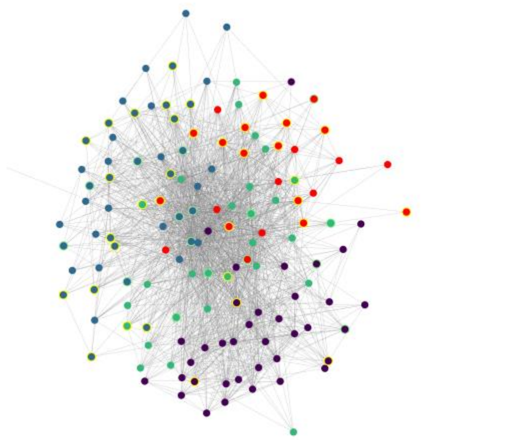


Figure 2: Visualization of the acquaintance network with clusters

these factors are statistically significant at the 5% significance, their p-values are close to the threshold, and the model demonstrates relatively low explanatory power, as indicated by a modest coefficient of determination. Other factors, including the presence of entrepreneurial "relatives", the number of entrepreneurial "friends", high school "Awards ", and "BTs" experience indicated no significant associations with either "planning" or "creation".

In the "aspects of the EE program that they found helpful" question, participants in the "creation" group demonstrated distinct tendencies in their reported benefits from the course and alumni community. This group indicated that "networking" was the most valuable component, whereas the other two groups primarily cited "inspiration from others" as the most influential factor.

Table 1: Multinomial logistic regression analysis (R^2=0.1030)

Independent var.	Regression Coefficient (creation)	Standard Error (creation)	Regression Coefficient (planning)	Standard Error (planning)
Degree Centrality	**-7.3835 (p=0.049)**	3.746	5.5582 (p=0.175)	4.100
Communication	**0.1404 (p=0.048)**	0.071	-0.0869 (p=0.327)	0.089
Relatives	0.5845 (p=0.216)	0.473	0.8571 (p=0.209)	0.682
Friends	0.0343 (p=0.785)	0.126	-0.1141 (p=0.538)	0.186
Awards	0.5080 (p=0.356)	0.508	-0.1408 (p=0.850)	0.743
Bridging Tutors (BTs)	0.7296 (p=0.224)	0.600	1.4129 (p=0.050)	0.720

4 DISCUSSION AND CONCLUSION

The regression analysis reveals that strategic networking, rather than a broad network of connections, is key to entrepreneurial progress, while centrality is less important in determining influence within the network. Additionally, cluster analysis suggests that the components of the creation "group" are relatively close to each other and form a community. Our findings indicate that general acquaintance networks do not correlate with venture progress, consistent with previous studies on social networks and entrepreneurship. However, continuous engagement with individuals who serve as connectors to key resources and opportunities proved critical for advancing students' business. This suggests that even when starting with limited personal networks, strategically building connections with individuals who can introduce a diverse range of contacts within and beyond the community is more impactful than broad networking efforts solely within the community. It also underscores the importance of "bridging" connections and supports the effectiveness of incorporating BTs into entrepreneurship courses.

This study makes several contributions to the field of EE evaluation. First, it provides a new direction for longitudinal tracking in EE assessment, addressing a previously underexplored area. Second, it elucidates the relationship between venture progress and network characteristics among students, highlighting the potential of designing programs that involve individuals who possess established networks rather than solely focusing on developing students' networking skills. This insight suggests that inviting well-connected individuals into EE programs could enhance program effectiveness, rather than offering prizes or featuring guest speakers.

REFERENCES

[1] Anubhav, K., Dwivedi, A. K., & Aashish, K. (2024). Entrepreneurship education in higher education (2002–2022): A technology-empowered systematic literature review. *The International Journal of Management Education*, *22*(3), 100993.

[2] Nabi, F. Liñán, A. Fayolle, N. Krueger and A. Walmsley, The Impact of Entrepreneurship Education in Higher Education: A Systematic Review and Research Agenda, Academy of Management Learning & Education, 16(2), 277–299 (2017)

[3] Schimperna, F., Nappo, F., & Collaretti, F. (2022). Universities and CSR teaching: new challenges and trends. *Administrative Sciences*, *12*(2), 55.

[4] Longva, K. K. (2021). Student venture creation: developing social networks within entrepreneurial ecosystems in the transition from student to entrepreneur. *International Journal of Entrepreneurial Behavior & Research*, *27*(5), 1264-1284.

[5] Hoang, H., & Antoncic, B. (2003). Network-based research in entrepreneurship: A critical review. *Journal of business venturing*, *18*(2), 165-187.

[6] Dubini, P., & Aldrich, H. (2002). Personal and extended networks are central to the entrepreneurial process. *Entrepreneurship: Critical perspectives on business and management*, 217-228.

[7] Birley, S., Cromie, S., & Myers, A. (1991). Entrepreneurial networks: their emergence in Ireland and overseas. *International Small Business Journal*, *9*(4), 56-74.

[8] Baldwin, T. T., Bedell, M. D., & Johnson, J. L. (1997). The social fabric of a team-based MBA program: Network effects on student satisfaction and performance. *Academy of management journal*, *40*(6), 1369-1397.

[9] Greve, Arent. "Networks and entrepreneurship—an analysis of social relations, occupational background, and use of contacts during the establishment process." *Scandinavian journal of management* 11.1 (1995): 1-24.

Aligning Analytics with Theory: A Customized Epistemic Network Analysis Rotation for the Practical Inquiry Model

Shen Ba
Education University of Hong Kong
bas@eduhk.hk

Yahan Chen
University of Wisconsin-Madison
chen2254@wisc.edu

Yuanru Tan
University of Wisconsin-Madison
yuanru.tan@wisc.edu

Guoqing Lu
Northwest Normal University
luguoqing@nwnu.edu.cn

David Williamson Shaffer
University of Wisconsin-Madison
dws@education.wisc.edu

ABSTRACT: This study explores the alignment of Epistemic Network Analysis (ENA) with the Practical Inquiry (PI) model, which describes the cognitive processes of critical thinking in educational contexts. The PI model specifically describes the phases of inquiry: Triggering Event, Exploration, Integration, and Resolution. To better understand how these phases manifest in discourse, we applied a customized ENA rotation to data from 108 pre-service teachers discussing "Artificial Intelligence in Education," with half using a GPT-4 chatbot. The customized ENA rotation was designed to better align with the PI model's dimensions. Results show that groups using chatbot made stronger connections between Exploration and Integration, while those without chatbot linked Exploration more to the Triggering Event. Individual analyses reveal that higher cognitive and social engagement correlates with greater progression through the inquiry phases, with only highly engaged students reaching the Resolution phase. This study offers a refined method for visualizing cognitive development in collaborative inuiry.

Keywords: Practical Inquiry Model; Epistemic Network Analysis; Situational Engagement; Artificial Intelligence; Inquiry-based Discussion

1 BACKGROUND

1.1 Practical Inquiry Model

The Practical Inquiry (PI) model describes the cognitive process of critical thinking and knowledge construction within educational contexts and is part of the broader Community of Inquiry (CoI) framework, which emphasizes collaborative meaning-making and reflective discourse (Kim & Gurvitch, 2020). The PI model consists of four phases in inquiry-based learning: (1) ***Triggering Event***, where a problem or question is identified; (2) ***Exploration***, where learners investigate by gathering information and brainstorming; (3) ***Integration***, where they synthesize information and form coherent solutions; and (4) ***Resolution***, where they apply their knowledge to real-world situations. These phases are organized along two cognitive dimensions: the action-deliberation axis, which represents the shift between practical application and reflective thinking, and the perception-conception axis, which spans from observing concrete information to forming abstract ideas. This dynamic back-and-forth movement along both axes is central to fostering deep understanding and critical thinking in collaborative inquiry.

1.2 Applications of Epistemic Network Analysis in Practical Inquiry

Epistemic Network Analysis (ENA) is an analytical method used to model and visualize the connections between concepts in discourse data. It quantifies how ideas co-occur within conversations or written exchanges, allowing researchers to understand how knowledge is constructed through interaction. In collaborative inquiry, researchers have employed ENA to analyze connections between cognitive development phases, uncovering groups' and individuals' unique inquiry patterns and strategies. In addition, studies have applied ENA to the broader CoI framework (Ba et al., 2024).

While existing studies have revealed the connections between phases of cognitive development, they have not aligned these analytical results with theoretical assumptions. Specifically, further exploration is needed to determine whether the dimensions (i.e., axes) of the PI model can be effectively modeled in ENA. This alignment will allow researchers to validate whether the observed patterns in the data reflect the underlying cognitive processes proposed by the theory.

1.3 Research Context

In this study, 108 pre-service teachers from a public university were divided into 16 groups to discuss "Artificial Intelligence and Applications in Education." Half of the groups used a GPT-4 chatbot, while the others did not. The 40-minute discussions took place on a digital chat platform, and students' interaction data was collected afterward. Surveys measured academic motivation, self-efficacy, situational engagement, and cognitive load. The discussion data (1,617 messages) was coded based on the CoI framework. Three coders achieved strong agreement (Krippendorff's alpha = 0.81) on an initial 20%, resolved discrepancies, and then coded the rest (Marzi et al., 2024).

1.4 A Customized Epistemic Network Analysis Rotation

We used the rENA R package to construct an ENA model aligned with the PI framework (Marquart et al., 2023). To represent the PI model, we customized the rotation by using a regression-based rotation instead of the standard SVD, creating a two-dimensional space aligned with the PI model's axes. The first axis was based on Behavioral and Social Engagement, capturing social abilities, while the second axis was based on Cognitive Engagement and Course Self-efficacy, reflecting cognitive engagement. This rotation ensured alignment with the PI model, enhancing interpretability. The ENA model units were set at the class, group, and student levels, with conversations categorized by class and group. The window size encompassed the entire conversation, and the selected codes—Triggering Event, Exploration, Integration, and Resolution—represented the four-phase progression.

2 PRELIMINARY FINDINGS

As shown in Figures 1(a) and 1(b), the overall mean networks reveal that while both groups made strong connections with Exploration, the group using ChatGPT connected Exploration more with Integration, whereas the group without ChatGPT connected it more with Triggering Event. This suggests that, with assistance from ChatGPT, students were more inclined to integrate their exploratory findings into subsequent steps.

Further, Figure 1(c) illustrates the network of Student A, who displayed low cognitive and social engagement, showing limited connections primarily between Triggering Event and Exploration. In contrast, Student B, in Figure 1(d), with higher social engagement than Student A, established additional connections between Exploration and Integration. Similarly, Student C in Figure 1(e), exhibiting high cognitive and moderate social engagement, progressed further by making more connections between Exploration and Integration than Student B, indicated by a thicker edge. Finally, Student D, rated high in both social and cognitive engagement, was the only one to reach the Resolution phase, as shown by the connecting edge.

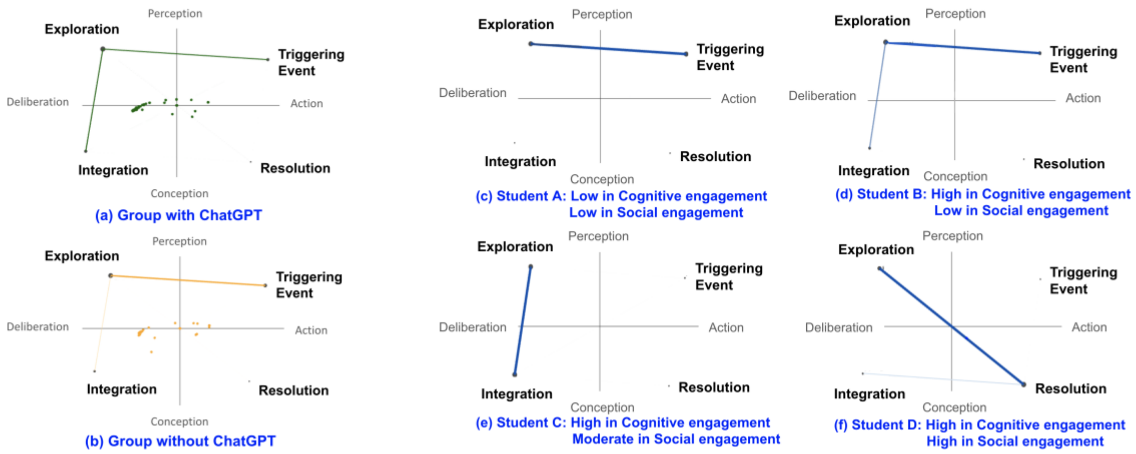


Figure 1: (a) ENA network for the group with ChatGPT; (b) ENA network for the group without ChatGPT; (c, d, e, f) ENA network for four representative individual students

3 DISCUSSION

This study shows that the customized rotation visualizes students' progression through inquiry phases. A pattern emerges where lower cognitive engagement correlates with more connections in Triggering Event and Exploration, while higher cognitive and social engagement aligns with Integration and Resolution. Future work could explore individual trajectories to further understand students' inquiry.

REFERENCES

Ba, S., Hu, X., Stein, D., & Liu, Q. (2024). Anatomizing online collaborative inquiry using directional epistemic network analysis and trajectory tracking. *British Journal of Educational Technology*, 55, 2173–2191. https://doi.org/10.1111/bjet.13441

Kim, G. C., & Gurvitch, R. (2020). Online education research adopting the community of inquiry framework: A systematic review. *Quest*, *72*(4), 395-409. https://doi.org/10.1080/00336297.2020.1761843

Marquart, L. C, Swiecki, Z., Collier, W., Eagan, B., Woodward, R., Shaffer, W. D (2023). *rENA: Epistemic Network Analysis*. 0.2.7, https://cran.r-project.org/package=rENA.

Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha Calculator–Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. *MethodsX*, *12*, 102545. https://doi.org/10.1016/j.mex.2023.102545

ReflectionApp: Scaffolding Student Reflection with AI

Gerti Pishtari[1], Alejandro Ortega-Arranz[2], Edna Milena Sarmiento-Márquez[3], Paraskevi Topali[4], Marlene Wagner[1], Tobias Ley[1,3]
[1]Danube University Krems, [2]Universidad de Valladolid, [3]Tallinn University, [4]NOLAI | National Education Lab AI, Radboud University
[1]gerti.pishtari@donau-uni.ac.at

ABSTRACT: Reflection is crucial for learning, enabling learning transfer, and developing transversal skills. Conversational agents are often used to scaffold reflection. Despite their benefits, agents may fail to boost students' engagement over time. Advances in Artificial Intelligence (AI) now offer potential for more interactive agents. We introduce ReflectionApp, an AI-based agent designed to scaffold student reflection on project challenges and investigate its impact on university students' reflection quality. Using ReflectionApp improved reflection quality in content, across dimensions like reporting, relating, and reasoning, as well as in depth. However, students found it challenging to reflect on future learnings.

Keywords: Student Reflection, Conversational Agent, Artificial Intelligence, Higher Education

1 INTRODUCTION

Reflection on experiences, emotions, or actions is key to learning, linking new and prior knowledge (Chan & Lee, 2021), and enabling transfer of learning to real-world situations. Through reflection, students build transversal skills like critical thinking, self-awareness, and problem-solving (Chan & Lee, 2021). Technology can scaffold the process by nudging students to reflect (e.g., through reminders), facilitating sharing and peer discussions on reflection outcomes, or by providing writing analytics (Buckingham Shum et al., 2016, Wolfbauer et al., 2022). Conversational agents have been used to guide reflection with sequential, predefined prompts/questions. While effective in enhancing reflection skills, student engagement with agents may decline over time (Wolfbauer et al., 2022), likely due to the perception of fewer new insights from prompts that do not adapt to responses.

Artificial Intelligence (AI), especially Large Language Models (LLMs), can enable agents that foster an interactive experience. Yet, the impact of AI agents on student reflection, how efficiently AI guides the process, and students' perceptions of AI agents remain open questions. To investigate, we use ReflectionApp[1], an LLM-based agent that supports student reflection on project-related issues. ReflectionApp prompts students with three questions (Q) adapted from Driscoll (1994): *Q1. What happened, and how did you address it? Q2. Why do you think your approach did not work? Q3. What lessons did you learn for the future?* Each question has predefined criteria that responses should match (see Figure 1), which are assessed by GPT-4o (with a Temperature=0 value for consistency). In case of unmet criteria, the model generates up to two follow-up questions per main question to prompt further details on them. This study explores the following research question: **How does ReflectionApp impact the content and depth of student reflections?**

[1] ReflectionApp source code: https://github.com/gertipishtari/ReflectionApp, last accessed in December 2024.

2 METHODOLOGY

We recruited students from the "Research in Human-Computer Interaction" masters' course at a Spanish university. In total, 8 students (5 male, 3 female) agreed to participate. The study spanned two phases, each during a lecture happening after students had submitted a corresponding (i.e. different) assignment on designing and evaluating a human-computer interaction system. In **Phase 1**, students (a) filled out a demographic questionnaire, (b) responded to a form with the same three open-ended questions about an assignment-related problem as in ReflectionApp, and (c) wrote an essay on what they learned from the assignment (without further guidance). In **Phase 2**, students (a) used ReflectinApp to answer the questions, and (b) wrote a similar essay.

To address our research question (tool impact), we manually assessed the quality of responses to the three questions in Phase 1, using the predefined criteria (see Figure 1). Additionally, we manually analyzed student responses using ReflectionApp in Phase 2, comparing our results with the automatic GPT-4os' classifications (see Introduction). We also compared essay quality across phases using content analyses based on the 5R framework (that examines if refection composition includes Reporting, Responding, Relating, Reasoning, and Reconstructing) (Bain et al., 2002) and the Reflection Continuum that classifies the depth of a reflection into four hierarchical categories: Non-Reflection, Understanding, Reflection, or Critical Reflection (Kember et al., 2008). Two of the authors of this paper conducted the analyses, discussing each case until reaching an agreement.

3 RESULTS AND DISCUSSION

AI-generated follow-up questions encouraged students to provide more detail, deepening their reflection. Student responses met more criteria when using ReflectionApp, largely due to the follow-up questions in Phase 2 (Figure 1). Essays in Phase 2 scored higher (Figure 2, left), but Reconstructing, the part of the 5R framework focusing on planning future actions, received a low score. Notably, the depth of the reflections improved significantly in Phase 2 essays, with all essays reaching Reflection or Critical Reflection levels on the Reflection Continuum (Figure 2, right). However, students still struggled to reflect on future learnings, as indicated by the progressive scores for Q3 and its follow-up questions (Figure 1) and a lower score in Reconstructing (Figure 2).

Question	Criteria	Phase 1	Phase 2	Main question	Follow-up 1	Follow-up 2
Q1	1.1. Problem identification	7 (87.5%)	8 (100%)	8 (100%)	/	/
	1.2. Understanding the context of the problem	0	8 (100%)	4 (50%)	3 (37.5%)	1 (12.5%)
	1.3. Steps taken to address the problem	6 (75%)	8 (100%)	8 (100%)	/	/
Q2	2.1 Actions that did not produce desired results	2 (25%)	7 (87.5%)	2 (25%)	4 (50%)	1 (12.5%)
	2.2. Internal missteps, errors, judgements	6 (75%)	8 (100%)	4 (50%)	3 (37.5%)	1 (12.5%)
	2.3. External factors that influenced	1 (12.5%)	8 (100%)	4 (50%)	4 (50%)	/
Q3	3.1 Key lessons learned	4 (50%)	8 (100%)	7 (87.5%)	0	1 (12.5%)
	3.2. Future application of learned lessons	4 (50%)	8 (100%)	6 (75%)	2 (25%)	/
	3.3. Specific improvements to be implemented	2 (25%)	8 (100%)	4 (50%)	4 (50%)	0
	3.4. Measures to avoid similar failures	2 (25%)	8 (100%)	4 (50%)	3 (37.5%)	1 (12.5%)
	3.5. Personal growth from this experience	2 (25%)	6 (75%)	1 (12.5%)	3 (37.5%)	2 (25%)

Figure 1: Number of students achieving response criteria across phases (N=8), with Phase 2 broken down by initial response (Main) and subsequent follow-up prompts

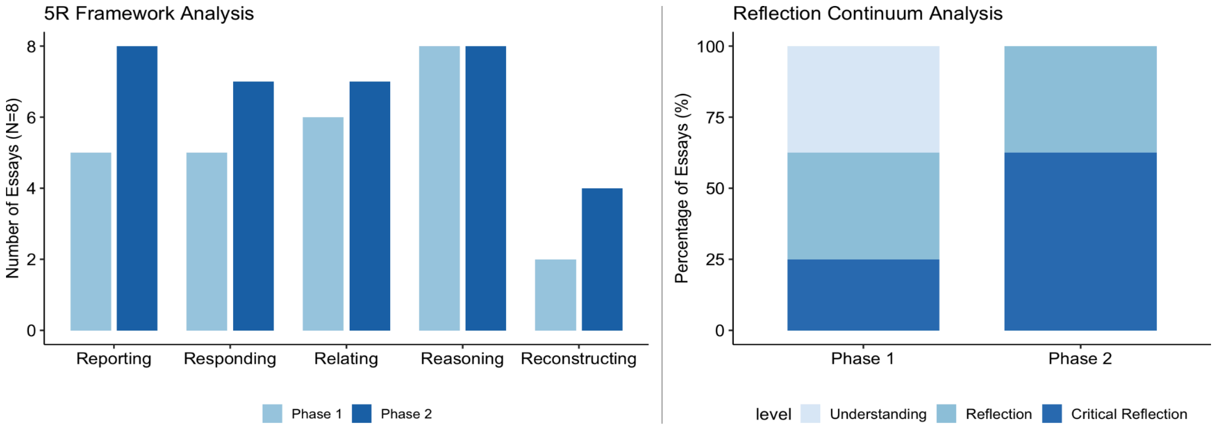


Figure 2: Essay quality using the 5R Framework (left) and the Reflection Continuum (right)

While larger longitudinal studies are necessary, these results suggest that follow-up questions alone may be insufficient to support reflection across multiple dimensions, indicating a need for more structured guidance. Implementing theory-based systems could enhance this structure, as we partially do by scoring follow-ups with specific criteria. Frameworks like the 5R model (Bain et al., 2002) could further aid this process. Integrating AI reflection agents with writing analytics (Buckingham Shum et al., 2016) could also improve students' reflection skills. Providing these analytics to teachers would help them support students and adapt courses to their needs. Future research will explore how sustained tool use impacts essay quality, reflection skills, and learning outcomes across diverse contexts, as well as how individual traits (e.g., academic performance, gender) influence reflection quality. While it is not the focus of this paper, future work will examine how accurately (vanilla and fine-tuned) LLMs can code student reflections using coding schemes deriving from related theoretical frameworks. Additionally, we will explore student perceptions of ReflectionApp, in terms of support, quality of follow-up questions it generates, and usability.

REFERENCES

Bain, J.D., Ballantyne, R., Mills, C., & Lester, N.C. (2002). Reflecting on practice: Student teachers' perspectives. Flaxton, QLD: Post Pressed.

Buckingham Shum, S., Sándor, Á., Goldsmith, R., Wang, X., Bass, R., & McWilliams, M. (2016, April). Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 213-222).

Chan, C. K., & Lee, K. K. (2021). Reflection literacy: A multilevel perspective on the challenges of using reflections in higher education through a comprehensive literature review. *Educational Research Review*, *32*, 100376.

Driscoll J. (1994). Reflective practice for practise. Senior Nurse, 13, 47 -50

Kember, D., McKay, J., Sinclair, K., & Wong, F. K. Y. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & evaluation in higher education*, *33*(4), 369-379.

Wolfbauer, I., Pammer-Schindler, V., Maitz, K., & Rosé, C. P. (2022). A script for conversational reflection guidance: a field study on developing reflection competence with apprentices. *IEEE Transactions on Learning Technologies*, *15*(5), 554-566.

Analysis of Individual Differences in Leg Movement Measurement for Multimodal Learning Analytics

Tatsuya Hamada, Masaki Kodaira, Tatsuya Ueda and Hironori Egi
The University of Electro-Communications
tatsuya.hamada@mail.uec.jp, hiro.egi@uec.ac.jp

ABSTRACT: Recent studies have highlighted a correlation between learners' leg movements and their learning behaviors. Nonetheless, this relationship is influenced by individual differences. To enhance multimodal learning analytics, we investigated these individual variances in leg movement. The growing focus on adaptive learning support for diverse learners underpins the importance of such analytics. Our study specifically concentrated on leg movement measurements and their individual disparities. We identified a significant correlation between the patterns of leg movements during the first and second halves of mental arithmetic tasks. This finding underscores the possibility of predicting individual differences based on early leg movement patterns in a learning session. Conclusively, our research confirms the presence of distinct individual differences in leg movements among learners. Acknowledging and integrating these differences into leg movement measurements can substantially improve the effectiveness of multimodal learning analytics.

Keywords: Multimodal learning analytics; leg movement; individual difference

1 INTRODUCTION

In recent years, the need for learning support that accommodates the diversity of learners has intensified. Understanding learners' behaviors, such as fatigue, concentration, and engagement, is crucial for providing effective individualized support. Multimodal learning analytics, which analyze physiological indicators using sensors to estimate learner behaviors, have emerged as a key tool in this context. Goldberg et al. (2021) demonstrated the ability to gauge learners' cognitive engagement and involvement through classroom videos, noting a correlation between students' engagement and observable behaviors like hand-raising and question-asking.

However, the adaptability of multimodal learning analytics across various learning environments is a challenge. For example, in computer labs, monitoring learners with classroom videos is impractical due to obstructions like computer monitors. An alternative approach in such settings is measuring leg movements. Many learning activities, especially those requiring seated positions like tests, necessitate maintaining proper posture. Elvitigala et al. (2020) presented a method to detect stress using foot movement and posture characteristics, employing a pressure-sensitive insole to discern stress and relaxation states. Additionally, Aikawa et al. (2019) found a correlation between learners' mental fatigue or concentration and their leg movements.

This study introduces a novel approach using a sensor placed under desks to measure leg movement. This method is less invasive, not requiring learners to wear any devices, and remains unobtrusive during learning activities. However, we acknowledge that leg movements vary among individuals. Some learners may move their legs consistently during learning sessions, while others may not. This study aims to identify these individual differences in leg movement measurements to enhance the accuracy and effectiveness of learning support through improved multimodal learning analytics.

2 METHODS AND EXPERIMENTS

To measure learners' leg movements, we developed a specialized device. This device integrates a Raspberry Pi 4 (a single-board computer), an Arduino Nano Every (a microcontroller), and a passive infrared (PIR) sensor (model EKMC2609112K). The PIR sensor, designed to detect infrared light emitted by the human body, is strategically placed under the learners' desks. This setup allows for the detection of leg movements while learners are seated and engaged in learning activities. The other components of the device are responsible for processing the collected data. Figure 1 illustrates the setup of our leg movement measurement device.

Figure 1: Leg movement measurement device

Our experiment aimed to discern individual differences in leg movement measurements. We conducted the study with 60 university students from a science and engineering background. The participants were required to complete mental arithmetic tasks while seated. These tasks were chosen to consistently impose cognitive load on the learners. The tasks included 300 addition and subtraction questions, involving numbers up to three digits. These questions were presented sequentially on a screen, and responses were made via mouse click. Participants were instructed to solve the problems as quickly and accurately as possible. To maintain a continuous cognitive load, the difficulty level of the tasks was dynamically adjusted based on the accuracy of the learners' responses. Completing all questions took approximately 25 min.

3 RESULTS

In our analysis, we segmented the leg movement data collected during the mental arithmetic tasks into two phases: the first and second halves. This division was based on the assumption that learners would experience increased fatigue in the latter half. Figure 2 displays the average leg movement for each learner across these two periods. Each data point reflects the leg movement measurements, with lines representing the results of a single linear regression (R^2 = 0.788, $F(1, 58)$ = 216.1, $p < 0.05$). The "leg movement amount" refers to the frequency of leg movements within a given time frame. Our findings highlight distinct individual differences in leg movements. Most learners exhibited a noticeable escalation in leg movement in the second half compared to the first. The correlation coefficient between the leg movements in the two halves was 0.89, indicating a strong association. This correlation suggests the feasibility of predicting leg movement during periods of fatigue by initially measuring leg movements. Consequently, early measurement of leg movements allows for the adaptation to individual differences among learners. In summary, this study not only identified individual variances in leg movement but also proposed the potential of enhancing learning support by considering these differences. Initial measurement of leg movements can provide valuable insights for personalized learning interventions. Our future work will focus on further data analysis and developing an estimation model. This model aims to contribute to more tailored and efficient multimodal learning analytics.

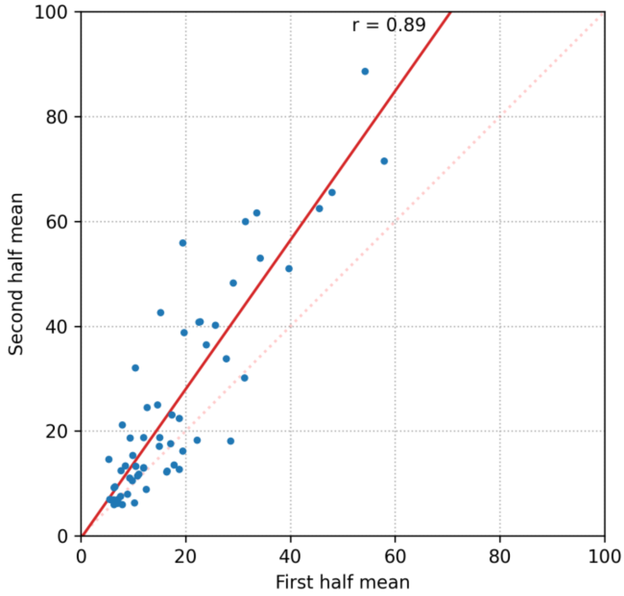


Figure 2: Means of leg movement amount for learners

ACKNOWLEDGEMENTS

This work has been partly supported by the Grants-in-Aid for Scientific Research (NOs. 21K18496 and 24K21451) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan.

REFERENCES

Aikawa, D., Asai, Y., Egi, H. (2019). Proposing an Estimation Method of Mental Fatigue by Measuring Learner's Leg Movement. In P. Zaphiris, A. Ioannou (Eds.), *Learning and Collaboration Technologies. Designing Learning Experiences. HCII 2019 (Vol. 11590).* Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-030-21814-0_17

Elvitigala, D. S., Matthies, D. J. C., & Nanayakkara, S. (2020). StressFoot: Uncovering the Potential of the Foot for Acute Stress Sensing in Sitting Posture. *Sensors, 20*(10), 2882. http://dx.doi.org/10.3390/s20102882

Goldberg, P., Sümer, Ö., Stürmer, K. et al. (2021). Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction. *Educational Psychology Review, 33*, 27–49. https://doi.org/10.1007/s10648-019-09514-z

The Promise of Learning Analytics for Improving Learning Design: A Systematic Review

Yukiko Maeda
Purdue University
ymaeda@purdue.edu

Jennifer Richardson
Purdue University
jennrich@purdue.edu

Masanori Yamada
Kyushu University
mark@mark-lab.net

Xiyu Wang
Purdue University
wang5608@purdue.edu

Xuewang Geng
Kyushu University
geng@mark-lab.net

Basori
Purdue University
fbasori@purdue.edu

Seyedeh Fatemeh Dadashipour
Purdue University
fatemeh@purdue.edu

Yoshiko Goda
Kumamoto University
ygoda@kumamoto-u.ac.jp

Kharon Grimmet
Purdue University
grimmet@purdue.edu

ABSTRACT: Of the types of actionable outcomes that learning analytics (LA) has been applied to there is limited research on the implementation of LA for the purpose of modifying learning design in higher education courses in blended or online settings. Faculty-designers or learning designers interested in using LA to enhance course design often face limited agency about the conception, design, and implementation of LA. Consequently, LA is rarely integrated into current course design practices. This systematic review aims to evaluate the integration of LA into Learning Design to facilitate data-driven (or evidence-based) decision-making in designing online and blended instruction. We will discuss the theoretical frameworks and models for integrating LA into learning design used to make pedagogical connections; types of data collected, and the practical actionable outcomes of learning design modification(s) that have been or will be implemented as a result of the LA.

Keywords: Learning analytics, learning design, instructional design, higher education, blended learning, online learning, systematic review

1 BACKGROUND

Learning Analytics (LA) needs to have actionable outcomes to be effective, in higher education environments many times those outcomes are related to retention or intervention systems that use predictive analytics for student success at the program or course level (Yan, et al., 2021). However, it is often the case that those outcomes do not provide practical implications related to learning design (Leitner, et al., 2017; Wood, 2023). Further, even when faculty-designers or learning designers are interested in the use of LA to enhance course design they often have limited agency about "how LA is conceived, designed, and deployed" (Wood, 2023, p. 32). If they have access to the technologies and systems related to LA at their institution there is the compounded factor of knowing the types of data that could be useful or how to request access to particular LA data, especially since LA is not a common consideration in current course design practices connection (Yan, et al., 2021). A result of limited agency and access can ultimately lead to challenges related to the data collection process, such as organization of LA and misalignment of timing between data collection and need for data (Wise & Jung, 2019) and pedagogical connection (Leitner, et al., 2017; Wood, 2023).

1.1 Research Purpose

This systematic review aims to evaluate the integration of LA into Learning Design to facilitate data-driven (or evidence-based) decision-making in designing online and blended instruction. Specifically, from existing literature, we seek to understand how LA informs the effectiveness of the learning and teaching process through instructional strategies, curricular design, and learning design of content and/or course design based on the collection and analysis of data from students, faculty, and other potential sources from where learning occurs. More specifically the research questions are:

1. What theoretical frameworks and models for integrating LA into learning design have been used to make pedagogical connections?
2. What types of data are collected to integrate LA with learning design?
    a. What is the nature of the data being collected to incorporate learning theories and insights from key stakeholders? (e.g., data about Faculty? Learners? Learning designers?)
3. What learning design modification(s) have been or will be implemented as a result of the LA?

Ultimately this research study will allow us to better understand how we can empower stakeholders, faculty-designers, and learning designers, in implementing LA at the course level for blended and online instruction.

2 METHODS

The target study for this systematic review are empirical studies that report on the integration of LA into instructional design decision-making. Our review is limited to studies reported from 2011 onward as this marks the year of the first international Conference on Learning Analytics and Knowledge (LAK). In addition, the 2011 Horizon Report (Johnson et al. 2011) highlighted LA as one of the six emergent technologies expected to have widespread adaptation in higher education in the next four to five years. Since then, LA has gained increased attention for its potential impacts on and use in teaching and learning.

2.1 Inclusion and Exclusion Criteria

We established a set of inclusion and exclusion criteria to define the scope of this systematic review. Studies were included if they met the following criteria: (1) the study was conducted on online or blended learning courses in higher education; and (2) the study applied LA to support learning design. Studies were excluded if they did not explicitly focus on the use of learning analytics with an outcome of modifying learning design within higher education.

2.2 Study Identification and Full Screening Process

To identify the relevant studies, the research team first conducted a brainstorming session to generate potential keywords. We also reviewed keywords used in previous relevant systematic reviews. Three researchers, two of whom have extensive experiences conducting systematic reviews and meta-analysis, piloted keyword combinations and search strings with multiple databases, including EBSCOhost, Web of Science, Scopus, Science Direct, PubMed, and ProQuest. Through iterative testing and refinement, we finalized our search strings with a combination of Boolean operators and keywords such as learning analytics, educational data analytics, big data analytics, teaching analytics, instructor analytics, instructional design, learning design, curriculum design, course design, course improvement, course evaluation, instructor's feedback, recommendation systems, decision-making, instructional design, and personal learning. Note that search string specifications were adjusted based on the requirements of each database we searched. After removal of duplicated studies, a total of 3568 were identified for title and abstract screening. All of the authors have been involved in the screening process. At least two reviewers were assigned to each of the studies for screening. The inconsistency of the screening was resolved by the third reviewer whose expertise is either LA or instructional design. Of those 234 were further reviewed at the stage of full screening. For the full screening process again involved at least two reviewers with a third reviewer available for inconsistencies or conflicts. For the full screenings, we obtained full texts and determined if the study fit into our inclusion criteria. Through this process, we identified 34 studies eligible to include the current synthesis.

2.3 Data Extraction

The coding protocol was drafted by two of the co-authors to extract the relevant information to address research questions. These two co-authors also referred to the coding protocols utilized in several relevant systematic reviews, including Drugova et al. (2023) and Hase & Kuhl (2024) to develop the draft. Then the draft was modified with pilot coding of randomly selected three studies from the final pool of the eligible studies for the current synthesis. After coding consistency is established among coders, a pair of coders are assigned to each study to extract relevant information from the document individually. Discrepancies in coding will be resolved with a third coder. We will report our preliminary findings at the presentation.

REFERENCES

Drugova, Elena & Zhuravleva, Irina & Zakharova, Ulyana & Latipov, Adel. (2023). Learning analytics driven improvements in learning design in higher education: A systematic literature review. Journal of Computer Assisted Learning. 40. n/a-n/a. 10.1111/jcal.12894.

Hase, A., Kuhl, P. Teachers' use of data from digital learning platforms for instructional design: a systematic review. *Education Tech Research Dev* 72, 1925–1945 (2024). https://doi.org/10.1007/s11423-024-10356-y

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—A literature review. In *Learning analytics: Fundaments, applications, and trends* (pp. 1–23). Springer, Cham.

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. https://doi.org/10.18608/jla.2019.62.4

Wood, E. (2023) *Instructors' lived experiences of using learning analytics to improve online course design*. [Doctoral dissertation, Northeastern University]. Northeastern University Digital Repository Service. https://doi.org/10.17760/D20483502

Yan, H., Lin, F. & Kinshuk. (2021). Including learning analytics in the loop of self-paced online course learning design. *International Journal of Artificial Intelligence,* 31, 878–895. https://doi.org/10.1007/s40593-020-00225-z

Bridging the Digital Divide: Gamification and AI in Higher Education Learning Analytics

Martha Hubertz
University of Central Florida
Martha.Hubertz@UCF.edu

Alisha Janowsky
University of Central Florida
Alisha.Janowsky@UCF.edu

Elizabeth Hubertz
WashU Law
ejhubertz@wulaw.wustl.edu

ABSTRACT: This study investigates the intersection of gamification, artificial intelligence (AI), and learning analytics in higher education, with a particular focus on digital equity (Selwyn & Facer, 2021). Drawing on Self-Determination Theory (SDT), we examine how gamification elements differentially support autonomy, competence, and relatedness based on students' technological access and comfort levels. Using comprehensive technology adoption data and implementation results from gamified course elements (N = 1,183), we examine how varying levels of technological access and AI comfort influence student engagement in gamified learning environments. Our findings reveal significant patterns in technology adoption and demonstrate how learning analytics can guide the optimization of gamification approaches to address digital disparities and create inclusive learning environments that accommodate diverse student populations and varying levels of technological access (Reich & Ito, 2017).

Keywords: Gamification, Learning Analytics, Digital Divide, AI, Higher Education, Engagement

1 Background and Motivation

The increasing integration of technology in higher education necessitates understanding of how learning analytics can inform equitable implementation of digital learning tools (Bond et al., 2020). This study addresses gaps in recent literature (Rodrigues et al., 2022) by investigating how learning analytics can optimize gamification and AI implementation, analyzing student responses to different gamification formats, measuring AI integration impacts (Holstein et al., 2020), and developing guidelines that address digital equity concerns (Selwyn & Facer, 2021).

1.1 Gamification and Self-Determination Theory (SDT) Impact on Active Learning

Gamification has shown promise in promoting active learning and increasing student engagement across various educational contexts. By incorporating elements such as points, leaderboards, badges, and levels, gamified learning environments can transform passive learning experiences into interactive, challenging, and rewarding activities (Mustafa & Karimi, 2021; Bovermann & Bastiaens, 2020). This approach aligns with the principles of active learning that emphasize student interaction with course content. Mustafa and Karimi (2021), for example, have shown that gamified online learning can significantly influence students' experience, engagement, and course completion rates.

SDT provides a crucial framework for understanding motivation in gamified learning environments by identifying three fundamental psychological needs (Ryan & Deci, 2020). SDT's three core psychological needs - autonomy, competence, and relatedness - naturally align with effective gamification elements as demonstrated through our learning analytics data. Autonomy is supported through choice in learning paths and engagement methods. Learning analytics can be used to track how different student populations engage with choice-based elements (e.g., multiple learning paths, optional challenges). Competence is developed through progressive skill development and achievement tracking. Relatedness is fostered through collaborative challenges and peer interaction.

Research Question #1: How can learning analytics inform the design of gamification elements to better support SDT needs across diverse student populations?

1.2 Technology Access and Digital Equality

The Technology Acceptance Model (TAM) provides a theoretical framework for understanding how users come to accept and use technology (Davis, 1989). TAM posits that a user's intention to use and actual use of technology is influenced by 1) perceived usefulness (PU) – how much a person believes using a particular system would enhance performance, and 2) perceived ease of use (PEOU) – how much a person believes using the system would be free of effort. In educational settings, TAM provides valuable insights into how students evaluate and engage with new learning technologies, making it particularly relevant for understanding the adoption of gamified learning systems and AI in education.

Research Question #2: How do varying levels of technological access and AI comfort influence students' engagement with gamified learning environments?

2 METHODOLOGY

2.1 Participants

1,183 students were recruited through the university's SONA research participation system at a large R1 Hispanic-Serving Institution in the southeastern United States during Fall 2022. Most participants (99.57%) were undergraduates (mainly freshman (52.78%) and sophomores (21.87%)), female (57.22%), with an age range of 18-54 years (M = 19.48, SD = 3.44). Participants identified as White (71.89%), Asian (11.79%), Black or African American (11.19%), and other racial categories (4.26%).

2.2 Procedure

This study was administered through an online Qualtrics survey in which participants answered demographic questions and completed several standardized measures assessing attitudes, beliefs, and experiences related to technology use and artificial intelligence. Participants received course credit for completing the survey. Engagement metrics from gamified course elements and comparative analysis of different gamification approaches were also used.

2.3 Measures

Participants reported their experience (e.g., first age of use, current usage, etc.) with various technologies (e.g., smartphones, laptops, and video game consoles). Perceptions of AI were assessed with open-ended questions, GAIS Attitudes Scale, GAIS Comfort Scale, and Trust in Technology scale.

3 PRELIMINARY FINDINGS

3.1 Technology Adoption and Autonomy

- Early technology adoption patterns: gaming devices (mean age: 9.20 years), smartphones (12.22 years), computers (13.90 years).
- Technology access suggests baseline autonomy in choosing learning tools.
- Learning analytics reveal significant differences in AI comfort levels between first-generation and continuing-generation students (23% gap), indicating potential barriers to autonomy.

3.2 Engagement Patterns, Competence, and Social Connections

- Gamified elements improve engagement (18.20%) when controlling for access.
- 76% of students view AI applications as beneficial.
- Learning analytics identifies correlations between early gaming device adoption and engagement with competitive gaming elements (+24%).
- First-generation students demonstrate stronger engagement with collaborative features when AI comfort barriers are addressed (+28%).
- Later computer adopters show higher engagement with self-paced elements (+31%).
- Trust metrics reveal ways to strengthen relatedness with transparent AI implementation.

REFERENCES

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. International Journal of Educational Technology in Higher Education, 17(1), 1-30.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340.

Holstein, K., McLaren, B. M., & Aleven, V. (2020). Co-designing AI-powered orchestration tools for K-12 classrooms. *International Journal of Artificial Intelligence in Education, 30*(4), 555-577.

Mustafa, A. S. & Karimi, K. (2021). Enhancing gamified online learning user experience (UX): A systematic literature review of recent trends. In N. Thakur & B. D. Parameschachari (Eds.), Human-Computer Interaction and Beyond – Part 1 (pp. 74-99). Bentham Science Publishers.

Reich, J., & Ito, M. (2017). *From good intentions to real outcomes: Equity by design in learning technologies*. Digital Media and Learning Research Hub.

Rodrigues, P., Bidarra, J., & Carvalho, T. (2022). Gamification in higher education: A systematic mapping study. *Education Sciences, 12*(1), 44.

Selwyn, N. & Facer, K. (2021). Digital technology and social justice: Research priorities for the post-pandemic era. *Learning, Media, and Technology, 46(2)*, 117-125.

MetaMap: a Conceptual Framework to Support Metacognitive Skills Development in Digital Environments

Edna Milena Sarmiento-Márquez[1], Gerti Pishtari[2], Paraskevi Topali[3], Alejandro Ortega-Arranz[4], Reet Kasepalu[1], Kairit Tammets[1]

[1] Tallinn University, [2] Danube University Krems, [3] NOLAI | National Education Lab AI, Radboud University, [4] Universidad de Valladolid

[1]msm@tlu.ee

ABSTRACT: Metacognitive skills are crucial for student success in complex tasks, needing explicit teaching through carefully designed instructional practices. While technology can support teachers in the design of these practices, creating tools that are theoretically sound, pedagogically meaningful, and feasible remains a challenge. As previous research animates collaboration between multiple stakeholders in designing tools for metacognitive skills, this paper introduces the MetaMap framework as a strategy to support the reflection in such collaboration. MetaMap builds on the learning design cycle and past research to foster stakeholder reflection on designing tools that promote and assess students' metacognitive skills. A preliminary assessment of MetaMap suggests areas for improvement, emphasizing the importance of adapting the framework to diverse stakeholder backgrounds.

Keywords: Metacognitive Skills, Learning Design, Reflection Framework, Digital Learning.

1 INTRODUCTION

Metacognitive skills (MS) help individuals think about, monitor, and regulate their learning processes, which is essential for sustained learning (Hamza et al., 2022). Yet teachers and curriculum designers often overlook MS (Ellis et al., 2014). Since MS involves internal processes that are difficult to measure, teaching and tracking MS require careful design. Learning Analytics (LA) and Learning Design (LD) could support MS' design and instruction: LA offer insights into student learning that might be linked to MS, while LD provides context for interpreting these insights (Mangaroska & Giannakos, 2019). Existing LA tools often lack theoretical grounding and prioritize technical over pedagogical value (Topali et al., 2023), factors especially important when addressing MS. Existing LA and LD guidelines give general advice about learning but fail to address the nuances of MS. Though prior research in LA has focused on student-centered analytics and the assessment of self-regulated learning (Law & Liang, 2020; Fan et al., 2023), there is a lack of guidance when designing LA tools that align learning theories, pedagogical goals, and technological capacities to support stakeholders in addressing MS effectively. As a first step in this direction, we propose the MetaMap framework (see Figure 1) and the LD rubric (https://bit.ly/LD4MS), grounded in (1) the LD cycle (Pishtari & Rodríguez-Triana, 2022), (2) the MS teaching principles (Ellis et al., 2014) and (3) a method for LA-integrated LD (Law & Liang, 2020).

1.1 The MetaMap Framework

MetaMap is conceived as a feedback loop involving reflection and improvement through a set of questions derived from the aforementioned principles and methods. It ensures that LA tools remain relevant and effective over time by articulating theoretical grounds, technical elements, and pedagogical stands. The framework involves three phases: *1) Planning and setup*: Teachers and researchers establish a theoretical foundation for MS by setting metacognitive goals, selecting learning activities, and identifying suitable analytics to capture MS. We have designed the LD rubric to organize all the relevant information about the LD. Developers then set up the digital environment. Developers support the setup of the digital environment. 2) Interpretation: Developers and researchers focus on data collection and LA tool quality. Developers ensure data alignment with pre-defined metrics and transfer it to researchers, who interpret the data concerning MS, being guided by the framework items. This phase may yield evidence-based suggestions to enhance LD and data collection. *3) Analysis*: Researchers and teachers evaluate LA insights resulting from the interpretation phase to assess the tools' effectiveness in supporting MS. This reflection considers the trustworthiness of the LA, their actionability and impact, and the usability of the dashboard. Roles in the MetaMap Framework were assigned based on stakeholders' expertise, ensuring efficient collaboration where their contributions are most relevant (Alfredo et al., 2024).

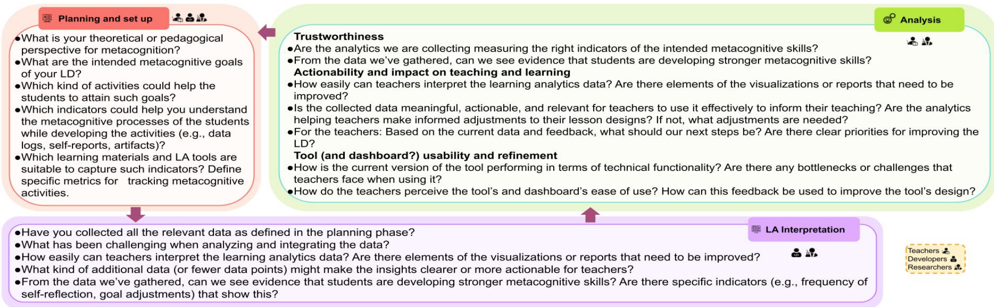


Figure 1. MetaMap Framework: reflection questions

2 METHODOLOGY

A preliminary evaluation of the framework was conducted with two teachers, selected based on a purposive sampling based on the availability of participants who designed MS-oriented activities in H5P. The study unfolded in two phases: first, a hands-on session where we introduced the framework, the teachers reviewed mockup dashboards from their previously designed H5P activities, and guided the redesign of an activity focused on a specific MS using the LD rubric. Next, we conducted semi-structured interviews to gather feedback on the teachers' experiences with the framework. We analyzed the transcripts using thematic analysis to address the research question: *How did teachers perceive the usefulness of the MetaMap framework to reflect on the design of LA solutions for MS?*

3 PRELIMINARY RESULTS AND DISCUSSION

Thematic analysis showed that teachers recognized the MetaMap's potential for facilitating communication with developers and guiding task design to support MS. Teachers' experience was shaped by their prior knowledge of MS and digital literacy. Teachers were not familiar with MS concepts and theories or the use of learning indicators before creating H5P activities. This further supports the need for multi-stakeholder collaboration to integrate a variety of expertise for MS-based activity design (Mangaroska & Giannakos, 2019). More specifically, teachers may benefit from the exchanges with researchers to improve their understanding of learning theories, while collaborating with developers may help them to understand better the technical underpinnings of LA. Fine-tuning the language of the framework could better align it with stakeholder backgrounds.

> "I didn't know what to write as metacognitive goals, the word metacognition is there, but it is not that intuitive" (T1).

Despite challenges, teachers responded positively to the LD rubric, with one noting its potential for facilitating communication with developers as also noted by Law & Liang (2020). Another teacher recommended adding a column to clarify terms like "metacognitive goals" for greater understanding. Both valued the reflection questions for guiding task design to support MS. Although they did not use the mockup dashboard data to refine analytics or adjust learning design, likely due to time constraints, they recognized a need to help teachers interpret such data. They found the questions in each phase relevant, though one teacher was unsure about the developers' ability to reflect on them. Lastly, a teacher noted the framework's potential to engage them actively in teaching and evaluating MS in digital environments. These results indicate that while the framework enhances teacher agency, shifting them from consumers to co-creators of LA tools, it requires adjustments to suit stakeholders with non-research backgrounds. A key strength of the framework is its ability to foster communication among stakeholders, integrating theory, technology, and pedagogy in LA tool development. Future iterations will involve a broader range of stakeholders, including developers and researchers, the application of the framework in authentic settings, and expert assessment of the questions.

ACKNOWLEDGEMENT

This study was funded by the Estonian Research Agency (grant PRG1634).

REFERENCES

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 100215.

Ellis, A. K., Denton, D. W., & Bond, J. B. (2014). An analysis of research on metacognitive teaching strategies. *Procedia-Social and Behavioral Sciences*, 116, 4015-4024.

Fan, Y., Rakovic, M., van Der Graaf, J., Lim, L., Singh, S., Moore, J., ... & Gašević, D. (2023). Towards a fuller picture: Triangulation and integration of the measurement of self-regulated learning based on trace and think aloud data. *J. of Computer Assisted Learning, 39*(4), 1303-1324.

Hamzah, H., Hamzah, M. I., & Zulkifli, H. (2022). Systematic literature review on the elements of metacognition-based higher order thinking skills (HOTS) teaching and learning modules. *Sustainability, 14*(2), 813.

Mangaroska, K., & Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies, 12*(4), 516-534.

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

Pishtari, G., & Rodríguez-Triana, M. J. (2022). An analysis of mobile learning tools in terms of pedagogical affordances and support to the learning activity life cycle. In *Hybrid learning spaces* (pp. 167-183). Cham: Springer International Publishing.

Topali, P., Chounta, I. A., Martínez-Monés, A., & Dimitriadis, Y. (2023). Delving into instructor-led feedback interventions informed by learning analytics in massive open online courses. *Journal of Computer Assisted Learning, 39*(4), 1039-1060.

Memory vs. Attention: Understanding Their Roles in Deep Knowledge Tracing Architectures

Indronil Bhattacharjee
New Mexico State University
indronil@nmsu.edu

ABSTRACT: This paper presents a comparative analysis of memory-based and attention-based deep learning models in Knowledge Tracing (KT), which is essential for educational technologies that track and predict student learning progress. The study delves into models such as Dynamic Key-Value Memory Networks (DKVMN), Sequential Key-Value Memory Networks (SKVMN), Self-Attentive Knowledge Tracing (SAKT), and Context-Aware Attentive Knowledge Tracing (AKT). It explores how these models maintain persistent records of student interactions and dynamically assess past activities to improve prediction accuracy. The analysis also prepares the ground for future work aimed at adapting KT models to effectively handle new students without historical data, thereby enhancing the personalization and accessibility of educational technologies. This initiative is vital for evolving KT models to support diverse educational environments and individual learner needs.

Keywords: Deep Knowledge Tracing, Memory Networks, Attention Mechanism

1 INTRODUCTION

Knowledge tracing in educational technology refers to the modeling and prediction of evolving knowledge states of students over time, based on interactions with learning materials. Currently, there is an ever-growing demand from online education platforms for a machine system that can track and adapt to the knowledge of students. This work also provides a review of the different KT models, whether memory-based or attention-based, along with their architecture and performance on diverse datasets. KT aims to monitor the knowledge states of students and predict their performances for future exercises as part of a series of learning interactions. This is an essentiality to personalized pathways that maximize educational efficiency.

2 DEEP KNOWLEDGE TRACING (DKT)

Deep Knowledge Tracing (DKT) is another step forward for the application of neural networks to the KT problem by Piech et al., in 2015. Unlike the previous models, DKT explicitly focuses on sequential tracking of students' learning; hence, it utilizes Long Short-Term Memory (LSTM) networks for the patterns of how students answer questions. Keeping successful performance aside, a lack of interpretability has been realized for DKT, and hence it is hard to obtain insight from the underlying learning processes.

3 MEMORY-BASED DKT MODELS

Memory-based models, including Dynamic Key-Value Memory Networks (DKVMN) and Sequential Key-Value Memory Networks (SKVMN), aim to improve upon DKT's limitations by integrating memory-augmented neural networks.

(a) Dynamic Key-Value Memory Networks (DKVMN): DKVMN was proposed by Zhang et al., in 2017, involving a memory-augmented neural network architecture with a key-value pair for encoding the relations between concepts and student mastery states. The model uses two kinds of key-value memory systems, where keys indicate concepts being learned and values denote the student's mastery of the learned concepts. The input includes two key operations: Read and Write. While the

Read operation computes the correlation weights telling it which concepts are most relevant to an exercise a student is currently working on, the Write operation updates the memory based on the responses from students. DKVMN keeps track of a student's continuous learning curve regarding the introduction of new concepts and reviewing previously learned concepts. It makes this model more effective to use in a structured learning environment where the curriculum does not change much.

(b) Sequential Key-Value Memory Networks (SKVMN): SKVMN, introduced by Abdelrahman et al. in 2019, follows the structure of DKVMN but embeds advanced mechanisms that help capture long-term dependencies much better. Similar to DKVMN, SKVMN uses a key-value memory architecture and integrates Hop-LSTM in order to enable adaptive jumping between memory states. This makes the model skip over less relevant interactions and focus on more relevant ones, which is very important during temporal knowledge development. Its novelty comes in through its handling of temporal dependencies, hence providing a fine-grained look at the learning trajectory that a student takes. This improves the performance of the model in scenarios where previous interactions have much to say about the current and future learning.

4 ATTENTION-BASED DKT MODELS

Attention-based models, such as Self-Attentive Knowledge Tracing (SAKT) and Context-Aware Attentive Knowledge Tracing (AKT), use attention mechanisms to prioritize relevant historical interactions when predicting future performance.

(a) Self-Attentive Knowledge Tracing (SAKT): SAKT was proposed by Pandey and Karypis in 2019, which relies on the transformer architecture that uses self-attention mechanisms when estimating student interactions. This model tackles how to give dynamic weight to the importance of past interactions when predicting future performances. It operates based on no fixed structure but only estimates learner needs with regard to the relevance of past interactions. It identifies the important knowledge concepts, considering a student's interaction history, on which adaptive responses to different student needs are based. SAKT addresses more flexible knowledge tracking and adaptation to individual learning paths and is particularly suited for personalized learning environments. It focuses efficiently on relevant past interactions, and thus can largely improve prediction accuracy.

(b) Context-Aware Attentive Knowledge Tracing (AKT): AKT, developed later by Ghosh et al. in 2019, integrates further advanced attention mechanisms with contextual embeddings to further enhance knowledge tracking. AKT utilizes a monotonic attention mechanism that links a learner's future responses to past assessment interactions by weighing their past performance effectively. This is used in combination with Rasch model-based embeddings for an even finer-grained context-aware analysis. With a forgetting mechanism, AKT can simulate natural cognitive processes to adapt knowledge states with respect to the recency and frequency of interactions. This feature makes for a more realistic estimate of what the student can or cannot remember after a certain period of time.

Table 1: Comparative representation of different Deep KT models

Model	Type	Key Features	Specialty
DKVMN	Memory-Based	Key-value architecture, Read/Write operations	Effective for structured learning environments
SKVMN	Memory-Based	Hop-LSTM for long-term dependency, adaptive jumping	Nuanced understanding of temporal dependencies
SAKT	Attention-Based	Self-attention mechanism, flexible interaction weighting	Adaptable to personalized learning paths
AKT	Attention-Based	Monotonic attention, context-aware embeddings	Realistic assessments of knowledge retention

5 PERFORMANCE COMPARISON

The performance of various KT models has been evaluated across several datasets, including multiple versions of the ASSISTments dataset which features detailed student performance data on various mathematical skills.

Table 2: Performance comparison among different Deep KT models (AUC Score)

Model	ASSIST 2009	ASSIST 2012	ASSIST 2015	ASSIST 2017
DKT	0.7561	0.713	0.707	0.7263
DKVMN	0.8157	-	0.7268	0.707
SKVMN	0.8363	-	0.7484	0.717
SAKT	0.848	0.735	0.7212	0.734
AKT	0.8346	0.755	0.7828	0.7702

These AUC scores indicate that attention-based models like SAKT and AKT tend to outperform traditional memory-based models in most scenarios, highlighting their effectiveness in dynamic and adaptive learning environments.

6 FUTURE WORK

Moving forward, our research will focus on developing Knowledge Tracing (KT) models that quickly adapt to new students without prior data, utilizing minimal inputs for wider application in diverse educational settings. We will integrate transfer learning to get insights from established cohorts, enhancing learning outcomes for new groups and optimizing educational content personalization. Additionally, we plan to explore few-shot learning techniques, enabling our KT models to make reliable predictions with fewer data points, addressing challenges in data collection and privacy. These efforts aim to refine and personalize learning paths more effectively, broadening the accessibility and impact of educational technologies across various learner demographics and needs.

7 CONCLUSION

The comparative analysis of Knowledge Tracing models, particularly between memory-based and attention-based approaches, demonstrates a clear progression toward more sophisticated methods that better accommodate individual learning patterns. Future research endeavors will undoubtedly drive further innovations in this vital area of educational technology, enhancing the learning experiences of countless students worldwide.

REFERENCES

Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J. Guibas, and Jascha Sohl Dickstein (2015). Deep knowledge tracing. In NeurIPS. 505–513.

Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung (2017). Dynamic Key-value memory networks for knowledge tracing. In International World Wide Web Conference Committee (IW3C2). 765–774.

Ghodai Abdelrahman and Qing Wang (2019). Knowledge tracing with sequential key-value memory networks. In ACM SIGIR. 175–184.

Shalini Pandey and George Karypis (2019). A Self Attentive model for Knowledge Tracing. In Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019), Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.). 384 – 389.

Aritra Ghosh, Neil Heffernan, and Andrew S. Lan (2020). Context-aware attentive knowledge tracing. In ACM SIGKDD. 2330–2339.

Generative AI in Coding: Elevating Efficiency or Eroding Skills? Insights from an Indian User Study

Omkar Joshi[1], **Syaamantak Das**[2]
[1,2]Centre for Educational Technology, Indian Institute of Technology Bombay, Mumbai, India
[1]23d1202@iitb.ac.in, [2]syaamantak.das@iitb.ac.in

ABSTRACT: As Generative AI (GenAI) tools become increasingly prevalent in programming, understanding their impact on skills development in time-constrained coding (TCC) (e.g. competitive programming) is essential yet understudied. This study explores the perceptions of Indian users—including students, faculty, and professionals—regarding GenAI's influence on coding efficiency and problem-solving skills within TCC contexts. Using a mixed-methods approach, we surveyed 145 participants, revealing significant variations in perceived utility across roles and experience levels. Notably, 66.2% of participants found GenAI tools helpful or extremely helpful for TCC, emphasising gains in coding efficiency and learning outcomes. However, 29.66% of stakeholders have concerns that emerged around potential over-reliance on AI, risks to independent problem-solving skills, and challenges in maintaining academic integrity. These findings contribute valuable insights to the broader discourse on AI integration in CS education and hiring, underscoring the need for responsible use frameworks that balance AI assistance with skill-building in educational and professional settings.

Keywords: Generative AI (GenAI), Time-Constrained Coding Challenges (TCC), User Perceptions, AI in Education

1. INTRODUCTION & RESEARCH QUESTIONS

The rapid evolution of Large Language Model (LLM) based Generative AI (GenAI) tools has begun transforming how programmers approach coding tasks, offering support for efficiency and code optimization (Ebert and Louridas, 2023 Cui et. al 2024). These tools in the domain of Computer programming (e.g. Amazon Codewhisperer), which generate code based on input prompts, hold particular relevance in time-constrained coding challenges (TCC), where speed and accuracy are critical. However, as GenAI tools gain traction, questions arise regarding their impact on essential programming skills, such as independent problem-solving and critical thinking (Idrisov and Schlippe, 2024). In educational and professional settings, these skills are foundational, and their development could be hindered by over-reliance on AI-generated solutions (Petrovska et.al, 2024). This study aims to investigate the perceptions of Indian users—students, faculty, and professionals—regarding the role of GenAI tools in TCC. India's ubiquitous presence in the global technology sector, combined with the increasing use of AI-driven tools in both educational and corporate environments, presents a unique context for examining these dynamics. In particular, the study seeks to understand if users find GenAI tools helpful in enhancing performance and learning outcomes and to what extent they believe these tools may limit the development of critical problem-solving abilities.

1.1 Research Questions

1. How do different stakeholders perceive the utility of GenAI tools in TCC?

2. What are the potential impacts on learning outcomes and problem-solving skills?
3. How can we promote the responsible use of GenAI tools in coding assessments?

2. METHODOLOGY

2.1 Research Design

This study adopted a mixed-methods approach, combining quantitative survey data with qualitative thematic analysis to explore how Generative AI (GenAI) tools impact time-constrained coding challenges (TCC) from the perspective of Indian users. This design allowed for both a broad statistical overview and in-depth qualitative insights, capturing the varied experiences of students, faculty, and professionals regarding coding efficiency, problem-solving, and ethical considerations.

2.2 Participants

A total of 159 programmers participated in the survey which was developed by the authors and validated through SMEs, from which 145 responses were retained after excluding incomplete or irrelevant entries. The sample consisted of three key groups: students (n=78), faculty members (n=35), and industry professionals (n=32). The majority of students reported 1–3 years of programming experience, reflecting a cohort of newer programmers. Faculty members showed greater diversity in experience, with a significant portion having 10–15 years in the field, while professionals were concentrated around 3–6 years of experience.

2.3 Data Collection

Data were gathered through a structured online survey designed to capture a comprehensive view of participants' experiences with GenAI in TCC contexts. The survey included sections on GenAI tool usage, perceived impacts on performance, problem-solving, and ethical considerations. It also collected demographic details to contextualize responses. The survey combined Likert-scale questions with open-ended prompts, enabling both quantitative analysis and qualitative exploration. Recruitment took place through social media, email networks, and online forums.

2.4 Data Analysis

Quantitative data were analyzed using descriptive statistics to summarize demographics and responses. Correlation analysis was conducted to identify relationships between variables, such as GenAI usage and performance impact. ANOVA tests were applied to detect significant differences in perceptions based on roles and experience levels, revealing notable distinctions in perceived utility ($p < 0.05$) and interaction effects between role and experience ($p < 0.05$). Qualitative data from open-ended responses were analyzed through thematic coding, which identified key themes related to GenAI's perceived benefits and concerns. The thematic analysis highlighted recurring insights, such as GenAI's positive impact on efficiency and concerns over dependency, academic integrity, and the risk of diminishing problem-solving skills.

3. RESULTS

3.1 Quantitative Findings

To gauge the perceived effectiveness of GenAI in TCC, participants rated helpfulness on a five-point Likert scale ranging from "Not Helpful at All" to "Extremely Helpful." The mean helpfulness rating was 3.91, indicating an overall positive perception. A majority of respondents (66.2%) rated GenAI tools as either "Helpful" or "Extremely Helpful." However, 29.66% remained mostly neutral with some concerns, and only a small minority (4.1%) found GenAI to be unhelpful. This spread reflects general optimism about GenAI's utility in TCC, albeit with some reservations. An ANOVA analysis identified significant differences in perceived helpfulness based on role ($F = 6.18$, $p = 0.016$), suggesting that students, faculty, and professionals view GenAI's utility differently. Additionally, while experience alone did not significantly impact helpfulness ratings ($F = 1.41$, $p = 0.236$), the interaction between role and experience was significant ($F = 3.17$, $p = 0.014$), implying that perceptions may vary within roles depending on experience level. These results highlight role and experience based nuances in how GenAI is perceived as a coding aid in TCC.

3.2 Qualitative Insight

Many participants noted that GenAI significantly boosts coding efficiency, especially under time constraints, with 80% agreeing that it enhances code quality and streamlines tasks by offering quick optimization suggestions and reducing debugging times. However, there are concerns about potential dependency on these tools, as students and professionals alike worry that over-reliance on GenAI could hinder the development of independent coding skills and critical thinking, potentially stifling creativity, particularly among early learners. Faculty members also raised ethical concerns regarding academic misconduct, such as plagiarism, and the reduced development of skills, with 17 expressing the need for responsible use guidelines. Additionally, worries about unequal access to GenAI tools highlight the risk of exacerbating educational inequities.

4. DISCUSSION & INTERPRETATION OF FINDINGS

The quantitative result's positive outlook is largely attributed to the tools' ability to enhance coding efficiency by streamlining repetitive tasks, providing quick optimization suggestions, and supporting debugging processes. While this study offers valuable insights, it is important to acknowledge certain limitations. The reliance on self-reported data introduces the potential for response bias, and the findings may not be fully generalizable beyond the Indian context. Future research could benefit from longitudinal studies that examine the long-term impact of GenAI on coding skills development, as well as expanded samples across different cultural and educational contexts to enhance generalizability.

References

Ebert, C., & Louridas, P. (2023). Generative AI for software practitioners. *IEEE Software*, *40*(4), 30-38.

Cui, K. Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2024). The Productivity Effects of Generative AI: Evidence from a Field Experiment with GitHub Copilot. An MIT Exploration of Generative AI.

B. Idrisov and T. Schlippe (2024), "Program code generation with generative AIs," Algorithms, vol. 17, no. 2, p. 62.

Petrovska, O., Clift, L., Moller, F., & Pearsall, R. (2024, January). Incorporating Generative AI into Software Development Education. In *Proceedings of the 8th Conference on Computing Education Practice* (pp. 37-40).

Classification Method Using Topological Feature Extraction with Point-Generated Persistent Homology

Hitoshi Inoue
Nakamura Gakuen University
jinoue@nakamura-u.ac.jp

Koichi Yasutake
Hiroshima University
ystake@hiroshima-u.ac.jp

ABSTRACT: We propose a new approach to topological feature extraction using persistent homology, a representative method in Topological Data Analysis (TDA). Traditional TDA methods typically group data by label and extracting topological features for each group to use in learning; however, it is challenging to compute persistent homology for individual data points, limiting TDA's applicability in classification and prediction tasks. To address this issue, we propose Point-Generated Persistent Homology (PGPH). PGPH generates multiple sets of points from a single data point and applies persistent homology to these sets, enabling the extraction of topological features even from individual data points. This method enables the utilization of TDA-based features even for individual new data points, facilitating real-time applications and personalized predictions that were challenging with traditional approaches. While the experimental results did not improve classification accuracy, PGPH demonstrated potential as a TDA method applicable to real-time prediction and individual data. This paper validates the characteristics of PGPH through experiments on specific datasets and discusses its future application possibilities.

Keywords: Topological Data Analysis, Persistent Homology, Classification

1 INTRODUCTION

Classification and prediction in Learning Analytics are widely recognized as essential for improving the quality of education through tasks such as tracking learning progress, early detection of at-risk students, and providing personalized support. Supervised learning and deep learning techniques are commonly used for classification and prediction, achieving significant success with large-scale datasets. However, improving prediction accuracy remains crucial, as misclassification can negatively impact learning support.

To address challenges in analyzing the topological structures of datasets, Topological Data Analysis (TDA) has gained attention. TDA is effective for capturing the characteristics of complex structures and high-dimensional data (Edelsbrunner et al., 2002, and Munch, 2017). One of the most representative methods in TDA, persistent homology, extracts persistent topological features within data, which can then be used as features in machine learning. Typically, data are grouped by label, and the persistent homology for each group is used as a feature. However, directly computing persistent homology for individual data points presents challenges, especially in classification and prediction tasks where independent evaluation of each new data point is required. Due to this limitation, machine learning models using TDA may fail to generate effective topological features for new individual data points during prediction, thus restricting their practical applicability. For example, in situations where real-time predictions are required or when new data points are added one by one, traditional TDA features may not be directly applicable.

To address this issue, this study proposes a new approach called Point-Generated Persistent Homology (PGPH). PGPH generates multiple point sets from a single data point and applies persistent homology to these sets, allowing for the extraction of topological features from individual data points. With this method, topological features based on TDA can be utilized even for a single new data point, making real-time applications and individual predictions more feasible, which were challenging with traditional methods. This study validates the practicality of PGPH using specific datasets, demonstrating its potential to enhance Learning Analytics through dynamic and individualized predictions.

2 METHOD

Let a point x consists of k-features $(x_1, x_2, \ldots, x_n)$. We extend x to a set X that consists of ordered pairs of features from x. For example, given x = (x_1, x_2, x_3, x_4), then X = { (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4), (x_3, x_4)}.

In this study, we propose two approaches using PGPH. The first approach measures similarity by calculating topological features through PGPH and classifies data using a majority voting method, like the k-nearest neighbors (kNN) approach. The second approach utilizes the topological features generated from PGPH as input for supervised learning models (e.g., SVM or Random Forest).

Figure 1 shows the computational workflow for implementing PGPH. The workflow begins with preprocessing the dataset, generating point sets, calculating persistent homology, and finally integrating topological features into a classifier.

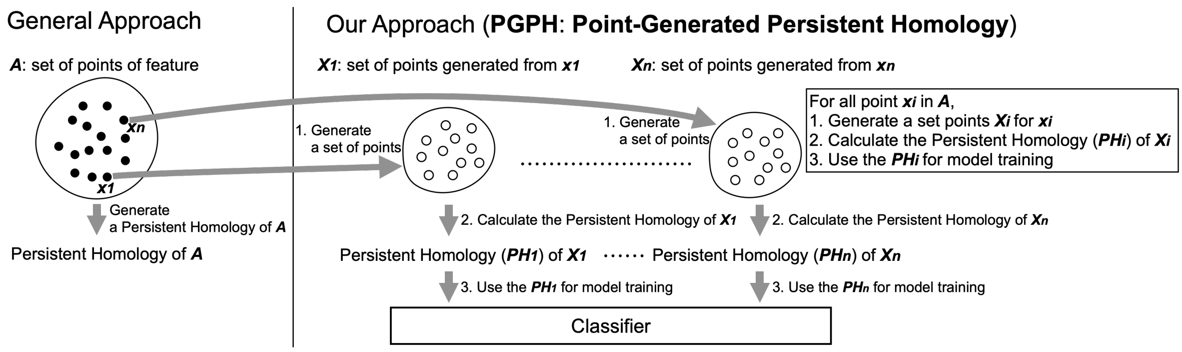


Figure 1: The computational workflow for PGPH

3 EXPERIMENTS AND RESULTS

To evaluate the proposed method, we compared it with existing methods using data from the Open University (Kuzilek et al., 2017). The evaluation data consisted of 34 features organized according to Zhang and Ahn (2023). As an indexed table, 561-point sets were generated from each individual point by selecting two features uniformly from the 34 features ($\binom{\mathbf{34}}{\mathbf{2}} = \mathbf{561}$). Persistent homology $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ were computed for each point set. For comparing two points, we use the Wasserstein distance for $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$, independently. Like k-nearest neighbors, the k points with the smallest Wasserstein distances were selected. Additionally, we evaluated several common methods such as SVM, Random Forest (RF), Gradient Boosting (GB), and Neural Network (NN) by using persistent images generated from persistence diagrams as training data instead of points. Persistent Images are also defined for both $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$.

The experimental results showed no improvement in classification accuracy when using PGPH in this case.

4 DISCUSSION AND FUTURE WORK

Our method and k-NN also share the use of information from data points close to the target point for classification. While k-NN measures “closeness” using distances like Euclidean or Manhattan, our method similarly evaluates “closeness” by calculating distances between the generated persistence diagrams. However, while k-NN evaluates proximity based on positional information between two points, this method utilizes “topological features” extracted through persistent homology, making it possible to capture geometric patterns related to data distribution and structure. For example, it is expected that this method can consider topological characteristics such as holes and cluster shapes, which k-NN cannot represent, allowing for classification based on the topological properties of data. Additionally, while k-NN’s accuracy can be improved by adjusting parameters like k, distance weighting, and axis weighting, PGPH’s classification accuracy is expected to improve through the selection of partial tuple sets. Moreover, SVM and Random Forest are effective when the original data is linearly separable, but they require adjustments for classifying data with nonlinear boundaries or complex distributions. By adding PGPH as a feature, more diverse patterns can be learned, including topological, nonlinear structures.

Future challenges include identifying the most discriminative features generated by persistent homology and excluding unnecessary features to improve model accuracy. The introduction of dimensionality reduction methods is also considered effective. Furthermore, because the calculation of persistent homology and Wasserstein distance can be time-consuming, implementing algorithms to improve computational efficiency will be necessary for practical applications.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Numbers JP22K18614, JP23K22315, and JP23K17619.

REFERENCES

Edelsbrunner, H., Letscher, D., & Zomorodian, A. (2002). Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4), 511–533. https://doi.org/10.1007/s00454-002-2885-2.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. S*ci. Data*, 4(170171) https://doi.org/10.1038/sdata.2017.171.

Munch, E. (2017). A User’s Guide to Topological Data Analysis. *Journal of Learning Analytics*, 4(2), pp.47–61. https://doi.org/10.18608/jla.2017.42.6.

Zhang, C., & Ahn, H. (2023). E-Learning at-Risk Group Prediction Considering the Semester and Realistic Factors. *Education Sciences*, 13(11), 1130. https://doi.org/10.3390/educsci13111130.

Bumpy Journey: Using Learning Analytics to Understand Undergraduate Computer Science Gateway Courses Performance and Major Switch Decisions

[1]Lujie Karen Chen, [2]Songling Ngo, [1]Safiatou Coulibaly, [1]Aditi Talwar
[1]Nishit Vyas, [1]Khalyl Konkobo
[1] University of Maryland, Baltimore County, [2] University of Washington
lujiec@umbc.edu

ABSTRACT: Computer Science (CS) gateway courses are critical milestones students must pass to proceed with the planned majors. Students may choose to leave computer science majors voluntarily or be forced to due to failing gateway courses. For some students, failing CS gateway courses and the subsequent major switch is one of the early “shocks" they must grapple with in their college experiences. These negative experiences could impact students’ self-efficacy, academic trajectory, and career outcomes. It is thus of interest to understand the factors that may predict gateway course outcomes and the relationships between gateway course performance and students' major switch decisions. In this project, we analyze the CS gateway course performance and major-switching data for students who initially plan to major in CS in a minority-serving institute. We report several learning analytics driven findings including (1) models to predict gateway course performance using information available at different points in time; (2) analysis to explore relationships between gateway course performance and major switch decisions and their association with demographic information such as gender and race, and family income level. We discuss the implications of those findings to support student advising and gateway course redesign and future work to support those inquiries.

Keywords: computer science education, gateway course performance, academic decision making

1 INTRODUCTION

Computer Science (CS) gateway courses are critical milestones students must pass to continue their computer science major. These courses traditionally serve as a filter to determine which students are prepared for the demanding studies ahead. At our institute, the gateway sequence includes CS1 (Python programming), CS2 (object-oriented programming with C++), and CS3 (discrete math). (Note: The exact course names are withheld for a blind review.) To pass these gateway courses, students must achieve at least a B in CS1 and CS2, and a C in CS3. Concurrently, they must also fulfill the math prerequisites, which include precalculus, calculus, and analytical geometry. Each of these courses can be retaken once if necessary. Historically, the passing rate for CS1 and CS2, with grades of B or higher, hovers around 60%. This has been a significant barrier for many students to proceed with the CS pathways. The project aims to understand better the interplay between gateway course performance and students' major switch decisions. This knowledge can guide us in offering more informed advice to students and, where necessary, help restructure the program to ensure a more inclusive computing pathway. In the next section, we briefly overview two strands of preliminary findings in (1) predicting the gateway course outcomes and (2) the interplay between gateway course performance and major switch decisions. We will conclude by discussing implications and ongoing research in this area.

2 Model for Predicting CS Gateway Outcomes

We explored a series of predictive models to answer the question: *To what extent and how early could we predict gateway course outcomes, i.e., whether or not a student could eventually pass all gateway courses and be eligible to proceed with CS majors?*. To answer this question, we assembled a cohort of 756 students who attempted gateway courses and enrolled as first-time freshmen in the computer science major from Fall 2015 to Spring 2018 in a Minority-Serving Institute on the East Coast of US, with academic performance data available until Fall 2022. The study was approved by the Institute Research Board at the University. We built three models with various input features available at three different points in time (Table 1). Feature set for Model 1 includes high school GPA and math placement level, which are available at the time of enrollment in the college; Model 2 feature set includes additional factors such as first-term GPA, available at the end of the first term; Model 3 features includes additional information on grades from math co-requisite courses. Using a 70% and 30% training and test split, the vanilla logistic regression yields Area Under Curve (AUC) scores of .69, .77, and .93 for Models 1, 2, and 3, respectively. We then perform model performance analysis using ROC (Receiver Operating Characteristic) Curves. We compare models' capacity to identify successful students (i.e. True Negative Rate) while maintaining a reasonable level of True Positive Rate of 80%, i.e. correctly identifying at least 80% failing students. Model 1 can only identify 36% successful students, which is suboptimal. However, we note that when adding the additional first-term GPA as in Model 2, the model can accurately identify 60% of the successful students, and when math grades were further added, 90% of the successful students can be identified. Overall, the model shows some promise in identifying students who might fail or pass gateway courses at relatively early time points.

Table 1: Model Performance for Predicting CS Gateway Outcomes

Model	Input	Time of Prediction	AUC score	TPR	TNR
Model 1	High school GPA + math-placement	at enrollment	0.69	80%	36%
Model 2	Model 1 features + First term GPA	end of first term	0.77	80%	60%
Model 3	Model 2 features + Math grades	end of first term	0.93	80%	90%

Note: TPR=True Positive Rate; TNR=True Negative Rate; Positive= Fail; Negative=Pass

3 Major Switch Decisions: "Forced" versus "Voluntary" and gateway outcomes

The CS program offers a series of three gateway courses, and students are allowed to retake each gateway course once. This structure gives students multiple checkpoints to assess whether they wish to continue in the CS major. By examining performance in these gateway courses, we can infer if students switch majors due to failure in these courses (i.e. "forced") or choose to depart the CS pathway even without encountering failures or choose not to repeat after first-attempt failure (i.e. "voluntary"). Voluntary decisions may be explained by other reasons to leave the CS major, such as a lack of belonging. In our cohort of 693 students who switched majors after taking at least one gateway course, only about 28.4% switched because they were "forced" out of CS due to gateway course failures, while the majority voluntarily chose to leave the CS program. We further analyzed the likelihood of "forced" versus "voluntary" switches based on demographics information such as race and gender. The result indicates that, on the whole, female students are more likely to voluntarily switch out of CS majors compared to male students. The likelihood of voluntary switch is 79% among 167 female students, compared with 69% among 441 male students. This gender difference is consistent across all race groups, as shown in Figure 1.

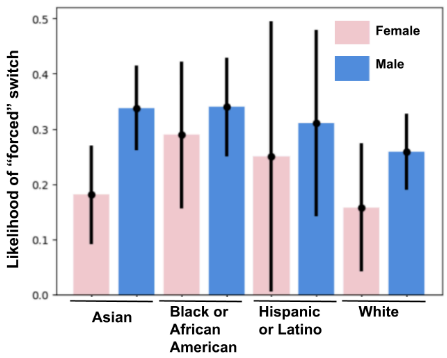


Figure 1: Likelihood of "forced switch" out of Computer Science major, grouped by gender and race, with 95% confidence interval.

We also examined the relationship between the nature of major switch decisions from the CS major and the student's family income level, which was estimated based on their high school zip codes. The result indicates that students from higher income brackets are less likely to voluntarily leave the CS major compared to those from lower income brackets. One reason could be that wealthier students have the financial resources to retake courses they fail, despite the added costs and time.

4. Conclusion and Future Work

The decision to pursue or abandon the CS pathway is multifaceted. Factors such as performance in gateway courses, personal experiences, financial considerations, or sense of belonging all play roles in students' choices of majors. As part of our ongoing research, we are gathering qualitative data from interviews and focus groups with students who have departed CS major pathways. This will offer a deeper understanding of the motivations behind their decisions. Combined with the predictive model of gateway course performance, we envision that those insights could inform data-driven decisions, such as refining student advising practices, gateway course structures, and the broader CS curriculum to improve persistence and inclusiveness (Norouzi 2023), especially for those from marginalized backgrounds.

REFERENCES

Norouzi, N., Habibi, H., Robinson, C., & Sher, A. (2023). An equity-minded multi-dimensional framework for exploring the dynamics of sense of belonging in an introductory CS course. *In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education*, 1, 131–137.

Connecting the Dots: Emerging insights into Classroom Practices from Automated Video-based Learning Analytics

Pawan Gupta, Peter Youngs
University of Virginia

ABSTRACT: Research on automatic classroom observation is nascent but growing with promises to create formative feedback opportunities for teachers at scale. To ensure this feedback is meaningful, it is important to detect the instructional setting or teaching-learning context in which the teachers make their instructional decisions. This study generates analytics on 24 fine-grained classroom activities using computer-vision techniques and uses exploratory factor analysis to test if meaningful themes of instructional settings may emerge from this seemingly discrete data. Further, it explores variation in these auto-detected themes across teachers' years of experience, grade levels (Kindergarten-to-6th grade), and subject (English language arts and mathematics) taught. Results are mixed with EFA models falling short of acceptable good-fit criteria, but cohesively translating analytics on student-teacher behavior into distinct, meaningful themes. Significant variations in some of the themes across subjects and grade-levels were also found. This work is significant because it is one of the first attempts to utilize video-based analytics and EFA in a novel manner for identifying overall teaching-learning practices in which instructions occur. Despite the limited results, it presents preliminary evidence of the effectiveness of using this technique to study variations in teaching-learning practices across different contexts like grades and subjects taught.

Keywords: Fine-Grained Learning Analytics, Computer Vision, Exploratory Factor Analysis, Neural Networks

1 INTRODUCTION

With increasing calls for providing formative feedback to teachers at scale, it is important to develop automated approaches for capturing and converting raw classroom data into useful insights on classroom instructions. Many studies have targeted and analyzed specific instructions like feedback, questioning, and uptake to develop tools for generating automated feedback using primarily audio data and machine-learning models (Wang et al. 2024). Promising results in this area have encouraged us to explore the possibility of using automated learning analytics to detect underlying instructional settings planned by teachers. Feedback on instruction is more meaningful when it is informed by the teaching-learning context in which instruction occurs. For example, knowing whether a specific questioning strategy is used by the teacher during whole class presentation or during individual support is important to evaluate the merit of instructional decision before generating meaningful feedback. This study aims to investigate: 1) How can computer-vision based learning analytics of in-person classroom recordings be used to identify meaningful and interpretable instructional settings or contexts? and 2) Can this automated analysis be used to detect variations in instructional settings employed by teachers based on their experience-level or across different grades and subjects taught?

2 METHOD

We applied enhanced background suppression neural networks (Bas-Net+) to generate video analytics (time-spent) on 24 observable fine-grained classroom activities (Foster et al., 2024) using ~41 hours of classroom videos representing 145 lesson segments (93 ELA and 52 mathematics) of 15-minutes each. The video data comes from a large corpus of classroom recordings collected in 2016-19 during a previous project, The Elementary Teacher Preparation Project (a pseudonym). We conducted correlational analysis, the Kaiser-Meyer-Olkin Test (KMO), and the Bartlett's Test of Sphericity to check the suitability of data for factor analysis. An Exploratory Factor Analysis (EFA) was conducted to investigate how the underlying structure of relationships among these activities capture meaningful classroom practices or teaching-learning context cohesively (See Figure 1 for example). We tested different extraction methods (i.e., Principal Axis Factoring & Maximum Likelihood) to determine the factor structure and different rotation methods (Varimax & Promax) for maximizing factor interpretability. Parallel Analysis, Scree plot and Velicer's MAP criteria were used to determine the optimal number of factors. Different factor models were compared using model-fit indices of the Mean Item Complexity, Tucker Lewis Index (TCI) and the Root Mean Square Error of Approximation (RMSEA). For the 2nd question, suitability of parametric or non-parametric method was checked by analyzing the distribution and variance of factor scores using the Shapiro-Wilk and Levene's test.

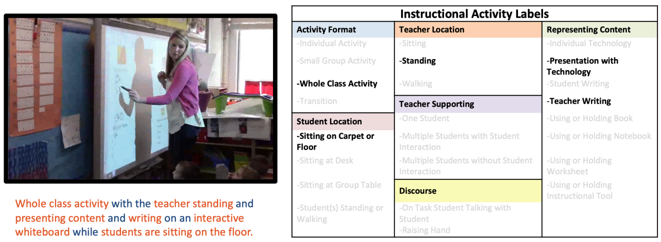


Figure 1: Activity Labels and their hypothetical representation of pedagogical strategy

3 RESULTS & DISCUSSION

RQ1: The results from Bartlett's test (χ^2=3410; *df* = 276; *p*<.01) indicated sufficiently large correlations among variables as confirmed by the correlation matrix. The overall measure of sample adequacy (0.8) with the KMO test indicated high suitability of data for factor analysis. Results from univariate Normality tests with histograms and Shapiro-Wilk test showed most variables in data deviated from normality leading to the choice of Principal Axis Factoring (PAF) for factor extraction. Expecting factors to have correlations and to minimize cross-loadings, we used an oblique rotation using the Promax technique (Thompson, 2004). The results from parallel analysis and Cattel's (1966) Scree plots revealed 5 as the optimal number of factors while Velicer's MAP criteria achieved minimum value of .05 with 4 factors. Thus, we compared the models with 4 and 5 factors. However, both models achieved TLI of 0.622 and RMSEA of 0.17 revealing an identical fit. While the models fell short of acceptable good-fit criteria, the most interesting findings were the emergence of clear, identical and meaningful themes for each factor in all models (see Table 1 for factor loadings and themes). A threshold value of absolute 0.4 for factor loading was selected for activity retention. The themes follow a clear structure representing distinct domains of classroom practices. One factor (F1) captures students' use of technology; three factors (F2, F3, F4) represent teacher-student interaction with learning material during distinct activities (individual, small-group, and whole class); and the fifth factor (F5) captures aspects of spatial pedagogy. Each factor cohesively translates student-teacher behavior into meaningful instructional settings. For example, the 2nd factor captures "teacher led whole class presentation with students participating by raising hands."

RQ2: Based on Shapiro-Wilk Test & Levene's test, we found F1 and F5 to be suitable for parametric and F2, F3 & F4 for non-parametric investigation. A one-way ANOVA produced statistically significant difference for F5 across Subjects ($F(1, 141) = [4.93]$; $p = .028$). The Kruskal-Wallis rank sum test on F4 revealed statistically significant differences across Grades ($\chi^2(6) = 18.03, p = .006$) and Subjects ($\chi^2(6) = 25.96, p = 10^{-7}$). This indicates that automated analysis may be used to detect variations in teachers' instructional strategies across subjects & grades. The variation across grades could be due to many factors such as the number of students in a class and the nature of the content taught. Figure 2 visualizes these differences in graphs.

Table 1. EFA: Factor Loadings of fine-grained classroom activities and emergent theoretical construct of each factor

Fine-Grained Classroom Activity	**F1**	**F2**	**F3**	**F4**	**F5**
Whole_Class_Activity Decimal		0.86			
Individual_Activity Decimal				0.92	
Small_Group_Activity Decimal			0.89		
Book-Using_or_Holding Decimal	-0.42	-0.57			
Instructional_Tool-Using_or_Holding Decimal			0.94		
Student_Writing Decimal	0.70				
Teacher_Writing Decimal		0.54			
Raising_Hand Decimal		0.60			
Presentation_with_Technology Decimal		0.82			
Individual_Technology Decimal	0.55			0.45	
Worksheet-Using_or_Holding Decimal	0.67				
Notebook-Using_or_Holding Decimal				0.66	
Student(s)_Carpet_or_Floor-Sitting Decimal	-0.81				
Student(s)_Desks-Sitting Decimal	0.83				
Student(s)_Group_Tables-Sitting Decimal	0.56	-0.48			
Student(s)_Standing_or_Walking Decimal					0.55
Teacher_Sitting Decimal				-0.47	
Teacher_Standing_(T) Decimal					
Teacher_Walking Decimal				0.45	0.62
Teacher_Supporting_One_Student Decimal				0.82	
Teacher_Supporting_Multiple_with_SS_Interaction Decimal			0.86		
Teacher_Supporting_Multiple_without_SS_Interaction Decimal			0.56		
On_Task_Student_Talking_with_Student Decimal				0.77	
Transition Decimal					0.59
Emerging theoretical themes of each Factor	*Students at their desks/group tables writing and using technology & worksheets*	*Teacher leading whole class activity (presentation or modeling) with active student participation*	*Teacher supporting group work with Instructional tools*	*Teacher moving in class to support students during individual activity*	*Physical movement in classroom*

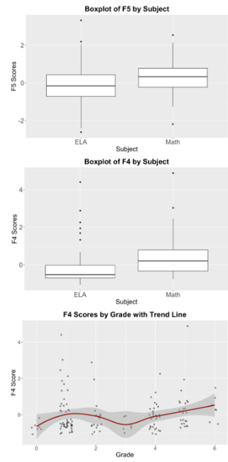


Figure 2: Visual Representation of differences in factor scores across Subjects and Grades

4 SIGNIFICANCE & LIMITATIONS

Begun as a theory to explore meaningful classroom contexts from discrete analytics, this study uses EFA in an innovative manner to statistically investigate its hypothesis. The results are mixed with the EFA model falling short of meeting an acceptable good-fit criteria, but detecting distinct, meaningful themes of instructional settings. The author believes that multiple factors might have contributed to limitations of this analysis such as small sample size, the inherent complex interactions between fine-grained activities, and the choice of activities themselves. It might also be argued that the activities may coincide to form meaningful themes by mere chance. But a synergic distribution of factor loadings and almost mutually exclusive themes found constantly across different models counter this argument. We believe further analysis and attempts at addressing the limitations should be made as detecting contexts in which instructional decisions occur can be important for investigating those decisions and advancing the work on automatic detection of classroom activities.

REFERENCES

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276.

Foster, J. K., Korban, M., Youngs, P., Watson, G. S., & Acton, S. T. (2024). Automatic classification of activities in classroom videos. *Computers and Education: Artificial Intelligence*, *6*, 100207.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association.

Wang, D., Tao, Y., & Chen, G. (2024). Artificial intelligence in classroom discourse: A systematic review of the past decade. *International Journal of Educational Research*, *123*, 102275.

Learning Analytics in AR Content Creation: Understanding Pre-service Teachers' Development as AR Designers

Author(s): Dr. Detra Price
The Ohio State University
price.151@osu.edu

Author(s): Mariah Hagadone-Bedir
The Ohio State University
hagadone-bedir.1@osu.edu

ABSTRACT: In this poster session, we present a case study examining how preservice teachers develop AR content creation skills through learning analytics. Using a specially designed AR Development Session Tracking Sheet, we collected data from six preservice teachers across five weekly one hour AR development sessions and documented their creation patterns, time allocation, challenges, and confidence progression. Complemented by semi-structured interviews, our analysis revealed distinct patterns in participant's learning trajectories, showing an evolution from basic AR elements to complex interactions, and a shift from instructor-dependent to peer and independent problem-solving. Results from this study can be used to inform a wider AR implementation study across teacher education programs in supporting preservice teachers as AR education content creators.

Keywords: augmented reality, teacher education, learning analytics, case study

1 BACKGROUND

This study aimed to explore the potential of learning analytics in understanding how pre-service teachers developed competencies in creating AR educational experiences. The research focused on developing a framework for analyzing teachers' learning progression as they acquired AR design and development skills. With AR, there is digital content that overlays a live view of the environment (Scavarelli et al., 2021). As AR technology continues to evolve, its applications in education have grown (Carreon et al., 2020), creating a need for teachers to effectively create and implement AR experiences. However, understanding how teachers, especially pre-service teachers, develop these skills remains a challenge. This study sought to address this gap by investigating how learning analytics can provide actionable insights into teacher's AR development processes and learning patterns. This study followed up on Scavarelli et al. (2021) underlining that the content creation of AR and VR is often portrayed as too complicated for instructors. In their review of AR/VR literature, Scavarelli et al. advocated for platforms to be available for instructors and learners without having to go through a developer.

Thus, by analyzing design behaviors, tool usage patterns, and learning outcomes, we aimed to develop a structured understanding of how teachers learn to create AR content. This research also provides practical guidelines for teacher education programs on supporting AR content creation skill development through analytics data. This approach aligns with the growing emphasis on teachers as technology creators and data-informed professional development. By bridging the gap between AR content creation and learning analytics, this study contributes to the ongoing efforts to prepare teachers who can effectively design and implement AR-enhanced learning experiences.

2 OBJECTIVES OF THE STUDY

This research aimed to establish a robust framework for implementing and analyzing learning analytics in AR content creation training within teacher education programs. The primary objectives were to identify patterns in preservice teachers' AR development behaviors, develop evidence-based principles for AR design learning, and create guidelines for integrating learning analytics tools within AR creation environments.

The study aimed to analyze how preservice teachers engaged with AR development tools and identified common patterns in their learning process through learning analytics data. Through these objectives, we summarized best practices and assessment frameworks that measured both technical proficiency and pedagogical understanding in preservice teachers who learned how to create AR experiences. Therefore, our research questions for this study were:

1. [RQ1] How do learning analytics from AR development sessions reveal preservice teachers' iterative design behaviors, and how do these behaviors relate to their AR creation competencies?

2. [RQ2] What relationships, if any, emerge between AR tool usage patterns and preservice teachers' pedagogical understanding of AR implementation in education?

3 METHODS

This qualitative case study investigated preservice teachers' learning progression using the Tinkercad app to deliver an AR learning activity. In this study, we analyzed the learning analytics data (i.e., the AR Development Session Tracking Sheet) to identify patterns and trends and then conducted semi-structured interviews to help explain why these patterns existed. The study used data from five AR development sessions over five weeks that each lasted one hour.

3.1 Data Collection

After obtaining IRB approval and receiving consent from participants, participants received a brief orientation on how to use AR. Learning analytics data was collected from an AR Development Session Tracking Sheet which included: AR elements created in each session, time spent on tasks, any challenges or assistance encountered and needed, any resources used for help, and a progress self-assessment of the participant's confidence proceeding through the development of the AR learning experience. Visual data representation was also done, and this was taking a picture of the participant's progress at the conclusion of each session. At the completion of the fifth session and tracking sheet, semi-structured interviews were conducted to provide valuable context for interpreting the analytics data and understanding the relationship between participants' AR development process and their pedagogical understanding of using AR as part of a learning experience. The interview protocol explored participants' reflections on their learning journey, specific challenges they overcame, and how their understanding of AR's educational applications evolved throughout the sessions.

3.2 Data Analysis

Data analysis combined individual learning trajectory analysis with visual data representation to understand the progression of preservice teachers' AR development skills. Individual learning trajectory analysis involved tracking each preservice teachers' development path in learning to create AR experiences across the five sessions. This analysis method, like a journey map, maps out each participant's unique learning journey by examining several key progression indicators from their tracking sheets which included technical skill development, time management, confidence progression, and work pattern changes. Visual representation of AR element creation showed the intervals of content creation and level of sophistication.

4 RESULTS

Analysis of learning analytics data from six preservice teachers using Tinkercad revealed distinct patterns in how they approached 3D design for AR implementation. Data showed preservice teachers initially struggled with spatial manipulation tools, spending an average of 40 minutes in their first session learning basic object placement and rotation. However, tool usage logs revealed a specific progression: preservice teachers who started with modifying existing shapes (rather than creating from scratch) developed faster proficiency with Tinkercad's core features. The analytics highlighted specific pivot points in preservice teacher learning - particularly when moving from basic shape manipulation to combining objects for more complex designs. Interface interaction data showed that successful preservice teachers frequently toggled between different viewports (top, front, side), suggesting that understanding 3D space visualization was a crucial competency. These findings indicate that preservice teacher training for 3D design tools might be more effective if structured around spatial reasoning skills rather than feature-by-feature instruction. The data suggests preservice teachers struggled not because they couldn't find or use specific tools, but because they had trouble visualizing how objects would interact in 3D space. They might know how to use the rotate tool, for instance, but still struggle to understand how rotation would affect their entire model from all angles.

So rather than teaching Tinkercad as a series of isolated features, the findings suggest we should structure learning around spatial thinking tasks that naturally incorporate multiple tools as needed. This mirrors how architects and engineers think about 3D design - they start with spatial concepts and then use whatever tools they need to realize their vision.

5 REFERENCES

Carreon, A., Smith, S. J., & Rowland, A. (2020). Augmented reality: Creating and implementing digital classroom supports. *Journal of Special Education Technology*, *35*(2), 109-115.

Scavarelli, A., Arya, A., & Teather, R. J. (2021). Virtual reality and augmented reality in social learning spaces: a literature review. *Virtual Reality*, *25*(1), 257-277. https://doi.org/10.1007/s10055-020-00444-8

A Visual Programming Approach to Enhance Spatial Computational Thinking Skills in Upper-Elementary Students

Zifeng Liu[1], Wanli Xing[1], Fan Zhang[1], Chenglu Li[2]
[1]University of Florida, [2]University of Utah
liuzifeng@ufl.edu, wanli.xing@coe.ufl.edu, f.zhang1@ufl.edu, chenglu.li@utah.edu

ABSTRACT: This poster presents a visual programming approach to enhancing spatial computational thinking (SCT) in elementary students by leveraging the capabilities of Minecraft: Education Edition (MEE). While MEE's built-in MakeCode platform offers basic functions such as cloning and object manipulation, it lacks the flexibility for more complex spatial transformations like rotation, reflection, and shape grouping—skills essential for SCT development. We introduced a specialized visual programming module with accessible, age-appropriate functions that enable students to engage in procedural and generative SCT paradigms within an immersive environment. The module includes seven core functions (Group, Move, Rotate, Reflect, CheckShape etc.) and a corresponding MEE map designed to support SCT tasks. A pilot study involving seven public elementary school students suggested potential benefits in enhancing understanding of spatial concepts following use of the module. By using a visual programming approach in an interactive environment, the study allows learning analytics to capture detailed data on how students engage with spatial problem-solving, enabling analysis of skill acquisition, progression, and learning patterns.

Keywords: Spatial computational thinking, elementary students, visual programming, immersive learning environment

1 INTRODUCTION

As computing technology has advanced, there is a growing need to develop students' spatial thinking skills, an area often underemphasized in STEM education (e.g., Newcombe, 2017; Khine, 2017). Spatial computational thinking (SCT) skills—such as abstracting spatial features, decomposing shapes, recognizing patterns, and applying computational tools to solve spatial problems—are crucial for students as they prepare for future careers in increasingly digital and data-driven fields.

2 METHODS

An enactive learning approach, which emphasizes autonomy, embodiment, and situated learning, offers a promising way to teach SCT skills (Addan et al., 2024; Hutto et al., 2015). This approach promotes active, hands-on engagement, allowing learners to refine actions and internalize knowledge through environmental interactions. Using concreteness fading theory (Skulmowski, 2023), we designed activities that transition from physical models (e.g., paper-based) to digital manipulation (e.g., tinkering with shapes), and finally to abstract spatial programming. We used Minecraft: Education Edition (MEE), which integrates self-directed exploration and encouraging student autonomy and curiosity. While traditional visual programming tools like Scratch and MakeCode have been effective in early computing education, they often lack the spatial programming features needed to fully support SCT. Additionally, essential spatial transformations like rotation and reflection are absent, despite being crucial for spatial reasoning and 2D/3D design. Moreover, MEE lacks a direct function for grouping and ungrouping shapes—a vital capability for decomposing and composing shapes programmatically and for automating complex designs. Advanced languages like Python and Java can handle complex spatial tasks but are not age-appropriate for young learners. To address this gap, we are developing a new spatial programming module within MakeCode, designed to provide upper elementary students (ages 9–12) with a high-quality, integrated SCT learning experience within MEE. This spatial programming module will benefit students by broadening access to SCT skill development in an engaging, age-appropriate way.

Table 1: Created functions for students' learning activity.

Function	Description	Coding block example
Group	Combine two shapes together to form a new shape.	set pyramidBase to Make platforms as a group
Ungroup	Detach shapes in a combined shape to be individual shapes.	set mayaPyramidComponents to Ungroup mayaPyramid
Move (1)	Move a shape along the X, Y, and Z axes by a given number of units.	Move mesh mayaPyramid towards axis x (East/West) by 1 units
Move (2)	Move a shape to a destination.	Move mesh mayaPyramid to destination world 200 65 17
Rotate	Rotate a shape along the X, Y, and Z axes by any given number of degrees.	Rotate mayaPyramid over axis x (East/West) by 90 degrees
Reflect	Reflect a shape over the X, Y, and Z axes.	Reflect pyramidStair1 over axis x (East/West)
Check Shape	Automatically validate a shape with predefined rules.	Validate shape of mayaPyramid

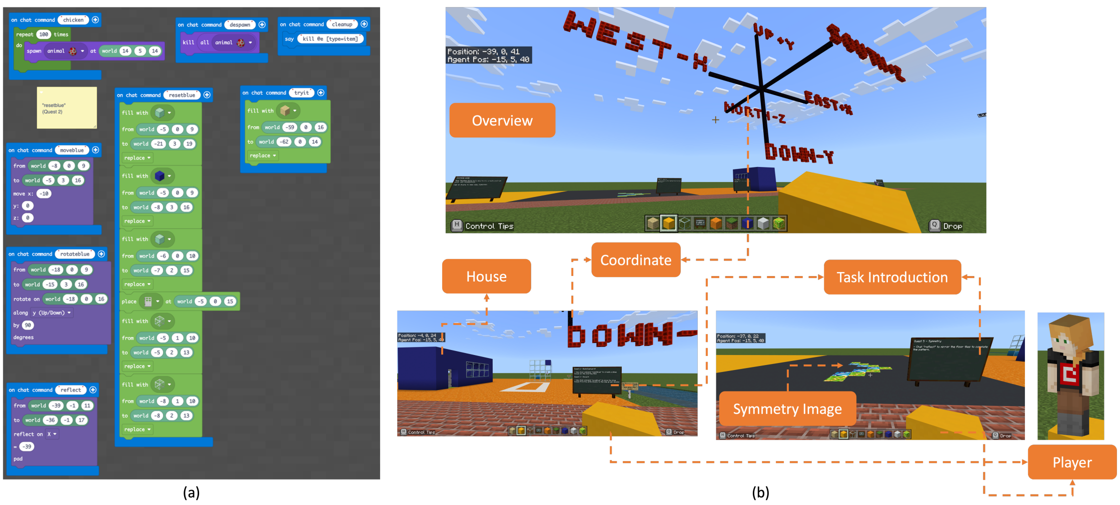


Figure 1: The developed module (a) and the spatial computational thinking learning map (b).

To help students practice procedural and generative paradigms of spatial computational thinking in MEE, we designed a module (see Table 1) to extend the programming functions of MakeCode. This design offered several unique affordances for students: (1) spatial manipulation of multiple objects simultaneously through programming; (2) construction of 3D objects via algorithmic design; and (3) optimization and strategic development of algorithms to realize 3D designs through effective decomposition and automation. The module included two functions for decomposition and automation (Group and Ungroup), three spatial transformation functions (Move, Rotate, and Reflect), and a validation function (CheckShape). The selection of spatial transformation functions was based on research into children's spatial development (e.g., Newcombe, 2017). Figure 1 shows the overview of the module (a) and the immersive learning map (b).

3 DATA COLLECTION AND PRELIMINARY RESULTS

A male teacher from a public elementary school implemented the module and map over a one-week classroom session. Data were collected from seven fifth-grade students (4 boys and 3 girls) to assess the module's effectiveness in enhancing SCT skills. The evaluation included three spatial concepts questions (ask about symmetry, direction and rotation) and three computational thinking questions, each rated on a five-point Likert scale: (1) "I am used to figuring out procedures step by step for a solution"; (2) "I usually try to find effective solutions for a problem"; and (3) "I usually try to lay out the steps of a solution." Results indicated that one student showed improvement in spatial concept understanding, while others remained at the same level. For SCT Question (1) and (2), minimal changes in responses suggested stable or slightly positive attitudes towards procedural and effective solution strategies. However, responses to Question 3 shifted towards disagreement in the post-survey, suggesting that students may have less need for structuring solution steps after the learning.

4 CONCLUSION AND FUTURE WORK

This poster presented a module designed and developed for students learning spatial computational thinking and provide foundation for learning analytics to capture detailed data on how students engage with spatial problem-solving. The module presented would enable students to handle intricate, composed shapes more easily, enhancing their understanding of decomposition and other core computational thinking concepts. The project's resources, including demo videos[1], detailed documentation, and the MakeCode Playground, equip educators with the knowledge and tools to seamlessly incorporate SCT into their curriculum. Future work will gather feedback from educators and students and expand analysis to support learning analytics.

REFERENCES

Addan Gonçalves, D., Caceffo, R., Armando Valente, J., Bonacin, R., Cesar dos Reis, J., & Calani Baranauskas, M. C. (2024). Enactive interaction in technology-based learning environments: A systematic literature review. *Educational Technology & Society*, *27*(2).

Hutto, D. D., Kirchhoff, M. D., & Abrahamson, D. (2015). The enactive roots of STEM: Rethinking educational design in mathematics. *Educational Psychology Review*, *27*, 371-389.

Khine, Myint. (2017). Spatial cognition: Key to STEM success. 10.1007/978-3-319-44385-0_1.

Newcombe, N. (2017), "Harnessing Spatial Thinking to Support Stem Learning", OECD Education Working Papers, No. 161, OECD Publishing, Paris. http://dx.doi.org/10.1787/7d5dcae6-en

Skulmowski, A. (2023). Do concreteness fading and guidance fading aid learning from perceptually rich visualizations? Changes in style lead to more cognitive load and interfere with learning. *Current Research in Behavioral Sciences*, *4*, 100112.

[1] https://youtu.be/Shqd9EMg3PY?si=DtTqvFFFwyNBISOC

Balancing the Scales: Using GPT-4 for Robust Data Augmentation

Mohammad Amin Samadi, Danielle Shariff, Raymond Ibarra, and Nia Nixon
University of California, Irvine, School of Education, Irvine, CA, USA
{masamadi, dsharif1, raymonai, dowelln}@uci.edu

ABSTRACT: This study examines the use of GPT-4 for data augmentation in a collaborative problem-solving (CPS) context to address class imbalance, specifically targeting the underrepresented “Cognitive Planning” class. Using three distinct prompting strategies, GPT-4 generated samples were compared against a back-translation baseline. Evaluations focused on Content Consistency (alignment with the original instance’s meaning and structure), Class Alignment (fit to class-specific patterns), and Semantic Similarity. Results showed that while back-translation achieved the highest Semantic Similarity, Prompt 2 balanced alignment with class-specific language and fidelity to the original context. Prompt 3, designed to generate original scenario-based examples, achieved near-perfect Class Alignment but faced challenges with Semantic Similarity due to conceptual departures from the original utterances. Key challenges included shifts in tone and added information, which occasionally reduced alignment with the Cognitive Planning criteria. Findings highlight GPT-4’s potential to generate diverse, contextually accurate data for improving model performance in minority classes.

Keywords: Data Augmentation, LLMs, GPT-4, Data Generation, Collaboration, NLP

1 INTRODUCTION & BACKGROUND

Despite advancements in Natural Language Processing (NLP), maintaining balanced representation within data samples remains a challenge in Machine Learning (ML). In educational research, data distribution is often skewed due to the diverse nature of student data (Fang et al., 2023). Machine learning algorithms are trained on overrepresented classes more frequently by nature. Consequently, they may struggle with classifying the underrepresented classes accurately, making robust data augmentation methods essential to improve classification outcomes with imbalanced datasets (Dai et al., 2023). Data augmentation enhances dataset balance by generating additional samples to address class imbalance without new data collection. Building on prior research that modeled collaborative problem-solving (CPS) skills using machine learning and NLP techniques (Samadi et al., 2024), recent advancements in large language models (LLMs), such as GPT-4, enable the creation of diverse, contextually accurate samples that improve classification in complex NLP tasks (Dai et al., 2023). In Learning Analytics, GPT-4 has been particularly effective in generating representative samples for underrepresented classes, improving automated scoring and other NLP applications (Liu et al., 2023).

This study explores using GPT-4 to augment underrepresented ‘Cognitive Planning’ instances within a collaborative problem-solving (CPS) context. Cognitive Planning refers to strategizing and organizing actions to achieve goals, a crucial component of effective collaboration. We aim to determine if GPT-4 can improve dataset balance by generating diverse, contextually accurate examples that enhance automated coding. The dataset comprises chat data from 516 undergraduate students working in teams of four on problem-solving tasks. Discussions were coded using a modified CPS Ontology (Andrews-Todd & Forsyth, 2020), covering social and cognitive skills. Among the codes, ‘Cognitive Planning (CP)’ was significantly underrepresented, appearing in only 38 utterances, compared to more frequent codes like ‘Sharing Information’ (2,638) and ‘Maintaining Communication’ (1,166),

Prompt 1	*"Paraphrase the following text in 10 different ways, separated by new lines. Only include the sentences, no numbers needed: {original_sample}"*
Prompt 2	*"You have been assigned the task of data augmentation for an underrepresented class. This involves generating new data belonging to that class by paraphrasing existing instances. Here are three random examples that belong to the class: {random_examples}. Please paraphrase the following text in 10 different ways, separating each paraphrase with a new line: {original_sample}"*
Prompt 3	*"In collaborative problem solving, planning is a cognitive process that involves setting shared goals, determining the steps needed to achieve those goals, and coordinating efforts among team members. This process requires organizing actions and ensuring that everyone's contributions are aligned with the overall objectives. For example, in a team project, planning might involve assigning specific tasks to each member, setting milestones for group check-ins, and revising the strategy based on collective feedback.* *Based on this definition, generate 10 original utterance examples of cognitive planning in collaborative problem-solving, each illustrating different ways a team uses planning to achieve shared goals. These examples should be very short and conversational style as if they happened during a short-term (plans should not be long-term) team discussion during problem-solving. Separate each example with a new line.* *Here are a few examples: {random_examples}"*

Figure 1: Prompts used for data generation with GPT4

highlighting a substantial class imbalance that motivated this study. Three distinct GPT-4 prompting strategies were tested and compared against a baseline back-translation method (M2M100 model). Evaluation focused on three criteria: semantic similarity, human-rated similarity to the original, and alignment with the Cognitive Planning code. Together, these three criteria provide both qualitative and quantitative evaluations of the generated data, allowing a comprehensive assessment of GPT-4's effectiveness in data augmentation for minority class enhancement. The research question is as follows: How well do different prompting techniques maintain both semantic similarity to the original instances and alignment with the Cognitive Planning coding criteria?

2 DATA GENERATION & EVALUATION

GPT-4 was used to generate data for an underrepresented CP, employing three distinct prompt strategies. The first prompt provided basic paraphrasing instructions to generate variations of the original text while maintaining semantic integrity, introducing linguistic diversity. The second prompt included examples from the Cognitive Planning class to guide GPT-4's paraphrasing towards class-specific language patterns. The third prompt departed from paraphrasing by generating original, scenario-based utterances inspired by a definition of cognitive planning, adding conceptual depth to the dataset (see Figure 1 for more detail). For baseline comparison, back-translation was conducted using the M2M100 model across a selection of languages to further diversify resamples. The code and implementation details are available at https://github.com/aminsmd/data-augmentation/.

To evaluate the quality of the generated data, resamples were scored on three main criteria: Content Consistency, Class Alignment, and Semantic Similarity. **Content Consistency (CC)** assessed how closely the resample matched the original instance, focusing on retaining core meaning and structure. This helped prevent significant semantic deviations that might distort the original message. **Class Alignment (CA)** measured the degree to which each resample fit the coding class of the original, ensuring it was consistent with linguistic patterns and features tied to that class. **Semantic Similarity (SS)** evaluated the depth of meaning overlap between the resample and the original, focusing on whether the resample maintained contextual relevance and thematic consistency with the source. Two researchers independently scored generated data to ensure objectivity, assigning scores based on these criteria and achieved strong agreement on CA (0.9) and adequate alignment on CC (0.73) indicating reliable quality assessments across the dataset using Cohen's Kappa.

3 RESULTS & DISCUSSION

Qualitative scores were independently assigned by two researchers based on the rubric for CC, CA. Common issues surfaced in the data augmentation process, revealing both the model's potential and limitations for enhancing class representation. Key challenges included inaccurate verb replacements, shifts away from action-oriented language, and added information that altered the original intent or context of CP utterances. As shown in Table 1, these issues sometimes led to semantic changes, which reduced alignment with the action-focused tone and context of CP.

Table 1. Common themes found during the qualitative assessment of the generated data

Issue	Explanation	Original sample	Resample
Verb replacement	Resampled text introduces a verb that alters the original meaning.	rank birthday venues	Prioritize the elite venues for celebrating birthdays.
Adherence to CP	Shifts from action-oriented to a declarative tone, reducing alignment with CP's language.	we're gonna have to compile the data	It's our job to put the data together.
Context Misalignment	Contains info misaligned with the original study's task and context.	-	Can you lead the research today? We'll review on Friday.

Table 2. Performance of the four data augmentation methods on Content Consistency, Class Alignment, and Semantic Similarity

Methods	Count	CC (M, SD)	CA (M, SD)	SS (M, SD)
Back-translation	330	0.427, 0.495	0.667, 0.472	0.761, 0.164
Prompt 1	330	0.621, 0.486	0.721, 0.449	0.638, 0.172
Prompt 2	330	0.688, 0.464	0.876, 0.330	0.671, 0.174
Prompt 3	300	-	0.997, 0.058	0.339, 0.080

To evaluate semantic similarity and other performance metrics across data augmentation methods, we compared Back-translation, Prompt 1, and Prompt 2 with each original instance, while Prompt 3 calculated similarity using the average embedding, making its SS score less directly comparable. Since Prompt 3 did not paraphrase specific instances, CC was not applicable. As shown in Table 2, Back-translation achieved the highest SS but lower CC, while Prompt 2 balanced high CC and strong CA. Prompt 3 excelled in CA but had lower SS.

REFERENCES

Andrews-Todd, J. and Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. Computers in human behavior, 104:105759.

Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., Zhao, L., Xu, S., Liu, W., Liu, N., et al. (2023). Auggpt: Leveraging chatgpt for text data augmentation. arXiv preprint arXiv:2302.13007.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., et al. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. Meta-Radiology, page 100017.

Samadi, M. A., Jaquay, S., Lin, Y., Tajik, E., Park, S., and Nixon, N. (2024). Minds and machines unite: Deciphering social and cognitive dynamics in collaborative problem solving with ai. In Proceedings of the 14th Learning Analytics and Knowledge Conference, pages 885–891.

Degrees of belonging: Gaining insights into university students' belonging through theory-informed learning analytics

Lisa-Angelique Lim, Simon Buckingham Shum
University of Technology Sydney
[Lisa-Angelique.Lim, Simon.BuckinghamShum]@uts.edu.au

ABSTRACT: This poster reports ongoing work to leverage learning analytics for enhancing students' sense of belonging in higher education. Despite the importance of belonging for student engagement, there is a significant research gap in how to monitor and support students' belonging throughout their degree programs. Using an innovative, theory-informed learning analytics approach, we conduct a study to gather and analyze both quantitative and qualitative data on students' belonging at scale via the SenseMaker® tool. This platform allows respondents to share narratives (referred to as 'stories') and then code these using signifiers grounded in theories of belonging. Currently conducted at an institution in [country blinded for review], we invited in-degree students across the university to share their stories of belonging (or alienation). The poster presents preliminary findings from the collected data and discusses possible interpretations and future directions, contributing to the emerging subfield of Belonging Analytics.

Keywords: belonging analytics, sensemaking, theory-informed learning analytics, participatory narrative coding

1 INTRODUCTION

Student belonging is crucial for academic success, retention, and well-being in higher education (Meehan & Howells, 2019). It is complex (Allen et al., 2024), dynamic (Kahu et al., 2022), and influenced by context, culture, and personal demographics (Gravett et al., 2023). Negative feelings of belonging can affect students' motivation to learn, making it essential to monitor and support student belonging throughout their studies. However, measuring belonging is challenging. Quantitative methods like self-report instruments and national surveys provide large-scale snapshots but lack deep insights and are often conducted too late for timely support. Qualitative methods, such as interviews, focus groups, and more recently vlogs (Gravett et al., 2023), offer deeper insights but are difficult to scale. The rise of online learning has further limited face-to-face interactions, reducing opportunities to observe behavioral cues related to belonging. There is a growing need for learning analytics to capture the complexity of student belonging and transform it into actionable insights for personalized support. Currently, data-informed approaches to support belonging are limited in learning analytics research (e.g., Benedict et al., 2022; Ramanathan et al., 2024).

To address the challenges, this study aimed to capture the complexities of student belonging in a higher education setting, via the use of an innovative research platform called SenseMaker® (Van der Merwe et al., 2019). SenseMaker® is an online mixed methods research tool that gives students a voice to tell their stories of belonging, and then index their own stories against the theoretical dimensions of this complex concept through 'self-signifiers'. Moreover, the platform's analytical function visualizes data through dashboards in a dynamic way. This pilot exploration has two goals: 1) To design a SenseMaker® framework grounded in theories of belonging; and 2) to examine patterns of student belonging derived from participatory narrative analytics in a higher education setting. This exploratory work contributes to the topic of 'Belonging Analytics' (Lim et al., 2023), an emerging field that leverages learning analytics to gain insights into belonging.

2 METHODOLOGY

This study is being conducted at the University of Technology Sydney, Australia, a research-intensive, public university, under an ethics-approved research project. All students were invited to participate through a general student newsletter. Students were presented with a flyer explaining the study and directed to a link to the SenseMaker® platform to share their stories anonymously. They could also share more than one story; individual stories were not visible to participants. Participants had the option to enter their institutional email addresses to enter a draw to win one of ten $50 gift vouchers.

We designed a SenseMaker® framework grounded in theories of belonging. In so doing, we reviewed the relevant literature on student belonging in higher education to identify key dimensions, indicators, and factors of student belonging. We also consulted with key stakeholders—faculty, staff, and students—to obtain their feedback on the design of the framework. In the platform, students were prompted to input their stories of belonging in a text box provided, describing the experience as fully as possible, in 1 to 2 paragraphs. An example of a signifier and how it is grounded in belonging theory is shown in Table 1. The signifiers were presented in the form of a triangle ('triad'), anchored by each of the conceptual dimensions. Students placed a marker within the space of the triad, to indicate how they felt about their experience in relation to the signifier; for instance, if their story was more about their sense of self-identity, they would place the marker closer to that point of the triad. Students could also indicate 'N/A' for any signifier(s) if these were not relevant to their story.

Table 1. Example of theory-grounded self-signifiers used in the Sensemaker framework.

Theoretical grounding	Question	Signifier
Academic belonging (Kahu et al., 2022)	The experience in my story had an impact on my…	**Triad:** • Sense of self-identity as a person • Sense of identity as someone working in my discipline • Willingness to persevere in my studies

3 RESULTS AND DISCUSSION

To date, 55 students across a range of disciplines have participated in this study, with 49% being in the first year of their degree. Figure 1 shows preliminary data from three of the signifiers used in this study, providing some insight into how students experience belonging. For example, students' stories of belonging expressed a desire for identity (Figure 1(L) and (Mid)). More students also wished they had more support for their mental health based on their stories (Figure 1(R)). On the whole, 72% of the respondents indexed their stories as being 'positive' or 'strongly positive'.

These preliminary results resonate with some of the existing empirical work relating to first-year students' belonging experiences, especially the initial focus on interpersonal belonging (e.g., Kahu et al., 2022). The finding that more students were hoping for more support with their mental health is especially notable; as an educational implication, support for psychological wellbeing is important for staving off the anxiety which comes with transitioning to a new academic and social culture. Possibly, inviting students to share their stories at regular intervals as a regular reflective activity can provide important information about their ongoing sense of belonging, to inform timely and personalized interventions. Furthermore, these visualizations could serve as dashboards to program leads, providing information about students' belonging as they progress through their degree programs.

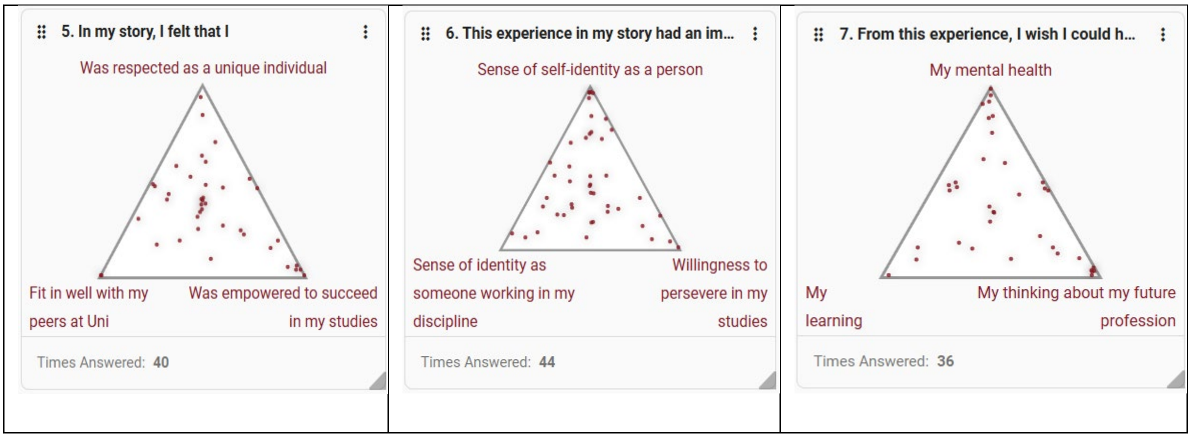


Figure 1: Example data visualisations from three self-signifiers used in the study. (L) Interpersonal belonging; (Mid) Academic belonging; (R) Impact of belonging experience on wellbeing

When the data collection for this study is completed, the text data gathered from respondents' stories will be qualitatively analyzed further for themes, and quantitative data from the signifiers will be explored to gain further insights into subgroups of students.

REFERENCES

Allen, K.-A., Slaten, C., Hong, S., Ma, L., Craig, H., May, F., & Counted, V. (2024). Belonging in Higher Education: A Twenty Year Systematic Review. *Journal of University Teaching and Learning Practice*, *21*(5), 1-55. https://doi.org/10.53761/s2he6n66

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). Association for Computing Machinery. https://doi.org/10.1145/3506860.3506903

Gravett, K., Ajjawi, R., & O Shea, S. (2023). Topologies of belonging in the digital university. *Pedagogy, Culture & Society*, 1-15. https://doi.org/10.1080/14681366.2023.2256342

Kahu, E. R., Ashley, N., & Picton, C. (2022). Exploring the complexity of first-year student belonging in higher education: Familiarity, interpersonal, and academic belonging [Other Journal Article]. *Student Success*, *13*(2), 10-20. https://doi.org/10.3316/informit.544244789917082

Lim, L.-A., Buckingham Shum, S., Felten, P., & Uno, J. (2023). "Belonging analytics": A proposal. *Learning Letters*, *1*, 4. https://doi.org/10.59453/EAXA8005

Meehan, C., & Howells, K. (2019). In search of the feeling of 'belonging' in higher education: undergraduate students transition into higher education. *Journal of Further and Higher Education*, *43*(10), 1376-1390. https://doi.org/10.1080/0309877X.2018.1490702

Ramanathan, S., Buckingham Shum, S., & Lim, L.-A. (2024). *To what extent do responses to a single survey question provide insights into students' sense of belonging?* Proceedings of the 14th Learning Analytics and Knowledge Conference, Kyoto, Japan. https://doi.org/10.1145/3636555.3636916

Van der Merwe, S. E., Biggs, R., Preiser, R., Cunningham, C., Snowden, D. J., O'Brien, K., Jenal, M., Vosloo, M., Blignaut, S., & Goh, Z. (2019). Making Sense of Complexity: Using SenseMaker as a Research Tool. *Systems*, *7*(2), 25. https://www.mdpi.com/2079-8954/7/2/25

Personal Learning Analytics: Integrating Data Science Education to Improve Data Literacy and Self-Regulated Learning

Lujie Karen Chen
University of Maryland, Baltimore County
lujiec@umbc.edu

ABSTRACT: This study explores the integration of Personal Learning Analytics (PLA) in an undergraduate introductory data science course to foster self-regulated learning (SRL) by engaging students to collect, analyze, and reflect on their own data. In this PLA framework, students tracked personal data, such as study hours, exercise, and wellness metrics, over 40 days. Subsequently, students posed questions about their data and conducted analyses using data visualization and inferential techniques learned in class. Grounded in constructivist theory, PLA encourages students to actively construct knowledge from familiar, personal contexts to enhance their sense-making and conceptual understanding. Qualitative analysis of students' work products and reflections reveals that students not only demonstrate a robust understanding of challenging statistical inference topics but also display an emerging pattern of self-awareness and self-regulation even though they are not explicitly discussed in the course. This study provides initial evidence of PLA as a promising tool to enhance SRL, feedback literacy, and data literacy. It concludes with future research to scale PLA for larger, more diverse cohorts.

Keywords: self-regulated learning, feedback literacy, data literacy, data science education

1 INTRODUCTION

Learning analytics (LA) use learner-generated data, data science, and analytical methods to inform educational practices, enhancing students' learning outcomes and, ultimately, flourishing. The ultimate goal of education is to empower students to become self-directed lifelong learners who can be in charge of solving complex learning problems of their own. This research explored Personal Learning Analytics (PLA), a framework that integrates learning analytics research with data science/analytics education. PLA is a kind of student-centered learning analytics where students are empowered to become their own "learning scientist" and "data scientist" by collecting, analyzing, and reflecting on data of their own, to improve their capacity to self-regulate in learning and beyond. This idea bears similarity with the framework of "learning analytics for learners" proposed by Knight and Anderseon (2016). In those settings, we use LA as a means to support students to discover their agentic power and voice to be in charge of their learning. With the PLA framework, we integrate learning analytics and data science/analytics education through teaching transferable data science/analytics knowledge and skills and transferable self-regulating learning (SRL) skills. In this paper, we report a subset of qualitative results from an analysis of students' work products and reflections in an introductory undergraduate data science course in which PLA has been an integral part of the curriculum for the last several years.

2 Course Background

The study was conducted in an introductory undergraduate data science course in a four-year minority-serving institute on the East Coast of the US and approved by the Institute Research Board of the university. This course is open to all majors with zero prerequisites and has attracted students from a diverse range of majors and backgrounds, with a moderate enrollment of 25-30 students in each cohort. This course teaches basic data science and analysis skills using a Python-based computing framework integrated with units in data wrangling, data visualization, and statistical inference using modern simulation techniques such as randomization and bootstrapping methods.

3 Personal Learning Analytics Activity and Theoretical Grounding

The PLA activity entails two components: (1) **Data collection**, in which students are invited to track some aspects of their life for about 40 days and instructed to log in to a spreadsheet; **(2) Data analysis:** we designed two open labs that allow students to pose their own interesting questions and answer them using data visualization and inference techniques learned in class. PLA is grounded in the theoretical framework of constructivism, which suggests that learners actively construct their own knowledge and understanding. "Knowledge-in-Piece," rooted in Constructivism, is an educational psychology framework that suggests individuals' understanding and knowledge is composed of numerous small, disconnected pieces or elements, which requires careful learning design to connect those pieces. Those frameworks underpin the growing literature studying students' agency in relation to the data while they are actively engaging in collecting or analyzing their own data, posing their own questions, or designing their own experiments. Studies have highlighted its benefit to foster deeper conceptual understanding and that personal data offers students a natural, familiar context, thus facilitating their sense-making.

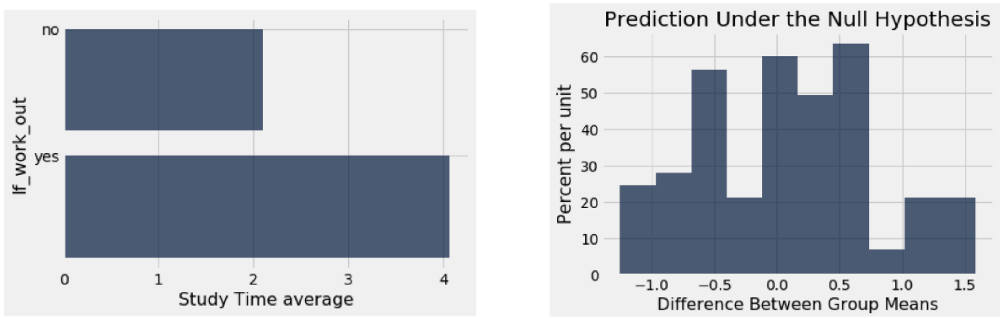
4 Results

4.1 Data Collected by students

As part of the PLA activity in their data science course, students used various data collection techniques, from wearable sensors to manual recording, and tracked various activities. This list included time management (e.g., studying, viewing YouTube, social media use, screen time, reading, listening to audiobooks, video gaming, and artistic pursuits), wellness metrics (e.g., sleep patterns, exercise, caffeine intake), social interactions (e.g., conversations with friends or family), and emotional states.

4.2 Analysis of students' work product and reflection

Due to the space limitation, we only report the analysis of one of the students' work products. This student collected six weeks' worth of data on daily hours spent on exercise and study. In one of the subsequent open-lab sessions, she used Python-based data manipulation and visualization techniques to investigate questions arising from her data. For instance, she created a plot (Figure 1(a)) summarizing average study hours on workout days versus non-workout days. She conducted a simulation-based analysis (Figure 1(b)) to test the hypothesis of whether a workout may impact the number of study hours. This exercise allows the student to reflect on how her exercise routine may have affected her productivity and helps her grasp the non-intuitive concept of hypothesis testing through a concrete dataset grounded in her own life experiences. This analysis can lead to a deeper discussion of advanced topics such as causal inference and prescriptive analytics.



(a) (b)

Figure 1: An example of one of the students' work product

Qualitative analysis of students' reflection reveals students' growth of data science knowledge and the emerging ideas of self-awareness and self-regulation as the result of analyzing their own data, even though those topics are not explicitly taught in the course. For example, one student commented, *"Collecting personal data was one of the interesting experiences I had in this class. … This is an interesting idea as I was able to explore more about myself from the collected data. Also, my classmates drew some incredible conclusions/observations/hypotheses from my data that I never realized."* Another student reflects on the personal experiment he did with himself to improve his learning. "*I find it helpful to learn as it gives better insight to what worked for me and what didn't through the collection of the data."*

5 Conclusion and Future Work

Promoting self-regulated learning (SRL) is becoming a popular research goal in learning analytics. While most of the work aims to build students' SRL through learning analytics deliberately designed to support scaffolding, little has been done to explore an alternative approach to promote SRL by empowering and engaging students in collecting, analyzing, and reflecting on their data. With the increasing attention to data science/analytics education and their natural connection to learning analytics, we argue that this is a viable approach to promote students' agency and increase data literacy and feedback literacy (Tsai 2022), which are essential for promoting SRL. Future research will explore systematic approaches to embed PLA in data science/analytics that may scale to large class sizes and diverse student populations.

REFERENCES

Knight, S., & Anderson, T. D. (2016). Action-oriented, Accountable, and Inter (Active) Learning Analytics for Learners. In LAL@ LAK (pp. 47-51).

Tsai, Y. S. (2022). Why Feedback Literacy Matters for Learning Analytics. In Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022, pp. 27-34. International Society of the Learning Sciences.

Enhancing Collaborative Learning with a Myers-Briggs Type Indicator (MBTI) -Driven Conversational AI Agent

Hanqi Li
New York University
hl4893@nyu.edu

Qianzi Wen
Savannah College of Art and Design
qiawen20@student.scad.edu

Ruiwei Xiao
Carnegie Mellon University
ruiweix@andrew.cmu.edu

Xavier Ochoa
New York University
xavier.ochoa@nyu.edu

ABSTRACT: The Myers-Briggs Type Indicator (MBTI) is a widely-used psychometric tool that categorizes personality types based on individual preferences, offering insights that can improve group dynamics by aligning each member's strengths. In collaborative learning, which emphasizes teamwork and interpersonal skill development, understanding each group member's communication style and decision-making approach is essential. MBTI plays a crucial role here by providing valuable insights into these characteristics, enhancing group collaboration. This project develops a conversational AI agent that leverages Myers-Briggs Type Indicator (MBTI) insights to enhance collaborative learning among college students. The AI agent provides personalized feedback based on student's MBTI results, including communication strategies and role suggestions, helping students use their strengths effectively, and also promotes social-emotional learning (SEL), making collaboration both productive and personally enriching. In this study, we collect multi-modal data in classroom experiments, including behavioral data (e.g., interaction patterns), learning data (e.g., task completion rates), and textual data (e.g., language patterns) to assess the agent's effectiveness and find insights.

Keywords: Collaborative Learning, Social Emotional Learning, Myers-Briggs Type Indicator, Conversational Agent, Artificial Intelligence

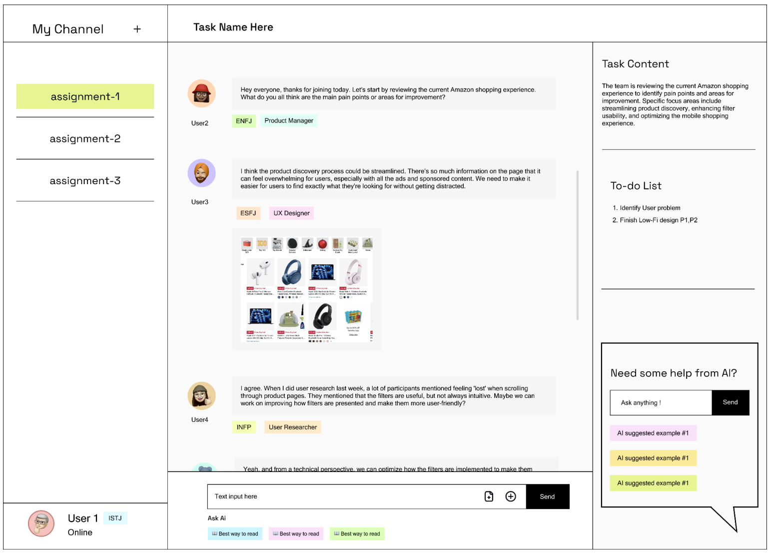


Figure 1: Overview of our project, the details of our demo can be accessed via Figma

1 Introduction

Collaborative learning is a popular pedagogical method that promotes academic success and interpersonal skills by encouraging students to work together toward common goals, leveraging diverse skills and perspectives (Laal & Ghodsi, 2012). Effective teamwork relies on understanding each member's communication style, decision-making approach, and behavior (Tan, Chen, & Chua, 2023). Personality assessments like the Myers-Briggs Type Indicator (MBTI) enhance collaboration by offering insights into these traits (Rodríguez et al., 2013).

MBTI is a widely used psychometric tool, categorizes personality types based on how individuals perceive, decide, and interact with others into 16 personalities (Pittenger, 1993). The MBTI assesses four primary dimensions of personality: Extraversion (E) vs. Introversion (I), Sensing (S) vs. Intuition (N), Thinking (T) vs. Feeling (F), Judging (J) vs. Perceiving (P). Each person's MBTI type is represented by a four-letter code (e.g., ENFP, ISTJ), based on their preferences in these four areas. By examining communication styles, problem-solving methods, and relational dynamics, MBTI helps predict group behavior and offers guidance for improving teamwork (Kwon & Kwag, 2010).

Our study proposes a conversational AI agent that leverages MBTI insights to enhance collaborative learning for students in higher education. Through MBTI-based recommendations, the AI agent encourages students to recognize and apply their unique strengths, ultimately improving teamwork, engagement,and emotional intelligence. The study will focus on two main **R**esearch **Q**uestions (**RQs**): **RQ1.** How does the integration of MBTI into a conversational AI agent enhance collaboration and group dynamics? **RQ2.** To what extent does personalized AI-driven feedback improve student engagement in collaborative learning?

2 Method: Integrating MBTI with Conversational AI

The core of this study is the development of a conversational AI agent that uses MBTI-based insights to enhance the collaborative learning experience. The AI agent will analyze each student's MBTI type based on the open source MBTI database and use this information to provide personalized tutoring and support. For example, the agent can offer targeted guidance on communication strategies, conflict management, and role delegation that align with each student's personality type. During the interaction, the AI agent adapts its feedback based on previous group conversations, as well as students' questions and engagement patterns, making it a dynamic and responsive tool (Figure 1). Beyond academic support, the AI also aims to foster social-emotional learning (SEL) by promoting self-awareness, empathy, and emotional regulation. Through this dual focus on academic and social-emotional outcomes, the AI agent represents an innovative application of MBTI in educational technology, providing both personalized guidance and support in collaborative settings.

3 DATA COLLECTION

To evaluate the effectiveness of the MBTI-driven conversational AI agent in enhancing collaborative learning, we will collect multi-modal data during classroom experiments from a significant sample size to ensure comprehensive analysis and generalizability. Key data sources include:

3.1 *Behavioral Data:*

Behavioral data such as keystrokes, mouse hovers, usage time, frequency, and interaction patterns will be collected to provide insights into user engagement, responsiveness, and interaction flow, revealing how students navigate tasks and interact with AI prompts.

3.2 *Learning Data*

Learning outcomes will be tracked through task completion rates, time on tasks, and group performance metrics. Metrics like conversation frequency, idea generation, and interaction length will assess whether students using the AI agent are more engaged and participatory, highlighting its impact on academic achievement and teamwork.

3.3 *Conversational Data*

Conversation logs will enable analysis of language patterns (e.g., collaborative language, emotional indicators), recommendations, and feedback. Correlating conversation content with learning outcomes will evaluate the AI's personality-based guidance and its influence on emotional intelligence.

The structured data will support further analysis, correlating interaction types and conversational dynamics. This multimodal data can inform future data mining and machine learning efforts, offering a comprehensive view of AI's role in collaborative learning and its effects on academic performance and social intelligence.

REFERENCES

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia-social and behavioral sciences*, *31*, 486-490.

Rodríguez Montequín, V., Mesa Fernández, J.M., Balsera, J.V. et al. Using MBTI for the success assessment of engineering teams in project-based learning. Int J Technol Des Educ 23, 1127–1146 (2013). https://doi.org/10.1007/s10798-012-9229-1

Tan, S. C., Chen, W., & Chua, B. L. (2023). Leveraging generative artificial intelligence based on large language models for collaborative learning. Learning: Research and Practice, 9(2), 125–134. https://doi.org/10.1080/23735082.2023.2258895

Pittenger, D. J. (1993). Measuring the MBTI… and coming up short. *Journal of Career Planning and Employment*, *54*(1), 48-52.

Kwon, Y. H., & Kwag, O. G. (2010). Effect of ready planned small group collaboration learning program through MBTI on interpersonal relationships and career identity of nursing college students. *Journal of the Korea Academia-Industrial Cooperation Society*, *11*(11), 4441-4448.

ActiveAI: Enabling K-12 AI Literacy Education & Analytics at Scale

Ruiwei Xiao[1], Ying-Jui Tseng[1], Hanqi Li[2], Hsuan Nieu[3], Guanze Liao[3], John Stamper[1], Kenneth R. Koedinger[1]
[1]Carnegie Mellon University [2]New York University [3]National Tsing Hua University
ruiweix@andrew.cmu.edu

ABSTRACT: Interest in K-12 AI Literacy education has surged in the past year, yet large-scale learning data remains scarce despite considerable efforts in developing learning materials and running summer programs. To make larger scale dataset available and enable more replicable findings, we developed an intelligent online learning platform featuring AI Literacy modules and assessments, engaging 1,000 users from 12 secondary schools. Preliminary analysis of the data reveals patterns in prior knowledge levels of AI Literacy, gender differences in assessment scores, and the effectiveness of instructional activities. With open access to this de-identified dataset, researchers can perform secondary analyses, advancing the understanding in this emerging field of AI Literacy education.

Keywords: AI Literacy, Learning Analytics, K-12 Education, Online Learning Platform

1 Introduction

K-12 AI literacy education has gained significant attention in the past year (Klopfer et al., 2024). While researchers have made considerable progress in designing learning materials and organizing summer camps, large-scale learning platforms (Tseng et al., 2024) and datasets (Almatrafi et al., 2024) remain limited. To provide accessible AI literacy learning materials for schools, as well as scalable datasets to support replicable research in the learning analytics community, we developed a K-12 AI Literacy learning platform. This platform offers evidence-based learning activities and assessments for classroom use, along with standardized data logging compatible with widely-used educational data repositories for secondary analysis. Over the past year, our efforts in instructional design, platform development, and school partnerships have resulted in the collection of AI literacy learning and assessment data from over 1,000 users across 4 learning modules.

2 Methods

2.1 System Design and Data Pipeline

The platform is implemented as a web application developed using Next.js and OpenAI APIs, supporting three types of users (***students, teachers, researchers***). Example interfaces and data flows for each role are illustrated in Figure 1. In a complete learning experience, **students** generate all the data, and they have access only to their own data. Students begin by completing a survey to provide de-identified demographic information, followed by a sequential process of a pre-test, learning module, and post-test on the assigned topic. The pre- and post-tests are isomorphic assessments targeting the same learning objectives, while the learning module includes interactive activities with an AI agent in simulated real-life scenarios, such as identifying LLM hallucinations in news summaries, to teach AI literacy learning objectives. Learners' interaction data from learning activities and assessments are logged into the learning management system and aggregated at varying levels of granularity for other user roles. ***Teachers*** can access survey and activity data from their classes. Specifically, teachers can view data on time spent, completion rates, and correctness, aggregated by activity, learning objective, module, student, and class. This real-time data aggregation allows teachers to make informed adjustments to their instructional plans and provide timely support to students in need. ***Researchers*** can access all types of de-identified data and their aggregated forms. Instead of providing interfaces and visualizations, the platform supports downloading standardized data logs compatible with widely-used educational data repositories (e.g., DataShop[1], LearnSphere[2]) and data analysis tools (e.g., RStudio, Tableau) for public access and analysis. With different levels of aggregation, our data supports learning analytics, including learner modeling, validation of existing learning sciences principles, and learning engineering within AI literacy as a new domain.

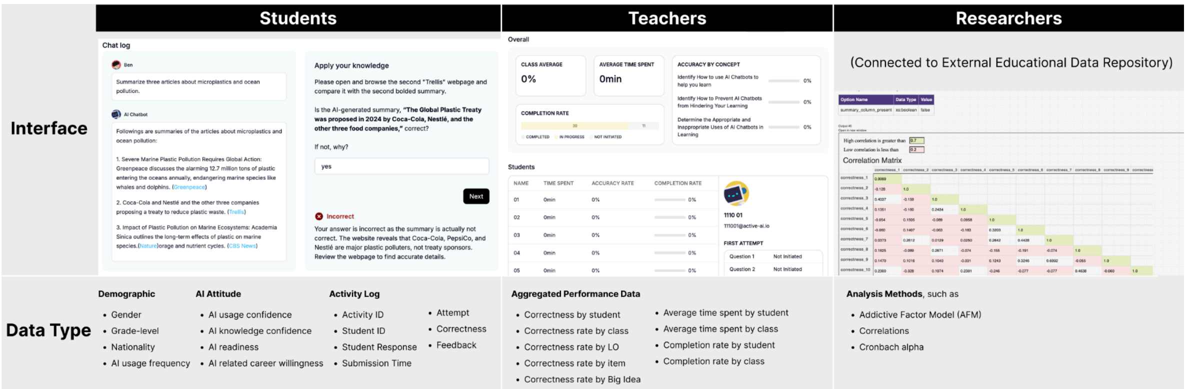


Figure 1: System Design and Data Pipeline

2.2 Learning Design on K-12 AI Literacy Modules

We use a backward design approach (Wiggins & McTighe, 1998) to create learning objectives, assessments, and activities aligned with AI literacy standards (Touretzky et al., 2023) and priority topics from partner schools. For example, to address teachers' interest in identifying LLM hallucinations, we map this skill to AI4K12's Big Idea #5: Societal Impact, design corresponding assessments, and develop interactive, scaffolded activities with feedback (Figure 1). Empirical examples are detailed in our prior work (Tseng et al., 2024).

3 Preliminary Results and Future Works

We partnered with 12 secondary schools across North America, Asia, and Australia. Eight learning modules were implemented, with four deployed in four schools along with surveys and pre- and post-tests. Over 1,000 unique learners used the platform, and 426 (171 for modules 1 and 2, 131 for module 3, and 114 for module 4) completed all components, providing complete data for analysis. To triangulate the results and enhance the interpretability of our findings, we conducted teacher interviews and student cognitive task analysis. Preliminary results indicate learning gains, gender differences, and variations across educational contexts.

[1] DataShop: https://pslcdatashop.web.cmu.edu/

[2] LearnSphere: https://learnsphere.org/

Across the 7 learning objectives in 4 modules, we observed significant learning gains in 4 of them, based on Wilcoxon tests on pre- and post-test scores on the non-normal distributed data. Learning gains correlated with cognitive engagement levels (ICAP framework, Chi & Wylie, 2014): objectives with significant gains involved interactive activities, while those with smaller gains were linked to passive reading. This highlights the importance of cognitive engagement, though further analysis is needed to identify areas for improvement.

In the latest experiment involving module 4 (Identify LLM Hallucinations), where gender data is available, non-male students achieved significantly higher assessment scores on both the pre-test ***(f=6.97, p<0.01, one-way ANOVA)*** and post-test ***(f=6.80, p=0.01, one-way ANOVA)*** compared to their male counterparts, along with slightly higher learning gains. These findings help researchers identify threats to activities validity and improve them through targeted interventions. In the future, as 2 partner schools implement 4 learning modules by year-end, we will collect AI literacy data over a longer duration. Additionally, while Asian schools offer standalone IT courses dedicated to AI literacy, schools in other regions integrate our materials into regular subjects or STEM clubs. These differences in learning contexts will also enrich our dataset.

4 CONTRIBUTIONS

This study makes three key contributions to the LAK community: 1. **A Novel Dataset**: We provide a dataset in the emerging and understudied domain of AI literacy, capturing learners' prior knowledge, interactions with LLM systems, and learning outcomes, enriched with demographic and contextual information. 2. **Open Access for Secondary Analysis**: The dataset will be made openly available on established educational data repositories, offering a valuable resource for secondary analysis and enabling broader research fields. 3. **Practical AI Literacy Resources:** A set of AI literacy learning activities are provided to support practitioners in integrating AI concepts into diverse classroom environments. These contributions aim to advance knowledge, research and practice in AI literacy education, fostering a deeper understanding and scalable approaches through learning analytics.

REFERENCES

Almatrafi, O., Johri, A., & Lee, H. (2024). A Systematic Review of AI Literacy Conceptualization, Constructs, and Implementation and Assessment Efforts (2019-2023). *Computers and Education Open*, 100173.

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

Klopfer, E., Reich, J., Abelson, H., & Breazeal, C. (2024). Generative AI and K-12 Education: An MIT Perspective.

Touretzky, D., Gardner-McCune, C., & Seehorn, D. (2023). Machine learning and the five big ideas in AI. *International Journal of Artificial Intelligence in Education*, *33*(2), 233-266.

Tseng, Y. J., Xiao, R., Bogart, C., Savelka, J., & Sakr, M. (2024). Assessing the Efficacy of Goal-Based Scenarios in Scaling AI Literacy for Non-Technical Learners. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2* (pp. 1842-1843).

Tseng, Y. J., Yadav, G., Hou, X., Wu, M., Chou, Y. S., Chen, C. C., ... & Koedinger, K. R. (2024, September). ActiveAI: The Effectiveness of an Interactive Tutoring System in Developing K-12 AI Literacy. In *European Conference on Technology Enhanced Learning* (pp. 452-467).

Wiggins, G., & McTighe, J. (1998). What is backward design. *Understanding by design*, *1*, 7-19.

Grade Prediction Using fastText Features Weighted Through Differential Pattern Mining

Ryota Tachi[1], Tsubasa Hirakawa[2], Takayoshi Yamashita[3], Hironobu Fujiyoshi[4]
Chubu University
{tatti[1], hirakawa[2]}@mprg.cs.chubu.ac.jp, {takayoshi[3], fujiyoshi[4]}@isc.chubu.ac.jp

ABSTRACT: With the digitization of educational materials, there is growing anticipation for predicting grade performance using learning log data. Previous studies have attempted to predict performance by inputting histogram features of the number of digital material operations into machine learning models. However, these approaches do not consider temporal sequences, making it difficult to reflect behavioral patterns in the performance predictions. To address this issue, we propose maintaining the time series using fastText, which embeds learning behaviors as features. Additionally, we employ differential pattern mining to detect behavior patterns that exhibit significant differences and then apply weighting to these patterns in fastText. Evaluation experiments show that our proposed method improves performance prediction accuracy compared to conventional methods and that weighting behavioral patterns proves effective.

Keywords: Grade Prediction, Differential pattern mining, Sequential Data

1 INTRODUCTION

As educational environments go digital, the use of machine learning models in education is drawing attention. Many studies predict performance from learning behaviors for early dropout detection and improving learning, but most rely on operation frequency histograms (Kohama et al., 2023), ignoring behavioral patterns. To address this, we propose a performance prediction method using fastText features weighted by differential pattern mining and verify the effectiveness of input data that reflects these behavioral patterns.

2 PROPOSED METHOD

We propose an operation log embedding method that applies weighting based on differential patterns. Figure 1 provides an overview of the method, which consists of three modules: differential pattern mining, operation log embedding, and classification prediction. First, differential pattern mining identifies patterns that differ significantly between high- and low-achieving students. Next, we use E2Vec (Miyazaki et al., 2024) preprocessing and embedding modules to generate operation log embedding features. E2Vec converts each operation log into a single character and aligns it with the NLP concepts of “character,” “word,” and “sentence,” then uses fastText for embedding. For instance, NEXT becomes ‘N,’ PREV ‘P,’ and OPEN ‘O.’ In E2Vec, logs within one minute and up to 15 operations are treated as a “word,” embedded, and averaged to obtain the operation log embedding vectors. In our proposed method, we acquire a sentence embedding vector that reflects learning patterns through a weighted average, where the difference in the proportion of students who performed different patterns is used as the weight. The embedding vector v_S of a sentence is computed as $v_S =$

$\frac{\sum_{i=1}^{m} w_i \cdot \frac{u_i}{|u_i|}}{\sum_{i=1}^{m} w_i}$. Here, w_i is the weight, and u_i is the embedding vector of the word. Using the operation log embedding features obtained as described above as input, we perform grade classification predictions using a classifier.

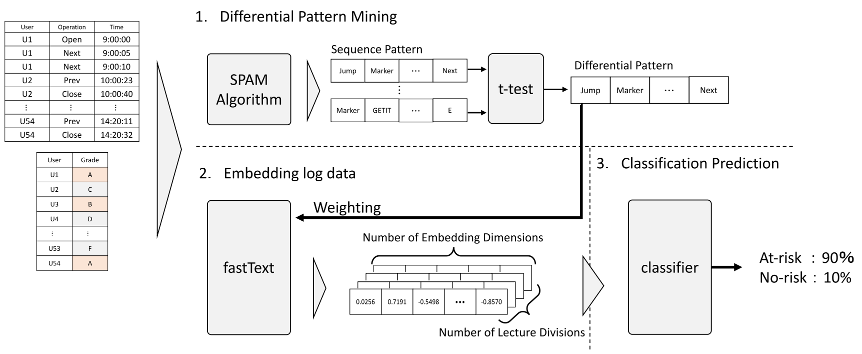


Figure 1：Overview of the proposed method

3 EVALUATION EXPERIMENT

We compare the proposed method with E2Vec, a conventional histogram-based method.

3.1 Experimental conditions

We use operation log data collected from six courses at Kyushu University. Among them, courses A-2020 and D-2020 are used solely for identifying differential patterns, while the remaining courses are used for the classification prediction task. For differential pattern mining, the SPAM algorithm is employed with a maximum pattern length of 15, a minimum support of 40%, and a maximum gap of 2. In this experiment, students are classified into two classes: No-risk and At-risk. The classification models used are RandomForest, XGBoost, and SVM.

3.2 Experimental results

We compare the grade classification accuracy of the proposed method with that of conventional methods in Table 1. Of the four courses, one is used as the training dataset, while the remaining three serve as evaluation datasets. Table 1 presents the average accuracy across evaluations conducted on each evaluation dataset. The results indicate that, except when using A-2022 as the training data, the proposed method achieved higher classification accuracy than the conventional method.

Table 1：Comparison of average accuracy per training data

Train	E2Vec	$\text{Propose}_{\text{RF}}$	$\text{Propose}_{\text{XGB}}$	$\text{Propose}_{\text{SVM}}$
A-2021	0.6268	**0.6280**	0.6134	0.6162
A-2022	**0.7237**	0.6087	0.5324	0.6213
D-2021	0.4869	**0.5745**	**0.6044**	**0.5528**
D-2022	0.6115	**0.6649**	**0.6696**	**0.6341**

4 ANALYSIS

This chapter analyzes patterns identified through differential pattern mining. Table 2 shows patterns that differ significantly between the two classes, as well as those that show minimal differences. According to Table 2, the patterns with a large difference between the two classes involve alternating "NEXT" and "PREV" operations. This suggests that after opening the material, students frequently use "PREV," indicating they are reviewing previously covered lecture content. In particular, high-achieving students repeatedly check preceding and following pages, implying a conscious effort to grasp the contextual flow of the material. This behavior suggests that their grade performance differences may stem from more active and intentional engagement with the course content. On the other hand, patterns with minimal differences involve repetitive "NEXT" or "PREV" operations—redundant actions often observed within the first 0 to 10 minutes of the lecture. These may reflect attempts to quickly navigate to specific pages used during the lecture. Moreover, some students may be simply tracing the instructor's own operations, repeatedly clicking "NEXT" or "PREV." Such repetitive sequences are particularly common among lower-achieving students and do not directly correlate with active learning. They may indicate a lower level of concentration or an attempt to mimic the instructor's actions rather than engaging deeply with the material.

Table 2: Patterns with Large and Small Differences

Pattern	Proportion difference	Pattern	Proportion difference
ONPPPN	1.000000	PPNNNNNNNNNNCO	0.020000
ONPPNPN	1.000000	NPPPPPPPO	0.024355
ONNPPNPN	1.000000	NNPPPPPPC	0.024355
NPNPNNC	0.669145	NNCJN	0.024355
ONNPNPNC	0.669145	PNPNNCO	0.024451

5 CONCLUSION

In this study, we demonstrated the effectiveness of weighting using differential patterns. Furthermore, we were able to identify patterns of active engagement in learning as positive features, and redundant patterns of continuous operations as negative features.

REFERENCES

Kohama, H., Ban, Y., Hirakawa, T., Yamashita, T., Fujiyoshi, H., Itai, A., & Usami, H. (2023). Recommending learning actions using neural network. In International Conference on Computers in Education 2023.

Miyazaki, Y., Švábenský, V., Taniguchi, Y., Okubo, F., Minematsu, T., Shimada, A. (2024). E2Vec: Feature Embedding with Temporal Information for Analyzing Student Actions in E-Book Systems. Proceedings of the 17th International Conference on Educational Data Mining, 434--442. https://doi.org/10.5281/zenodo.12729854

Analyzing Lecture Slides Generated by AI Presentation Makers with RAG

Itsuki Sano, Kazutoshi Sumiya
Kwansei Gakuin University, Japan
gil06518@kwansi.ac.jp, sumiya@kwansei.ac.jp
Yuanyuan Wang
Yamaguchi University, Japan
y.wang@yamaguchi-u.ac.jp
Yukiko Kawai
Kyoto Sangyo University, Osaka University
kawai@cc.kyoto-su.ac.jp

ABSTRACT: Generative artificial intelligence (AI) is now widely used in educational contexts and includes tools such as AI presentation makers that automatically produce slides from simple keywords or documents. However, certain prompts can lead to slides that deviate from the intended content. This study focuses on generating slides that accurately reflect textbook information using retrieval-augmented generation (RAG), a model combining information retrieval with text generation, which along with its extended version, GraphRAG, was used to create draft slides. A comparison of the draft slides generated by RAG and GraphRAG suggests the effectiveness of incorporating structured graph data to improve slide accuracy and content consistency.

Keywords: Lecture Slides, Generative AI, RAG

1 INTRODUCTION

In recent years, generative artificial intelligence (AI) has streamlined many educational tasks in a wide range of applications, such as English conversation partners and programming guidance, and is expected to promote student understanding and active participation in class. In addition, the use of AI presentation makers, who automatically generate slides by inputting words or documents, can improve the efficiency of class preparation for instructors. In this study, we propose and compare two methods for generating the draft slides that are input into AI presentation makers, namely, retrieval-augmented generation (RAG) and GraphRAG. The textbook content is complex, with words related to each other across chapters. Therefore, a GraphRAG-based method that uses textbook content is proposed as graph data and compared with the RAG-based method.

2 UTILIZING RAG FOR AI PRESENTATION MAKERS

2.1 Prompt Generation Using RAG

Recently, artificial intelligence (AI) presentation-makers have garnered attention for their ability to create slides using simple drafts. The quality of the generated slides depends on the input, and detailed inputs yield more accurate slides although increasing the burden of slide creators, whereas sparse inputs can result in inaccurate slides. To address this issue, we propose the use of retrieval-augmented generation (RAG) (Gu et al., 2020), which is a framework that combines information retrieval and text generation and is especially useful for tasks requiring open-domain question answering. RAG consists of two steps:

- **Search**: Finds relevant information from external documents using vector-based search.
- **Generate**: Produces natural language text based on retrieved information.

Graph retrieval-augmented generation (GraphRAG) (Peng et al., 2023) builds on RAG by using graph-structured data to represent relationships, allowing more complex answers by connecting entities and their relationships. It has the steps:

- **Graph Search**: Retrieves related entities and their relationships from a knowledge graph to capture indirect as well as direct information.
- **Generate**: Creates context-aware responses based on graph data, often combined with vector search.

These methods also easily adapt to different textbooks or revisions by changing the external data without modifying the language model itself

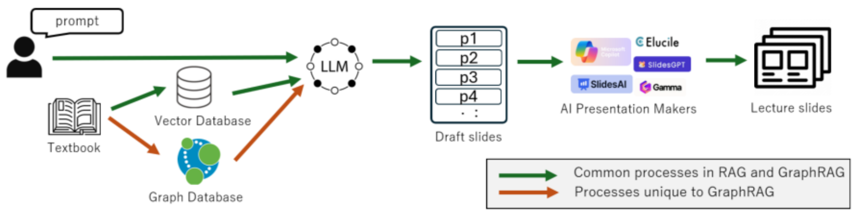


Figure 1: System Overview

As shown in Fig. 1, we used RAG and GraphRAG to create draft slides for AI presentation makers. The slide generation process using GraphRAG involves combining a large language model (LLM) with a graph database (Neo4j) to effectively retrieve information and produce accurate answers. Key steps include:

- **Data Preparation**: Document data are tokenized and divided into chunks (512 tokens each) for easy analysis.
- **Data Structuring**: Using the GPT-4-based model "gpt-4o-mini," documents are converted into nodes and relationships, then stored in a graph database (Neo4j) to maintain information.
- **Structured Search**: Nodes are extracted from a query, identifying related data in Neo4j using a full-text search.
- **Unstructured Search**: Documents and queries are vectorized using Neo4jVector and OpenAI's embedding model, and semantically related documents are retrieved by comparing vector similarities.
- **Answer Generation**: Both structured and unstructured data are input into the language model to generate a natural language response that combines highly relevant information.

In addition, the RAG model is defined only as an unstructured search model and does not perform a structured search among the above steps.

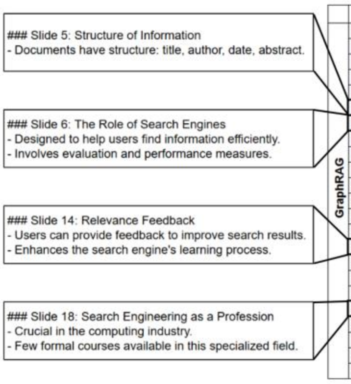
2.2 Comparison of Prompt Generation Methods

This section compares the drafts generated by RAG and GraphRAG. In this study, drafts were generated for each textbook[1] chapter. We used the slides[2] published as appendices in textbooks as

[1] W. Bruce Croft, Donald Metzler, and Trevor Strohman. 2010. Search Engines: Information Retrieval in Practice (1st ed.). Addison-Wesley, Boston, MA.

[2] "Search Engines Book." Accessed 9 Dec. 2024, http://www.search-engines-book.com/.

reference slides for our analysis. The prompt used to generate the drafts was: "Create a 22-slide presentation titled 'Search Engines and Information Retrieval' covering the sections: 'What Is Information Retrieval? 'The Big Issues,' 'Search Engines,' and 'Search Engineers." We set the number of slides equal to the reference slides and included section titles from the textbook. Each slide's content, for both the RAG and GraphRAG outputs, was represented as a 550-dimensional term frequency–inverse document frequency (TF-IDF) vector, with 550 total number of unique words across all slides. These vectors were compressed into 300-dimensional representations using Word2Vec (word2vec-google-news-300). We then calculated the cosine similarity between the reference slides and RAG/GraphRAG slides. Fig. 2 shows the cosine similarity between the GraphRAG and reference slides.



GraphRAG \ Reference Slide	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	p22
p1	0.60	0.72	0.70	0.64	0.73	0.68	0.62	0.66	0.71	0.67	0.55	0.67	0.67	0.67	0.69	0.70	0.72	0.69	0.71	0.64	0.63	0.66
p2	0.61	0.76	0.88	0.71	0.74	0.77	0.73	0.73	0.73	0.71	0.61	0.78	0.73	0.75	0.76	0.76	0.76	0.66	0.73	0.75	0.69	0.75
p3	0.54	0.78	0.74	0.72	0.72	0.72	0.74	0.68	0.66	0.70	0.73	0.75	0.71	0.70	0.71	0.73	0.72	0.67	0.72	0.72	0.63	0.75
p4	0.50	0.66	0.74	0.76	0.75	0.74	0.73	0.68	0.74	0.63	0.68	0.70	0.73	0.73	0.67	0.73	0.65	0.64	0.73	0.65	0.65	0.70
p5	0.54	0.62	0.67	0.68	0.70	0.65	0.62	0.52	0.63	0.58	0.67	0.63	0.60	0.64	0.59	0.63	0.59	0.55	0.66	0.58	0.56	0.55
p6	0.63	0.75	0.76	0.76	0.77	0.76	0.77	0.72	0.77	0.74	0.67	0.75	0.79	0.82	0.81	0.81	0.78	0.77	0.81	0.74	0.71	0.76
p7	0.62	0.74	0.69	0.67	0.73	0.70	0.69	0.63	0.72	0.77	0.64	0.74	0.76	0.67	0.74	0.74	0.76	0.74	0.72	0.71	0.62	0.73
p8	0.59	0.65	0.65	0.63	0.69	0.65	0.65	0.60	0.68	0.66	0.60	0.66	0.67	0.73	0.69	0.68	0.73	0.71	0.71	0.65	0.65	0.67
p9	0.62	0.64	0.63	0.62	0.63	0.65	0.64	0.62	0.62	0.67	0.57	0.68	0.62	0.65	0.74	0.68	0.68	0.66	0.65	0.64	0.65	0.61
p10	0.60	0.78	0.76	0.70	0.71	0.74	0.71	0.76	0.75	0.78	0.66	0.72	0.76	0.77	0.78	0.74	0.77	0.75	0.77	0.79	0.68	0.71
p11	0.63	0.65	0.66	0.61	0.64	0.66	0.65	0.69	0.70	0.75	0.59	0.67	0.67	0.68	0.75	0.69	0.74	0.71	0.72	0.74	0.67	0.66
p12	0.62	0.74	0.76	0.70	0.74	0.75	0.73	0.75	0.79	0.81	0.67	0.78	0.74	0.78	0.79	0.77	0.80	0.76	0.78	0.79	0.72	0.75
p13	0.56	0.72	0.71	0.61	0.75	0.72	0.70	0.69	0.72	0.68	0.64	0.69	0.73	0.70	0.77	0.71	0.71	0.68	0.71	0.69	0.66	0.70
p14	0.67	0.74	0.73	0.69	0.77	0.75	0.73	0.73	0.77	0.76	0.64	0.76	0.77	0.75	0.79	0.78	0.78	0.76	0.77	0.76	0.74	0.77
p15	0.60	0.71	0.72	0.66	0.69	0.75	0.75	0.74	0.71	0.75	0.73	0.71	0.72	0.75	0.75	0.66	0.73	0.70	0.75	0.76	0.68	0.68
p16	0.60	0.76	0.76	0.71	0.79	0.77	0.74	0.73	0.79	0.76	0.65	0.75	0.78	0.75	0.77	0.80	0.77	0.75	0.80	0.79	0.70	0.80
p17	0.49	0.75	0.69	0.65	0.72	0.69	0.75	0.70	0.69	0.71	0.67	0.70	0.74	0.73	0.72	0.69	0.69	0.64	0.69	0.69	0.59	0.72
p18	0.63	0.78	0.79	0.70	0.78	0.79	0.73	0.76	0.79	0.77	0.64	0.74	0.77	0.78	0.76	0.77	0.75	0.75	0.76	0.73	0.70	0.75
p19	0.60	0.74	0.75	0.70	0.70	0.71	0.74	0.72	0.78	0.73	0.66	0.70	0.74	0.73	0.77	0.75	0.74	0.70	0.76	0.73	0.72	0.72
p20	0.53	0.70	0.68	0.70	0.72	0.69	0.67	0.68	0.65	0.69	0.68	0.73	0.68	0.74	0.69	0.71	0.72	0.66	0.73	0.72	0.64	0.67
p21	0.54	0.68	0.69	0.69	0.71	0.70	0.69	0.64	0.66	0.62	0.65	0.72	0.68	0.69	0.66	0.73	0.69	0.63	0.71	0.65	0.64	0.65
p22	0.42	0.57	0.56	0.62	0.63	0.60	0.63	0.54	0.60	0.53	0.54	0.61	0.59	0.56	0.59	0.61	0.55	0.54	0.61	0.54	0.54	0.57

Figure 2: The cosine similarity between GraphRAG and a reference slide (bottom.)

3 DISCUSSION AND CONCLUSION

In calculating the average similarity between corresponding pages, RAG was 0.68 and GraphRAG was 0.71. The results indicated that GraphRAG's output was closer to that of the reference slides. The differences were significant for P5, P6, P14, and P18. Because P5 and P18 are referenced by the unstructured data used in both methods, the graph was not effective. However, some content was unique to GraphRAG and was not output by the standard RAG. The P6 of the GraphRAG included sentences such as "Designed to help users find information efficiently" and "Involves evaluation and performance measures," which align with the structured data in the graph (Search Engines - MEASURE -> Effectiveness, Search Engines - MEASURE -> Efficiency.) "Relevance Feedback" on P14 corresponded to the structured data from the graph (Search Engines - AFFECTED-BY -> Relevance Feedback and Web Search Engines - USED-IN -> Relevance Feedback). These results suggest that using graphs enables the inclusion of details that might otherwise be overlooked, thereby enhancing the draft's alignment with complex information. Future work will include an evaluation of the educational effectiveness of these lecture slides, in addition to cosine similarity. In addition, we verified whether the same results could be obtained for other textbooks.

REFERENCES

Gu, K., Lewis, P., Mohammad, H., & Yih, W. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Advances in Neural Information Processing Systems (NeurIPS).

Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., Zhang, Y., & Tang, S. (2023). Graph Retrieval-Augmented Generation for Knowledge-Enhanced Language Models. arXiv preprint arXiv:2107.07578.

This work was supported by JSPS KAKENHI Grant Number JP24K21464 and MEXT as "Developing a Research Data Ecosystem for the Promotion of Data-Driven Science".

Leveraging Learning Analytics to Investigate the Relationship between Course- and Work-based Learning Performance

Wanruo Shi[1], Abhinava Barthakur[2], Vitomir Kovanovic[2], Shane Dawson[2], Xibin Han[1]
[1]Tsinghua University; [2]University of South Australia
swr22@mails.tsinghua.edu.cn

ABSTRACT: This study investigates the relationship between course- and work-based learning in Initial Teacher Education (ITE) programs in Australia. While these programs aim to integrate theoretical and practical learning, how these components interact to build professional competence in ITE students remains unclear. Using clustering techniques, the study identified four distinct student profiles based on their performance in academic and workplace settings. Cluster 1, "Hands-on" students, demonstrated strong workplace performance despite lower academic achievement, while Cluster 2, "Well-rounded" students, excelled in both academic and professional settings. Cluster 3, "Theoretical" students showed high academic achievement but struggled in the workplace, and Cluster 4, "Struggling" students underperformed in both areas. These findings suggest that academic success does not always align with practical effectiveness, revealing multiple pathways of professional development. This diversity highlights the need for adaptable curricula that support professional expertise and competencies development, facilitating students' transition from academic settings to professional roles.

Keywords: Clustering analysis, Initial teacher education, Course- and work-based learning

1 INTRODUCTION

Course-based and work-based learning are essential in preparing students for professional roles, particularly in programs like Initial Teacher Education (ITE) (Collin & Tynjälä, 2003). In this study, Course-based learning focuses on building theoretical knowledge and practical skills within an academic setting, while work-based learning offers real-world experience to apply these skills. Bridging the gap between academic knowledge and workplace application remains challenging, especially for novice ITE students, as strong academic performance doesn't always equate to practical effectiveness (Baartman & De Bruijn, 2011). Although previous research has explored this theory-practice connection, it mainly relies on qualitative methods, leaving a need for quantitative and data-driven insights. This study addresses this gap by analyzing learning assessment records from ITE programs at a large Australian university. Using k-means clustering, it examines the relationship between students' course performance and workplace outcomes, measured against the Australian Professional Standards for Teachers (APST) (AITSL, 2022). The results aim to offer a data-driven understanding of how course- and work-based learning interact, supporting well-prepared future educators.

2 METHODS

The current study analyzed the assessment data from four-year ITE programs at a large Australian public university. The ITE programs aim to equip students with the practical skills and theoretical knowledge needed to excel as qualified educators. Data on ITE students' learning performance were gathered from two datasets on Moodle platforms: the course learning assessment dataset and the work-based learning assessment dataset, encompassing 1,344 students. Both datasets evaluated students' learning performance based on the APST, which includes three key dimensions and seven professional standards, as illustrated in Table 1. The course assessment dataset comprises all assignment grades throughout the students' studies covering the first four professional standards (S1 to S4), while the work-based learning assessment dataset reflects practical performance during placements covering all the seven professional standards (S1 to S7).

Table 1: Australian Professional Standards for Teachers (APST)(AITSL, 2022)

Dimension	Professional standard
Professional Knowledge	**S1**: Know students and how they learn
	S2: Know the content and how to teach it
Professional Practice	**S3**: Plan for and implement effective teaching and learning
	S4: Create and maintain supportive and safe learning environments
	S5: Assess, provide feedback and report on student learning
Professional Engagement	**S6**: Engage in professional learning
	S7: Engage professionally with colleagues, parents and the community

Note: See more details at https://www.aitsl.edu.au/standards/graduate

We matched the two datasets, and the combined data set was standardized. We then employed the *k*-means clustering algorithm, a standard centroid-based clustering algorithm, using the `cluster` R package. The optimum number of clusters (*k*) was determined using the Elbow Method and validated by the average silhouette width and Davies-Bouldin Index. Finally, we labelled and interpreted the classification of each cluster.

3 RESULTS

Aiming to illustrate the connection between course and work-based learning performance, this study combined these two datasets and applied *k*-means clustering analysis. To find the optimum number of clusters (*k*), we conducted the Elbow Method and looked for the "elbow point" where the rate of decrease sharply slows. Therefore, we chose k=4 as the optimum cluster number. To validate the consistency within clusters, we plotted the average silhouette width and Davies-Bouldin Index for different numbers of clusters ranging from *k*=2 to *k*=9, reaffirming k=4 as the optimal number. Figure 1 graphically represents the characteristics of the four clustering groups.

Cluster 1: "Hands-on" (26%) - This cluster is characterized by relatively low achievement at course-based assignments but above-average performance in the workplace placements.
Cluster 2: "Well-rounded" (23%) - Students in this cluster showed above-average performance at course-based assignments and very high achievement in work placements, identifying them as good academic and professional performers. However, this was the smallest group.
Cluster 3: "Theoretical" (24%) - Mirroring Cluster 1, this cluster is characterized by high achievement in course-based assignments. However, their performance in the workplace placements was below average.
Cluster 4: "Struggling" (27%) - Mirroring Cluster 2, this cluster comprises students who underachieve both academically and professionally, with below-average performance at school and relatively low performance in work placements.

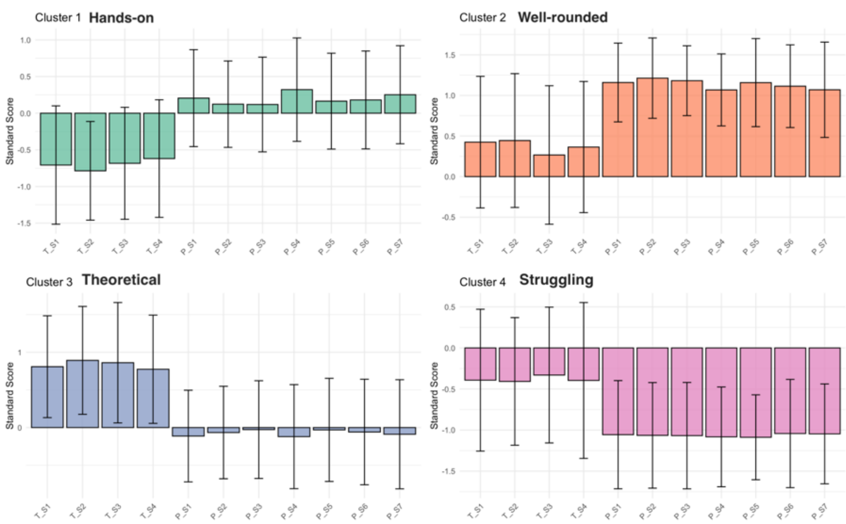


Figure 1: Clustering results by k-means (The vertical whiskers represent the standard deviation)

4 IMPLICATIONS AND FUTURE WORK

This study examined the relationship between coursework and work-based learning, revealing varied performance patterns. Half of the students showed consistent results: they either excelled or struggled in both academic and professional settings, suggesting a link between theoretical knowledge and practical competence. Strong academic performers often succeeded in the workplace, while underperformers faced challenges in both areas. However, the other half exhibited contrasting patterns: one-quarter excelled academically but struggled in the workplace, while another quarter performed well professionally despite weaker academic results. These findings indicate that academic success doesn't always translate to practical effectiveness, and some students with lower academic performance can still excel in real-world settings. This suggests further research needed to examine specific coursework elements to better support students' transition to professional roles, recognizing diverse pathways for integrating theoretical and practical skills. Program directors and stakeholders should consider these nuances, offering flexible curriculum options that address varied educational needs.

REFERENCES

Australian Institute for Teaching and School Leadership （AITSL）. (2022). Australian Professional Standards for Teachers. https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf

Baartman, L. K. J., & De Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. In Educational Research Review (Vol. 6, Issue 2). https://doi.org/10.1016/j.edurev.2011.03.001

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

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

Facilitating data-informed teaching decisions in Learning Design using Learning Analytics: A Toolkit Experiment

Chathura K. Sooriya-Arachchi
Queen Mary University of London
c.k.sooriyaarachchi@qmul.ac.uk

ABSTRACT: Despite the growing importance of learning analytics in higher education, educators face significant challenges in effectively incorporating learning analytics into their teaching practices. This poster presents a systematically developed toolkit that bridges this gap through a structured approach to data-informed teaching decisions in learning design. The toolkit integrates intended learning outcomes, pedagogic intentions, and learning design activity types with corresponding learning analytics markers and analysis methods. An experimental study evaluated the toolkit's effectiveness with 12 educators across multiple disciplines (66% Computer Science, 34% from Social Sciences, Law and Medicine). Evaluation results showed strong adoption potential, with 100% of educators finding the toolkit useful, 89% reporting interface reliability, and 78% noting improved learning outcomes. Areas for enhancement were also identified, such as interface simplicity and VLE integration (67% compatibility). The toolkit's systematic approach to learning analytics integration was validated through a perceived usability questionnaire. This demonstrates its potential for facilitating data-informed teaching practices. This poster presents the toolkit, empirical evaluation findings, and evidence-based improvements for supporting wider adoption of learning analytics in higher education teaching practice.

Keywords: learning analytics for learning design, toolkit experiment, actionable learning analytics, data-informed decision-making

1 OVERVIEW

Despite the potential of Learning Analytics (LA) in higher education, educators face significant challenges in effectively incorporating it into their teaching practices, resulting in reduced adoption rates (Macfadyen, 2022). The LA lifecycle emphasises 'acting' on data to enhance student learning experiences (Khalil, M; Ebner, 2015), yet existing frameworks often lack practical implementation guidance. While tools like AL4LD (Hernández-Leo et al., 2019) and The Loop (Bakharia et al., 2016) offer analytics integration approaches, they primarily focus on theoretical frameworks or conceptual models rather than practical tools for educators to implement in their teaching practice. This study addresses this gap through a systematic toolkit that integrates LA with Learning Design (LD) decisions. The toolkit operates within existing Virtual Learning Environment (VLE) contexts, allowing educators to map learning analytics markers (e.g. engagement and performance metrics) to intended teaching decisions. Built on TPACK framework (Liu & Koedinger, 2017) and pedagogical intention alignment (Blumenstein, 2020), the toolkit provides a structured matrix that maps LA markers to pedagogical decisions, incorporating LD activity types from OULDI (Rienties & Toetenel, 2016) while extending it to include analytics-driven decision support. To validate the toolkit's effectiveness in systematising LA integration, its implementation was evaluated with educators through an experimental study, guided by the following research question and objectives:

RQ-1: How can learning analytics data be effectively systematised to support data-informed teaching decisions in learning design?
RO1.1: Construct a toolkit integrating learning analytics data with learning design decisions in teaching
RO1.2: Evaluate toolkit usability and effectiveness in actioning learning analytics within teaching practice

2 TOOLKIT DESIGN AND EVALUATION METHODOLOGY

A comprehensive toolkit matrix was constructed based on findings from a systematic literature review and LA adoption survey, mapping learning analytics indicators to learning design constructs within teaching decisions across planning, delivery, and reflection phases. This matrix formed the foundation for the Learning Activity Planner, designed to guide educators in integrating learning analytics into their teaching practice, addressing RO1.1.

2.1 Toolkit Learning Activity Planner implementation

The toolkit facilitates educators to make informed teaching decisions within existing VLE environments by allowing them to map available learning analytics data to their learning design choices. It provides a structured workflow for aligning learning activities with learning outcomes, pedagogical intentions, and LD activity types with corresponding analytics markers and methods (Figure 1). Educators can access VLE analytics visualisations and metrics based on the toolkit's mapping process to analyse learner engagement patterns, monitor assessment performance, and reflect on learning outcomes achievement using analytics-informed insights.

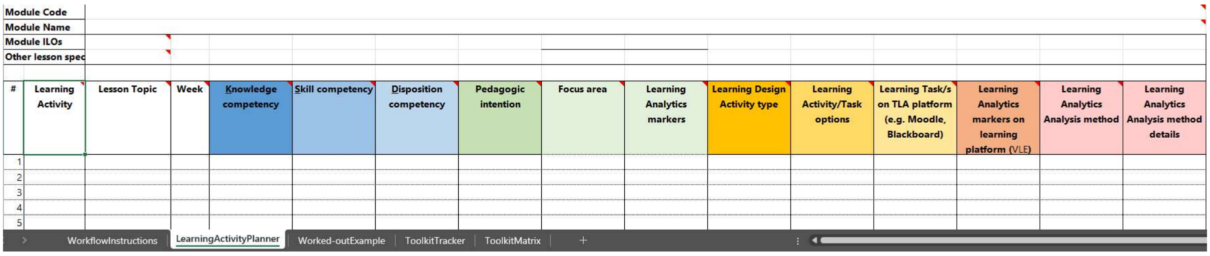


Figure 1: Toolkit spreadsheet screen capture

2.2 Evaluation Framework with Perceived Usability Questionnaire

The toolkit's effectiveness in facilitating educators' learning analytics actionability within teaching practice was evaluated using a perceived usability questionnaire. The questionnaire, consisting of 30 structured items and 8 open-ended questions, assessed the toolkit's practical utility in systematising learning analytics integration into planning, delivery, and reflection tasks across the teaching cycle, addressing RO1.2. It measured key aspects of the toolkit, including core usability, implementation capabilities, pedagogical integration, and teaching practice impact. This enabled a systematic assessment of toolkit's functionality and its effectiveness in supporting data-informed teaching decisions while providing educators an opportunity to share their experiences engaging with toolkit.

3 RESULTS

The mixed-methods evaluation, combining quantitative usability metrics and qualitative thematic analysis, validates the toolkit's effectiveness. Despite the small sample size (n=12), the diversity of disciplines and consistent positive results suggest wider applicability. Key findings include strong adoption potential (100% found useful), enhanced analytics alignment with learning outcomes (89%), and improved activity planning (78%). Qualitative analysis revealed benefits in structured planning, learning outcome alignment, and student monitoring across disciplines, demonstrating the toolkit's effectiveness in systematically connecting learning analytics with pedagogical decisions in various contexts.

Table 2: Key Findings from participant profiling and usability evaluation

Category	Key findings
Participant demographics	- Predominantly Computer Science educators (66% Computer Science, 34% Social Sciences, Law and Medicine) - Experienced academic staff (up to 20 years academic experience) - Teaching across undergraduate and master's levels
Usability Metrics	- High usefulness rating (100% found useful) - Strong interface reliability (89% reliable/easy-to-use) - Good VLE compatibility (67% VLE compatible) - Scaling capabilities (78% good scaling) - Time management concerns (78% time-consuming)

Table 3: Summary findings from thematic analysis of toolkit evaluation

Toolkit positives	Areas for improvement	Usage experience
- Structured approach to activity planning - Clear objective setting and outcome mapping - Enhanced student tracking and engagement monitoring	- Need for interface simplification - Additional LA guidance and support features - Dashboard visualisation and VLE integration requirements - Need for interface simplification	- Initial time investment required - Teaching strategy alignment and activity diversification - Increased LA motivation and enhanced student monitoring - Initial time investment required

4 CONCLUSION

This study demonstrates the systematic integration of LA into teaching practice through a structured toolkit approach. Evaluation validates the toolkit's effectiveness, with educators successfully utilising the toolkit for evidence-based teaching adjustments. While feedback indicates a need to address the learning curve, future development could explore on interface optimisation and interactive support. This work contributes to the LA field by providing a replicable framework for implementing data-informed teaching practices in higher education, effectively bridging the gap between analytics actionability and learning design integration.

REFERENCES

Bakharia, A., Corrin, L., De Barba, P., Kennedy, G., Gaševíc, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. *ACM International Conference Proceeding Series*, *25-29-April-2016*, 329–338. https://doi.org/10.1145/2883851.2883944

Blumenstein, M. (2020). Synergies of learning analytics and learning design: A systematic review of student outcomes. *Journal of Learning Analytics*, *7*(3), 13–32. https://doi.org/10.18608/JLA.2020.73.3

Campbell, J. P., & Oblinger, D. G. (2007). Academic Analytics. In *EDUCAUSE Publications* (Issue October). https://doi.org/10.4018/978-1-4666-5202-6.ch004

Hernández-Leo, D., Martinez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, *50*(1), 139–152. https://doi.org/10.1111/bjet.12645

Khalil, M; Ebner, M. (2015). Learning Analytics: Principles and Constraints. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, *JUNE*, 1326–1336
. https://doi.org/10.13140/RG.2.1.1733.2083

Liu, R., & Koedinger, K. R. (2017). Closing the loop: Automated data-driven cognitive model discoveries lead to improved instruction and learning gains. *Journal of Educational Data Mining*, *9*(1), 25–41.

Macfadyen, L. P. (2022). Institutional Implementation of Learning Analytics - Current State, Challenges, and Guiding Frameworks. In *Handbook of Learning Analytics* (2nd Edition, pp. 173–186). Society for Learning Analytics Research. https://doi.org/10.18608/hla22.017

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333–341. https://doi.org/10.1016/j.chb.2016.02.074

On the Design and Evaluation of an Interactive Study Planning Tool

Johannes Wagner-Schirrmeister [a], René Röpke [b], Tobias Johnen [c], Rike Carpantier [a], Ulrik Schroeder [c], Maren Scheffel [a]

[a] Ruhr University Bochum
{johannes.wagner-schirrmeister, rike.carpantier, maren.scheffel}@rub.de

[b] TU Wien
rene.roepke@tuwien.ac.at

[c] RWTH Aachen University
{johnen, schroeder}@cs.rwth-aachen.de

ABSTRACT: This poster[1] outlines the formative development and evaluation process of an interactive study planning tool that combines AI-based feedback and process mining to improve comprehensive, informed, and autonomous long-term study planning for students in higher education. The tool integrates rule-based AI with Learning Analytics to provide context-aware feedback, leveraging historical student data to inform planning decisions. The formative Human-Centered Design process integrates Process Mining and AI while incorporating interdisciplinary perspectives of computer science, pedagogy, social sciences, and ethics.

Keywords: Study Planning, Feedback, Artificial Intelligence, Process Mining, Evaluation

1 CONTEXT

Study planning in higher education is a complex challenge, particularly when students deviate from recommended plans. Existing resources, such as exam regulations and module handbooks offer limited insights into successful study pathways due to their static nature. This poster elaborates on an innovative interactive study planning tool that integrates Artificial Intelligence (AI) based feedback and data-driven insights from process mining on curriculum data. The challenges of study planning are multifaceted. Schulte et al. (2017) highlight the complexity of study programs and the need for personalized guidance, while different research projects explore prospects for interactive study planning tools for students (e.g., Hirmer et al., 2022; Judel et al., 2023; Weber et al., 2022). This work addresses gaps in existing study planning tools by combining technological, didactical, pedagogical, and ethical considerations to support students in higher education comprehensively. The goal is to provide a more effective and supportive solution for students, enabling informed, autonomous, context- and cohort-aware study planning. Our research[2] and development aim to explore the integra-

---

[1] After the conference, the presented poster (including screenshots of the tool) will be available here: doi.org/10.13154/294-12135. A related demo will be also presented (see the following chapter).

[2] Acknowledgment: This work was funded by the German Federal Ministry of Education and Research.

tion of Learning Analytics (LA), Process Mining (PM), and rule-based AI to develop a comprehensive study planning support system for students.

2 DEVELOPMENT AND EVALUATION PROCESS

Initially, a comprehensive pre-study was conducted to collect requirements from prospective users. A survey (n=674), as part of said pre-study, revealed that students (78%) generally follow recommended study plans while existing digital tools are mostly used as information sources (82%). However, only 12% of respondents use tools that facilitate personalized planning based on prior experiences, consistent with findings by Judel et al. (2023). Guided by these findings and common design principles for study planning applications (Hirmer et al., 2022), we designed an interactive study planning tool. The core component is a study planning timeline spanning the entire study life cycle, featuring an intuitive, semester-based interface with different forms of feedback and detailed information. A rule-based AI component translates program rules into formal notation, enabling immediate, context-aware feedback. Additionally, a process-mining component provides data-driven insights using historical student data on demographics and academic performance to support planning.

The iterative Human-Centered Design process based on the EFLA framework (Scheffel et al., 2017) examines *Usability*, *Acceptance*, *Ethics*, *Privacy and Data Protection*, *Pedagogy*, and *Improvement Potential*. Iterative user tests, surveys, and interviews with students and stakeholders ensure comprehensive evaluations and impact assessments, enriched by interdisciplinary perspectives from computer science, educational technology, ethics, and pedagogy. This process addresses shortcomings highlighted by Zawacki-Richter et al. (2019), continuously integrating ethical considerations instead of deferring them. Addressing ethical considerations regarding student autonomy and long-term collective impacts, especially considering AI-generated recommendations, students should be able to control recommendation settings and access algorithm explanations for transparency. Consultations with students as stakeholders, in user tests and interviews, with researchers and university administration inform ongoing development and evaluation of our tool, addressing autonomy, accessibility, and inclusion across various demographics. In this way, they also assist in contextualizing our understanding of the user group.

3 DISCUSSION

Our development and research aim to contribute to the field of LA and study planning assistance by integrating AI-based feedback with PM insights to provide personalized and dynamic feedback and support for individual study planning. We attempt to address existing gaps by developing and evaluating components for complex, context-aware feedback and recommendations that move beyond static recommended study plans. In the process itself, a framework for evaluating AI and LA implementations in educational technology, considering ethical and pedagogical implications has been devised and revised, which we will outline in future work along with corresponding results.

Integrating PM and LA for study planning poses several challenges, including ensuring high-quality data across diverse educational settings (within and between universities) and adapting the tool to various institutional structures, owing to the fragmented structure of the German higher education system, which may affect its generalizability. Balancing guidance with student autonomy is essential to avoid hindering self-regulated learning processes and competences while privacy concerns require strong data protection measures. Challenges may also arise in the course of technical system integration and widespread institutional adoption. Addressing these issues is essential for the tool's long-term success and scalability.

Future work will have to include conducting and presenting comprehensive evaluation studies to assess the short- and medium-term impacts of the tool on collective student planning behavior, student autonomy, as well as academic performance, and satisfaction. Development work will be done on the refinement of AI and LA components based on evaluations and best practices in educational data mining. Additionally, potential applications in curriculum design and institutional decision-making will be explored, leveraging insights gained from student planning data as well as historical student performance and experience to inform study program improvements.

4 CONCLUSION

The interactive study planning tool as well as the formative development and evaluation process outlined represent a novel and comprehensive approach to supporting students in higher education regarding study planning. By combining AI-based feedback, PM, and LA, we aim to empower students to make informed, more autonomous decisions about their academic paths. Our ongoing development and research align closely with emerging discussions within the LAK community by addressing the integration of AI in education, learning design, and complex data-driven decision-making. It aims to contribute to broader discourse on the ethical use of AI and LA in supporting student success and satisfaction in higher education. Moving forward, we invite collaboration and feedback from the LA community to further develop and reflect on this approach to study planning support.

REFERENCES

Hirmer, T., Etschmann, J., & Henrich, A. (2022). Requirements and Prototypical Implementation of a Study Planning Assistant in CS Programs. In *2022 International Symposium on Educational Technology (ISET)* (pp. 281–285). IEEE. https://doi.org/10.1109/ISET55194.2022.00066

Judel, S., Roepke, R., Azendorf, M., & Schroeder, U. (2023). Supporting individualized study paths using an interactive study planning tool. In *21. Fachtagung Bildungstechnologien (DELFI)* (pp. 225–230). Gesellschaft für Informatik e.V. https://doi.org/10.18420/delfi2023-36

Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., & Specht, M. (2017). The Proof of the Pudding: Examining Validity and Reliability of the Evaluation Framework for Learning Analytics. In É. Lavoué et al. (Eds.), *Data Driven Approaches in Digital Education* (pp. 194–208). Springer International Publishing. https://doi.org/10.1007/978-3-319-66610-5_15

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. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 538–539). ACM. https://doi.org/10.1145/3027385.3029446

Weber, F., Schrumpf, J., Dettmer, N., & Thelen, T. (2022). A 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.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int J Educ Technol High Educ, *16*(1). https://doi.org/10.1186/s41239-019-0171-0

Is Digital Game-Based Learning Effective for Enhancing Intrinsic Motivation in Mathematics Education?

Jeewon Choi [1], Young Hoan Cho [2]
Seoul National University
[1] woofwoof99@snu.ac.kr , [2] yhcho95@snu.ac.kr

ABSTRACT: Digital games have been proven to be effective for improving students' intrinsic motivation, but examination on whether this motivation is sustained after students stop playing the games is limited. This study investigates the effects of digital game-based learning (DGBL) on intrinsic motivation for mathematics with 73 third-year middle school male students, aged 14 to 15, in South Korea. Following a design-based research approach, the DGBL intervention was implemented over a period of three sessions. The results indicate a significant enhancement in intrinsic motivation for mathematics, suggesting that the motivation initially stimulated by the game continued to increase even after gameplay ceased. Moreover, students who engaged with the game through specific goal-setting and strategically used in-game tools experienced greater increases in intrinsic motivation. These findings imply that instructional support can effectively guide students in setting well-defined goals and fostering metacognitive awareness, helping them use game features appropriately based on their prior knowledge and task difficulty.

Keywords: Digital game-based learning, Mathematics education, Intrinsic motivation, Design-based research

1 INTRODUCTION

How can we prevent students from giving up on mathematics? The increasing number of students giving up on mathematics leads to learning gaps, which negatively impacts educational equality. Digital games have emerged as an effective tool for increasing students' interest in mathematics. However, a significant challenge remains. The interest generated through games often fails to sustain over time. Previous study has shown that gamification effectively strengthens learners' intrinsic motivation (Hanus & Fox, 2015); however, it has limitations in verifying whether this motivation persists after gameplay ends. Therefore, this study aims to examine whether the motivation initially stimulated continues even after the gameplay ceases as well as the impact of the digital game-based learning (DGBL) on intrinsic motivation for mathematics. To investigate the potential of DGBL in fostering and retaining intrinsic motivation in mathematics, this study addresses the following research questions: (1) How does DGBL impact intrinsic motivation in mathematics? (2) What activities play an important role in improving intrinsic motivation in DGBL?

2 METHOD

This study followed the design-based research (DBR) methodology (Wang & Hannafin, 2005) to develop and evaluate DGBL in math aimed at enhancing and maintaining intrinsic motivation in mathematics. The game was developed by researchers according to Self-determination Theory (Deci & Ryan, 2000), including level-based courses to satisfy competence, setting their own goals to fulfill autonomy, and interacting with NPC to address relatedness (See *Figure 1)*.

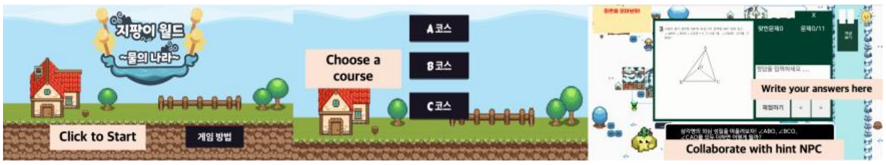


Figure 1: *Wand World* Gameplay Interface

We used a game called *Wand World* over 7 sessions, 45 minutes each, with 73 third-year all-boys middle school students in South Korea (See *Figure 2*). Of the 7 sessions, sessions 2–4 included gameplay using *Wand World*, while the others were solely for data collection purposes without additional mathematical content. Students' mathematics intrinsic motivation was measured using the intrinsic motivation inventory (IMI) questionnaires with 7 Likert scales (Ryan, 2006) - pretest, posttest, and delayed-posttest.

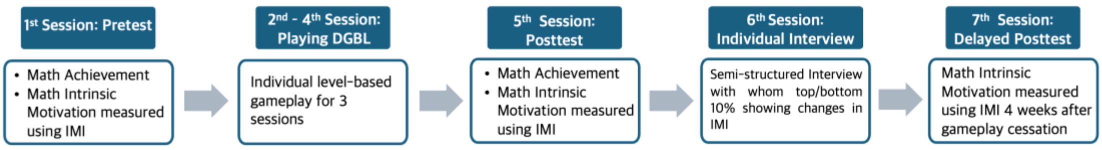


Figure 2: Procedures

Repeated measures ANOVA were conducted to examine the changes in intrinsic motivation scores through pretest, posttest, and delayed posttest. Furthermore, interview responses were analyzed using thematic analysis (Braun & Clarke, 2006) to explore the differences in learning activities between groups with high and low intrinsic motivation improvement. Initial coding was conducted by clustering data into higher-level themes through interactive coding.

3 RESULTS

3.1 The Effects of DGBL on Intrinsic Motivation in Mathematics

The descriptive statistical analysis revealed the intrinsic motivation in mathematics increased overall across the pretest, posttest, and delayed- posttest. Furthermore, repeated measures ANOVA revealed that these differences were statistically significant (F (2, 70) = 10.57, p < .001).

Table 1: Repeated measures ANOVA Results for Intrinsic Motivation in Mathematics

Pretest M(SD)	Posttest M(SD)	Delayed Posttest M(SD)	F	*p*
3.43(.85)	3.65(.69)	3.79(.75)	11.2	< .001

3.2 Differences in learning activities with intrinsic motivation improvement

As a result of the thematic analysis, gameplay style was a key factor in enhancing intrinsic motivation for mathematics, categorized into 'goal setting' and 'strategy'. In terms of goal setting, when students used the self-directed goal setting function, those who set specific goals, especially on the number of problems to solve- “I’d like to solve more than 5 problems” or “I will make sure to get 100 points”, showed greater motivation improvement. Conversely, students who didn’t show motivation improvement did not set goals or set them in an insincere manner. For example, they set trivial goals like “I’ll try more than 0 problems” or wrote nonsensical sentences by randomly typing without forming coherent words or phrases.

Second, strategy involved how students utilized various in-game tools to approach and solve problems. For instance, students who increased motivation were more likely to seek assistance from NPCs for hints or review relevant mathematical concepts provided within the game. They strategically utilize them based on factors such as task difficulty or their prior knowledge. In contrast, the students who didn’t show improvement relyed on hints or randomly guessing answers. Rather than strategically using in-game tools, they showed a tendency to show meaningless usage patterns of tools.

4 DISCUSSION

The results show that DGBL can significantly cultivate and sustain students' intrinsic motivation for mathematics. While DGBL was previously thought to mainly enhance extrinsic motivation, the findings proved its positive effect on strengthening intrinsic motivation as well (Ke, 2008). To maximize this effect, instructional support is crucial to foster intrinsic motivation. Teachers should guide students in setting specific goals or utilizing in-game tools to solve problems. It is essential to promote metacognitive support, enabling learners to effectively utilize game features based on their own abilities to solve problems appropriately. These insights highlight the importance of instructionally supported DGBL in fostering sustained intrinsic motivation.

REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Hanus, M. D., & Jesse, Fox. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & education, 80*, 152-161. https://doi.org/10.1016/j.compedu.2014.08.019

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay?. Computers & education, 51(4), 1609-1620. https://doi.org/10.1016/j.compedu.2008.03.003.

Ryan, R.M. (2006). Intrinsic Motivation Inventory (IMI). http://www.psych.rochester.edu/SDT/measures/intrins.html.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development, 53*(4), 5-23. http://doi.org/10.1007/BF02504682.

Context-Aware Synthetic Data Generation for Learning Analytics using Generative AI

Bineeth Kuriakose, Anna Mavroudi, Andres Moya, Jan Dolonen, and Crina Damşa
Department of Education, University of Oslo, Norway
{bineeth.kuriakose, anna.mavroudi, a.a.a.moya, j.a.dolonen, crina.damsa}@iped.uio.no

ABSTRACT: The integration of artificial intelligence in educational research has opened new avenues for addressing challenges related to data management and privacy. Building on this potential, our research tackles data scarcity and privacy issues by leveraging Generative Artificial Intelligence (GAI) to create context-aware synthetic data. Specifically, we propose a framework that employs Generative Adversarial Networks (GANs) to mimic student learning activities using data from digital learning platforms. The proposed framework also incorporates key contextual factors—such as demographic and academic background—to ensure the generated data reflects the diversity and variability of real-world datasets. This approach demonstrates that synthetic data preserves the statistical and behavioral properties of original data, enabling its application in real-world educational contexts, and significantly contributing to privacy-compliant and resource-efficient learning analytics.

Keywords: generative artificial intelligence, synthetic data, learning analytics, data privacy.

1 INTRODUCTION

The growth of digital learning has generated valuable data that can enhance student outcomes and improve learning systems. However, researchers face constraints due to data scarcity and privacy regulations, such as GDPR, complicating the acquisition of comprehensive datasets needed for robust educational research. Additionally, using original student data raises ethical concerns regarding privacy and the potential exposure of sensitive information (Isak, 2020).

To address these issues, this research leverages Generative Artificial Intelligence (GAI) (Lixiang et al., 2024) to produce synthetic data that preserves the statistical and behavioral characteristics of original student data. Specifically, we utilize Generative Adversarial Networks (GANs) (Divya and Jiannong, 2021). This approach incorporates detailed contextual information from the input seed data, making the synthetic data more realistic and comparable to the original data.

While previous research (Qinyi et al., 2024), have made strides in synthetic data generation, they often fall short in capturing nuanced contextual properties of input student datasets. Our approach addresses this limitation by preserving these detailed contextual characteristics, leading to more reliable and institution-specific synthetic student data. Hence the proposed method reduces reliance on original student data, mitigates privacy risks, and simplifies the data collection process. In this research, we analyze data both from the same cohort of students and across different populations over five years. This dual approach ensures a robust evaluation of our synthetic data generation framework. Through this approach, we aim to enhance the quality and impact of educational research while ensuring compliance with privacy standards and addressing ethical considerations.

2 METHODOLOGY

This research employs a systematic approach to generate synthetic student activity data using GANs. The process encompasses data collection and preprocessing, model training and data generation, and concludes with the evaluation and validation of the quality of the generated synthetic data. The process flow of the proposed framework is illustrated in Figure 1.

In the first phase of this research, we specifically focus on a bachelor course across one semester, sourced from a Learning Management System (LMS) platform and the Student Information System (SIS) to create realistic synthetic data replicating historical student activities. This example case is depicted in Figure 2, and an input dataset excerpt from the LMS platform is presented in Table 1.

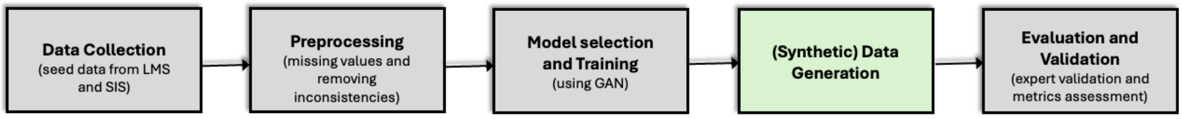


Figure 1: Process flow for proposed framework.

2.1 Data Collection and Preprocessing:

Data is collected from a single semester (Autumn 2023) for the bachelor course, including detailed student activity data from the LMS and aggregated exam results with demographic data from SIS covering the years 2019 to 2023. The preprocessing stage involves addressing missing values, normalizing data formats, and resolving inconsistencies to facilitate effective model training. These input variables are included to ensure that the synthetic data realistically reflects the diversity and variability of the original dataset.

2.2 Model Training and Data Generation:

A Generative Adversarial Network (GAN) is employed to generate synthetic data that mimics patterns in the original dataset, referred to as seed data. The GAN architecture consists of a generator, which creates synthetic data, and a discriminator, which assesses its authenticity. By leveraging contextual information from the collected data, the GAN iteratively refines its outputs to match features of original student activities across assignments, discussions, and exams. This method ensures the preservation of statistical properties and captures the nuanced behaviors present in the input dataset. During the generation of synthetic data for previous cohorts, we ensure to account for and normalize differences in curriculum design, teaching methods, and cohort-specific behaviors to maintain validity.

2.3 Evaluation and Validation:

The quality of the synthetic data is evaluated using statistical measures such as distributional similarity and error metrics, along with validation by domain experts in education to ensure it realistically reflects student behaviors. This dual-validation approach ensures the reliability of the generated data while adhering to ethical standards and privacy regulations, showcasing its consistency and effectiveness under practical conditions.

Table 1: Sample dataset excerpt from LMS and SIS.

id	userid	courseid	assignment	first deadline	last deadline	no. of tries	no. of feedback
10021	1019	1110	task7	05-11T23:59:00+02:00	05-05T08:56:59.359944+02:00	1	1
10022	1034	1110	task1	02-02T23:59:00+01:00	02-10T19:46:13.660977+01:00	2	2
10023	1034	1110	task2	02-16T23:59:00+01:00	02-22T17:38:29.002045+01:00	2	2
10024	1034	1110	task3	03-02T23:59:00+01:00	03-10T21:02:42.403235+01:00	2	2
10025	1034	1110	task4	03-16T23:59:00+01:00	03-16T23:42:55.542435+01:00	1	1

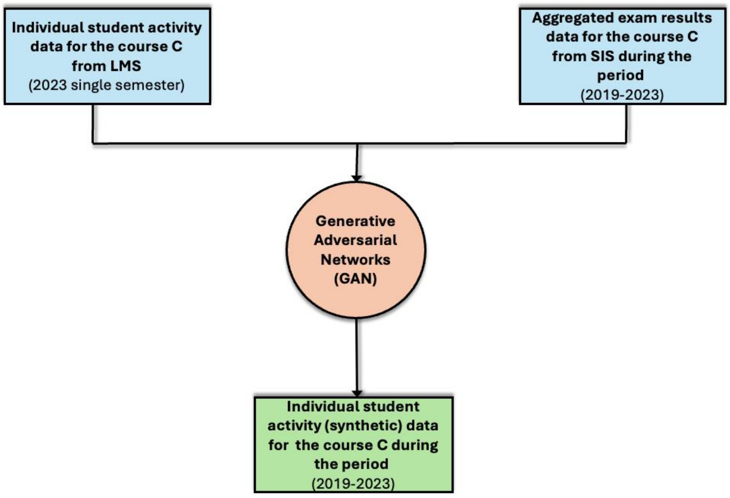


Figure 2: An example of the proposed conceptual framework.

3 DISCUSSION AND CONCLUSION

The proposed framework demonstrates the potential of GAI to address data scarcity and privacy concerns in educational research. By employing GANs to generate context-aware synthetic datasets, it is possible to replicate student learning data without compromising sensitive information. This approach addresses the challenges of large-scale data collection while providing valuable resources for developing models and systems in learning analytics. Our ongoing research prioritizes addressing ethical concerns and ensuring compliance with data privacy regulations to foster trust and acceptance of synthetic data practices. It lays the foundation for future studies that leverage the strengths of GAI to promote resource-efficient research, deliver innovative educational solutions, and improve learning outcomes.

REFERENCES

Qinyi Liu, Mohammad Khalil, Jelena Jovanovic, and Ronas Shakya. 2024. Scaling While Privacy Pre- serving: A Comprehensive Synthetic Tabular Data Generation and Evaluation in Learning Analytics. In Proceedings of the 14th Learning Analytics and Knowledge Conference. ACM New York, NY, USA, 620–631.

Isak Potgieter. 2020. Privacy concerns in educational data mining and learning analytics. The International Review of Information Ethics 28 (2020).

Divya Saxena and Jiannong Cao. 2021. Generative adversarial networks (GANs) challenges, solutions, and future directions. ACM Computing Surveys (CSUR) 54, 3 (2021), 1–42.

Lixiang Yan, Roberto Martinez-Maldonado, and Dragan Gasevic. 2024. Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In Proceedings of the 14th Learning Analytics and Knowledge Conference. 101–111.

Validity of Automated Analysis for Learner-AI Collaboration

Bookyung Shin, Hyeeun Kim, Yewon Han, Yeono Son, Anna Kim, Young Hoan Cho
Seoul National University
[shinbrooke, kim5039, ywhan, yeonoson, dkssk228, yhcho95]@snu.ac.kr

ABSTRACT: This study investigates the automated coding results of the Learner-AI Collaboration (LACo) assessment system, an automated tool for analyzing learners' competency for collaboration with AI. Using trace data and ChatGPT interactions from 31 university students' AI-enhanced writing tasks, we conducted automated analysis of learner-AI interaction patterns and prompt content. Validation comparing automated and manual coding showed high correlation for interaction analysis, and moderate inter-rater reliability in prompt analysis. This research contributes to the growing need for process-based, scalable assessment tools in AI-enhanced learning environments.

Keywords: learner-AI collaboration, competency assessment, automated assessment system

1 INTRODUCTION

With the advent of generative artificial intelligence (AI) services, the ability to collaborate effectively with AI agents has become a crucial competency. Human-AI collaboration competency entails working with AI towards a common goal, through continuous interactions that may involve aspects such as AI-oriented communication and task regulation (Song & Cho, 2023). To help learners enhance their human-AI collaboration competencies, it is essential to efficiently detect their current states via process data and thus assess potential areas for improvement.

Automated assessment systems can provide a scalable approach to analyzing learner-AI collaboration competencies. Trace data can be collected from learners' interactions with AI in online environments, then aggregated as action libraries to automatically track meaningful interactions during learning processes (Cheng et al., 2024). Additional content analysis can also be streamlined using generative AI, which has been shown to efficiently code discourse data (Garg et al., 2024). Building upon these possibilities, this study aims to explore automated coding results as a first step towards developing and validating the Learner-AI Collaboration (LACo) assessment system, an automated system that analyzes learner-AI collaboration using trace data and generative AI-based content analysis. [1]

2 METHODS

2.1 Data Collection and Analysis

Data was collected from 31 South Korean university students (19 female, age M=23.58), who used ChatGPT while performing two writing tasks (one simple, one complex). Trace data was acquired in

[1] This research was funded by the Korean Ministry of Education and the Korean Research Foundation (NRF-2022S1A5A2A01045587).

the form of event logs (i.e. clicks, keystrokes, etc.) from writing tasks and ChatGPT conversation logs exported from ChatGPT. Participants' screens were recorded during the writing tasks.

The LACo assessment system's analysis was divided into two primary dimensions: analysis of learners' actions when interacting with ChatGPT, and analysis of the contents of their prompts to ChatGPT. Coding schemes to analyze these actions and content were developed based on computer-supported collaborative learning and human-AI collaboration literature. Analysis of interactive actions (i.e. *interaction analysis*) was conducted with a coding scheme including the categories 'metacognition' (subcategories 'planning', 'human monitoring', and 'AI monitoring'), 'task' (subcategories 'writing' and 'revised writing'), and 'interaction' (subcategories 'prompting', 'AI response', and 'copy'). Automated coding was conducted with actions extracted from the combined trace data (event logs and ChatGPT conversation logs), organized by participant, task, and timestamp, which was then processed through a Python code that applied the interaction coding scheme. In addition, manual coding using participants' screen recordings was performed separately by two researchers with Atlas.ti software. Inter-rater reliability was high at 0.94 (Cohen's kappa).

Content analysis of learners' prompts to ChatGPT (i.e. *prompt analysis*) utilized a coding scheme with the categories 'content' (subcategories 'information seeking', 'metacognition-planning', 'metacognition-evaluation', 'request to perform task', 'negotiation', and 'miscellaneous'), 'type' (subcategories 'new', 'repeated', and 'additional'), 'elaboration', and 'socioemotional expression'. The categories 'elaboration' and 'socioemotional expression' were each coded into four levels: 'very low', 'low', 'high', and 'very high'. Automated prompt analysis was performed using ChatGPT's 4o model. Instructions given to ChatGPT to apply the coding to an uploaded excel file included an explanation of the coding scheme based on strategies such as Few-Shot-Chain-of-Thought. Furthermore, two researchers performed manual coding of the prompts (Cohen's kappa = 0.93 for binary codes, 87.18% accuracy for 'elaboration' and 'socioemotional expression').

2.2 Validity Analysis

To validate the LACo assessment system's automated analysis, we compared manual coding performed by the research team with the automated coding results, using correlation and inter-rater reliability analysis. Pearson's correlation was calculated in terms of frequency and summed duration to compare manual and automated coding of interaction analysis. Inter-rater reliability with Cohen's kappa was calculated for the coding results of the content analysis. If correlation coefficients are both significant and high, and if kappa > 0.65 (Swiecki et al., 2020), we can surmise that the LACo assessment system is performing similarly to the human coders and thus providing valid results.

3 FINDINGS

3.1 Validity of Interaction Analysis in the LACo Assessment System

Correlation analysis of manual and automated coding yielded significant results ($p<0.05$) and high correlation coefficients (Figure 1). Coefficients of correlations ranged from 0.356 to 0.996, with almost all p values below 0.001. Only coefficients for 'AI monitoring' (r=0.356, p=0.0495) or 'writing' (r=0.449, p=0.011) frequencies were relatively low. Coefficients for summed duration were lower in 'AI monitoring' (r=0.564, p=0.001) and 'human monitoring' (r=0.417, p=0.020).

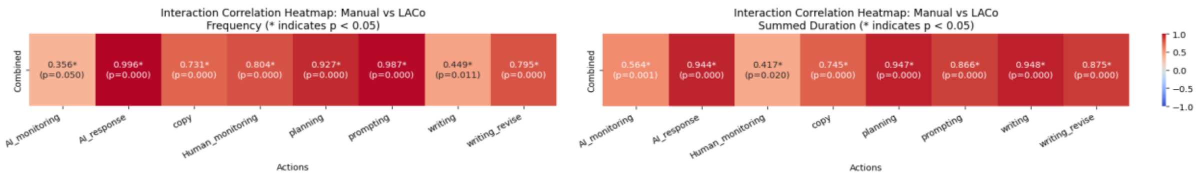


Figure 1: Correlations between manual and automated coding for interactions

3.2 Validity of Prompt Analysis in the LACo Assessment System

Inter-rater reliability for the binary codes when comparing manual and automated prompt analysis was substantial, nearing the threshold of reliability with Cohen's kappa at 0.646. For assessment of 'elaboration' and 'socioemotional expression', ChatGPT provided the same scores as human raters for 47.78% and 15.40% of the prompts, respectively. ChatGPT tended to overestimate the level of elaboration, and to underestimate the level of socioemotional expression compared to human raters.

4 DISCUSSION AND CONCLUSION

Our preliminary results show promising potential for the automated coding of interactions and prompts, which can be used for further development of the LACo assessment system. Automated assessment systems can decrease costs incurred by manual coding (Garg et al., 2024), making assessment applicable on larger scales. They can also aggregate data in real time to provide teachers with meaningful action indices of learner-AI collaboration competencies. For instance, copy-paste events, easily tracked with trace data via automated coding, could be used to present indices of AI-dependency (or 'knowledge-telling'; Cheng et al., 2024). Real-time presentation of assessment is particularly important for learning AI collaboration competencies. Collaboration is a temporal process (Swiecki et al., 2020) which targeted intervention such as scaffolding by teachers can enhance; in addition, presenting results to learners can lead to repeated reflections, which improves competency (Song & Cho, 2023). Future research on the LACo assessment system can develop process-based action indices that combine interaction and content analyses to quickly detect difficulties and improve competencies while numerous learners participate in AI-enhanced learning environments.

REFERENCES

Cheng, Y., Lyons, K., Chen, G., Gašević, D., & Swiecki, Z. (2024, March). Evidence-centered Assessment for Writing with Generative AI. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 178-188).

Garg, R., Han, J., Cheng, Y., Fang, Z., & Swiecki, Z. (2024, March). Automated Discourse Analysis via Generative Artificial Intelligence. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 814-820).

Song, H., & Cho., Y. H. (2023). A developmental study on design principles of activity-based instruction for improving Human-AI collaboration competency. *Journal of Korean Association for Educational Information and Media*, *29*(1), 145-173.

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to collaborative problem solving: a network analysis approach. *Computers in Human Behavior*, *104*, 105876.

Network Analysis of Solution Processes in Math Online Tests

Asahi Kurihara
Graduate School of Informatics, Nagoya University

Yasuyuki Nakamura
Institute of Liberal Arts and Sciences, Nagoya University
nakamura.yasuyuki.f1@f.mail.nagoya-u.ac.jp

ABSTRACT: This study visualized student answer data from the lecture *Introduction to Mathematics*, conducted using the automatic equation scoring system STACK, as a network diagram and analyzed it by calculating network features. Students can resubmit answers multiple times after an incorrect response, allowing data collection on their answering process. Based on this data, solution transitions were represented in directed graphs, and network features were calculated to quantitatively evaluate the effect of feedback on answer processes and nodes corresponding to wrong answers. Network features also facilitated comparisons between questions, revealing patterns in students' answering tendencies and feedback effects. Combining these indicators with traditional metrics, such as percentage of correct answers and number of attempts, provides a clearer understanding of class learning situations, supporting more tailored educational opportunities.

Keywords: Math online test, Network analysis, Question-solving

1 INTRODUCTION

There are several automated mathematics marking systems, including STACK, Möbius, and WeBWorK. In this study, answer data were analyzed using STACK, a system widely used in Europe and increasingly adopted in Japan. STACK is an online automated scoring system where students input mathematical expressions, which are scored for correctness. Teachers can set potential answers based on the required knowledge elements, enabling effective classification of responses. In an earlier work, incorrect answer patterns were categorized using e-learning log data, while Nakamura et al. developed a classification method using a potential response tree (PRT) (Nakamura et al., 2021). STACK provides immediate feedback on whether a response is correct or incorrect. If incorrect, students can retry the question until they succeed, enabling analysis of their answering processes and knowledge acquisition. While studies on classifying incorrect answers within STACK's PRT exist, analyzing solution processes could offer deeper insights into knowledge acquisition and help improve learning environments. Additionally, inadequate answer classification might reflect flaws in question design, but methods to identify such issues remain unclear.

This study aims to evaluate question quality by visualizing students' answering processes as directed graphs, using STACK data, and calculating network features. Previous educational studies using network features, such as Yasutake et al. (2011), analyzed e-learning log data and interaction networks among learners. These studies highlighted differences in network characteristics based on teaching methods, underlying factors, and their links to learners' knowledge levels.

2 ANALYSIS METHOD

2.1 Visualization of solution processes using directed graphs

Based on the results of incorrect or partially correct answers classified by STACK's PRT, a weighted directed graph was used to visualize answer trends. This was achieved by representing the types of incorrect or partially correct answers that led to a correct answer in a directed graph, which was then aggregated for an entire class. The data analyzed were the responses of approximately 100 participants in the introductory course *Introduction to Mathematics* (covering differential and integral calculus) offered at Nagoya University in 2021.

Figure 1 (left) illustrates the answer transition of a single student, showing the progression from an unclassified wrong answer (node 0) through a classified wrong answer (node 1) to the correct answer (node c) in a directed graph. Figure 1 (right) shows a weighted directed graph that aggregates the answer transitions for 20 students. The arrows indicate the direction of the answer transitions, and the thickness of the edges corresponds to the frequency of transitions between nodes.

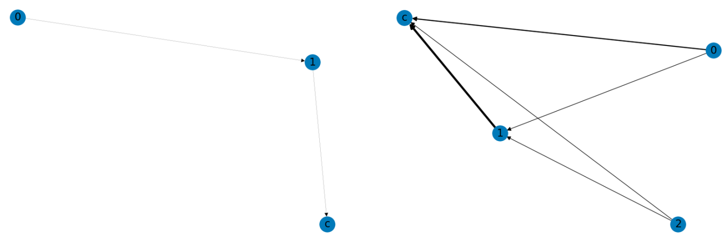


Figure 1: An example of an answer transition of one student (left) and a weighted directed graph superimposed for 20 students (right).

2.2 Feature analysis of answering processes using network features

For each of the constructed networks, various network features were calculated to quantitatively evaluate the solution processes. These features were visualized using numerical values and histograms. The analysis focused on identifying which types of incorrect responses were more likely to lead to correct responses. The following network features were analyzed: **Degree Centrality**: Nodes with high degree centrality represent points that are easier for students to transition through. **In-Degree Centrality**: Nodes with high in-degree centrality gather many transitions toward them, indicating common errors. **Out-Degree Centrality**: Nodes with high out-degree centrality have numerous outgoing edges, reflecting points from which students move onward to other responses.

3 RESULTS AND DISCUSSIONS

As an example, four questions (Q4-1 to Q4-4) were analyzed, and the relationships between the directed graphs and network features are summarized in Figure 2. The integral questions and their corresponding solution processes are as follows: 4-1. $\int (x-3)^7(2x+1)\,dx = \frac{(16x+15)(x-3)^8}{72} +$

C, 4-2. $\int (x-3)^7(2x+1)\,dx = \frac{(16x+15)(x-3)^8}{72} + C$, 4-3. $\int \frac{x}{\sqrt{2x+1}}\,dx = \frac{(2x+1)^{3/2}}{6} - \frac{\sqrt{2x+1}}{2} + C$, 4-4. $\int (x-1)^5 x^2\,dx = \frac{(x-1)^6(21x^2+6x+1)}{168} + C$. For question Q4-2, the node representing unclassified wrong answers (node 0) had a high in-degree centrality, indicating a large number of unclassifiable incorrect responses. This suggests that the PRT needs improvement for more accurate classification. Conversely, the low in-degree and out-degree centralities of nodes 2 and 3 imply that students quickly moved past these nodes after receiving effective feedback, leading to correct answers. For Q4-3, the relatively high degree centrality across all nodes suggests that the PRT effectively captured a variety of errors. This indicates that the feedback system was well-designed, allowing students to transition efficiently toward correct solutions.

The use of directed graphs provided insights into students' knowledge acquisition processes. Network features calculated for each question highlighted trends in student responses and the effectiveness of feedback provided by the PRT. By analyzing directed graphs and their associated network features, it is possible to identify the role of each node in the PRT and assess whether the feedback mechanisms at specific nodes are functioning effectively.

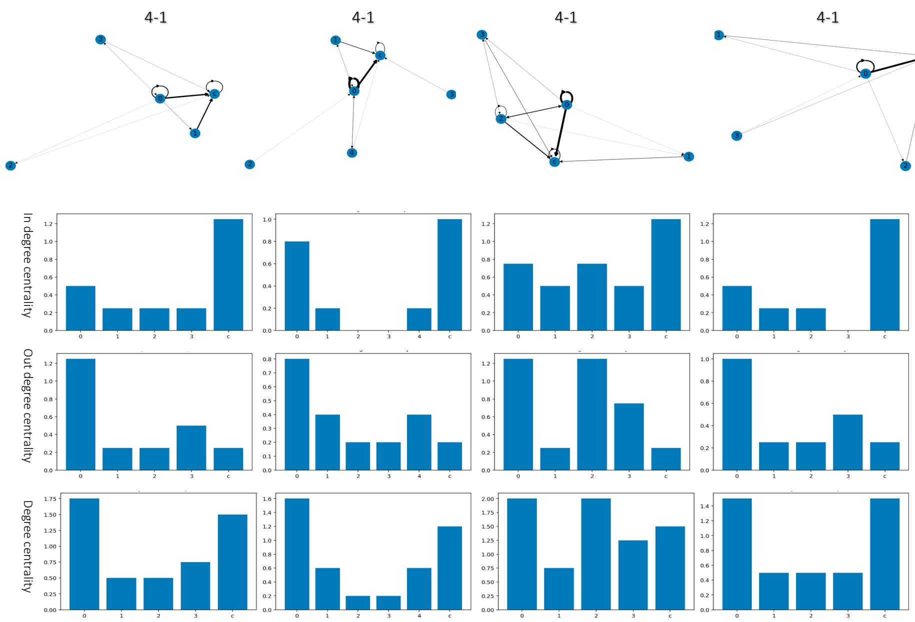


Figure 2: Directed graphs for the four questions and the corresponding network features for each.

REFERENCES

Nakamura, Y., Higuchi, S., Yoshitomi, K., Miyazaki, Y., Ichikawa, Y., Nakahara, T. (2021). *Automatic classification of incorrect answers to differentiation questions using Potential Response Tree*, Proceedings of International Meeting of the STACK Community 2021, 8 pages. http://dx.doi.org/10.5281/zenodo.4915994

Yasutake, K., Yamakawa, O., Tagawa, T., Sumiya, T., Inoue, H. (2011). A Simulation Analysis on Learning Effects through Interactions in Networked Communities, Transactions of Japanese Society for Information and Systems in Education, https://doi.org/10.14926/jsise.28.50

Dropout Analytics using Undergraduate course and program-level Digital Trace data – A Privacy-centered Approach

Andrés Araos Moya, Bineeth Kuriakose, Crina Damşa
Department of Education, University of Oslo, Norway
a.a.a.moya@iped.uio.no, crina.damsa@iped.uio.no, bineeth.kuriakose@iped.uio.no

Rachelle Esterhazy
Oslo Metropolitan University, Norway
rachelle.esterhazy@oslomet.no

ABSTRACT: First-year students' (FYS) dropout remains a pressing issue for higher education due to its negative implications for both students and institutions. Leveraging data from digital platforms presents new opportunities to understand and mitigate this problem. This contribution outlines a study aimed at collecting and analyzing aggregated data at the program, cohort, and course levels for studying FYS dropout, while ensuring the protection of students' data privacy. The study involves collecting data from every cohort and program at one major university in Norway from the period 2019-2023 and the corresponding courses, sourced from a learning management system and the university's digital system. The study's focus is on discussing the relevance of indicators, which includes using generative AI to extract features from text, and analytical approaches. The study expects to contribute with methodological advancements in studying FYS dropout, which can in turn set the basis for the development of openly available dashboards for administrative and teaching staff, and leadership.

Keywords: student dropout, learning analytics, first-year students, digital traces

1 BACKGROUND

First-year students' (FYS) dropout, namely students leaving their studies before degree completion, has become a growing concern for higher education institutions (HEIs). This issue not only has consequences for students' future but also presents substantial pedagogical and administrative challenges for HEIs. With the introduction of digital technologies in HEIs, the field of learning analytics has tried to exploit the opportunities for investigating dropout arising from the large amounts of data about students' activity and program progression. Several studies have developed various dropout analytics models using statistics, machine learning and AI, by capitalizing on such data to predict students' dropouts (Bond et al., 2024). Still, most studies focus solely on data from individual students, while only a few pays attention to dropout rates at the program level or how course design affects dropout rates (e.g. Poellhuber et al., 2023). Program and course-level statistics are key in the work of both teaching and administrative staff, as they can be used to inform their decisions about curriculum development. Nonetheless, the standards for the protection of personal data introduced by the General Data Protection Regulation (GDPR) have translated into new challenges for the development of such statistics[1].

[1] See: Statistical confidentiality and personal data protection - Eurostat

The present study is part of a project at a large university in Norway. The project explores the potential of data harvested by various digital platforms at the university to understand FYS dropout and its relation to social and academic integration. This contribution has two main aims. First, to use aggregated student data and course design data to develop program and course-level FYS dropout indicators and insights that avoid any direct or indirect identification of individual students. And second, to use these indicators to differentiate between dropout patterns across programs and courses.

2 METHODOLOGY

We start by defining methodological considerations for both the data collection and indicator development processes. First, we defined the main unit of analysis, the program-cohort. This unit of aggregation refers to students enrolled in the same program in the same year/semester. Relating course-level data to a program-cohort adds an extra layer of complexity. Many programs use flexible curricular structures, making course enrollment to be heterogeneous in terms of both program and cohort. Second, we defined indicators that, neither directly nor indirectly by means of data triangulation, allow for individual students to be identified. For this purpose, we used indicators that avoid collecting or reflecting any exact cohort size, course enrollment, participation or interaction numbers.

The data originates from two main platforms/databases: the information system Fellestudentsystem (FS); and the university's learning management system (LMS). We collected data from every program-cohort at the university from the period 2019 to 2023. Using this data, we developed different indicators using both individual and aggregated student data for each program-cohort. Meanwhile, to characterize the courses' design, we developed four main indicators: type(s) of final examination, number of obligatory assignments, and types of learning outcomes (LOs). For the latter, we provided a codebook based on Bloom's taxonomy to an in-house generative AI model to automatically classify the courses' LOs (Chew et al., 2023). A cross-check of a random sample of the results will be performed to ensure quality. To generate insights, we consider semester registration to be the main operationalization of dropout and, based on it, contemplate two main analytical approaches. First, graphical analyses of each indicator for each program-cohort and across them. Second, Bayesian statistics analyses (van de Schoot et al., 2021), intended at providing general insights into the impact of each indicator over the probability of dropout.

3 PRELIMINARY RESULTS AND CONTRIBUTIONS

Preliminary results of this study include the successful design and extraction of indicators (see Table 1), and the successful piloting of the extraction of types of LOs using generative AI. We expect that this study's contribution will help in the study of FYS dropout in two main ways. First, by advancing the field of learning analytics with new methodological approaches focused on program and course-level data and indicators that also ensure students' data privacy in all stages of the methodological design. Second, by exploring how differences at both the program and course design level affect FYS dropout. We expect that the findings of this research will set the basis for the development of FYS dropout dashboards that are openly available for higher education leadership, administration, and teaching staff.

Table 1: Indicators for each program-cohort

Indicator	Source	Level	Description
Cohort size	FS	Program	Number of students enrolled in a program
Semester registration	FS	Program	Rate of students who registered for a semester
Sex distribution	FS	Program	Rate of female/male students
Age distribution	FS	Program	Rate of students younger than 22
Geographic distribution	FS	Program	Rate of students coming from the university's main region(s)
Course enrollment	FS	Course	Rate of students from a program-cohort enrolling in a course
Exam grade(s)	FS	Course	Rate of pass/fail
Assignment results distribution	LMS	Course	Rate of complete/incomplete assignments
Assignment submission distribution	LMS	Course	Rate of late and missing assignment submissions
Online discussions participation	LMS	Course	Weekly rate of students participating and participations per student
Online quizzes participations	LMS	Course	Weekly rate of students participating and participations per student
LMS pageviews	LMS	Course	Weekly rate of students viewing and rate of views per student

REFERENCES

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., ... & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21(1), 4.

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. arXiv preprint arXiv:2306.14924.

Poellhuber, L. V., Poellhuber, B., Desmarais, M., Leger, C., Roy, N., & Manh-Chien Vu, M. (2023). Cluster-based performance of student dropout prediction as a solution for large scale models in a Moodle LMS. In LAK23: 13th International Learning Analytics and Knowledge Conference (pp. 592-598).

van de Schoot, R., Depaoli, S., King, R. et al. Bayesian statistics and modelling. Nat Rev Methods Primers 1, 1 (2021). https://doi.org/10.1038/s43586-020-00001-2

Striking the Balance: Exploring Levels of AI Tutor Proactivity in Enhancing Online Self-Paced Learning

Eunseo Yang, Yeonji Jung*
Ewha Womans University, The University of Memphis*
y.eunseo@ewhain.net, yeonji.jung@memphis.edu

ABSTRACT: This study examined the proactive role of Artificial Intelligence (AI) tutors in students' online lecture experiences. To achieve this, we designed and developed an AI tutor with four distinct proactivity levels: Reactive Support, Notification, Suggestion, and Active support. The AI tutor was embedded into the online lecture platform, where students could access and chat with the AI tutor during their lectures. Experiments were conducted with 8 students, where each participant engaged in a 15-minute online lecture with the AI tutor at different levels of proactivity. Using interaction log data, quiz scores, and survey and interview responses, students' learning ownership, engagement, and outcome were measured and compared across groups with four proactivity levels for the AI tutor. Findings showed that the AI tutor's higher proactivity positively influenced students' learning outcomes, while leading to a decrease in students' sense of ownership and engagement. The findings highlight the importance of a balanced approach to AI tutor proactivity, where tailored, adaptive interventions can enhance learning outcomes without compromising learner autonomy and engagement in self-paced online environments.

Keywords: generative AI, conversational agents, AI tutor, proactiveness, online self-paced learning, personalized learning, learning ownership, learning engagement

1 Introduction

With the expansion of remote education systems, Artificial Intelligence (AI) tutors play an increasingly essential role in supporting effective, self-paced online learning. In typical educational settings, teachers often play a critical role in monitoring student progress and offering timely support to guide student focus. AI chatbots are expected to take a teacher role in online self-paced learning through on-the-spot interactions (e.g., asking questions to AI and receiving instant responses generated), helping students stay engaged and address challenges simultaneously (Baillifard et al., 2024). Particularly, the *proactive* aspect of AI tutors received attention with the potential of fulfilling learner needs and providing timely information, often without explicit user requests (Meurisch et al., 2020; Deng et al., 2024). However, in most practices of leveraging AI chatbots into online learning, AI tutors often operate in a *passive* and *reactive* manner, failing to fully replicate the dynamic, responsive teacher-student interactions found in traditional face-to-face classroom settings (Baillifard et al., 2024).

This study explores the potential of how AI tutors can function *proactively* in providing timely support for online self-paced learning, with a particular focus on determining the appropriate level of proactive intervention for effective support. A generative AI-based tutor was designed with four levels of proactivity based on the literature review (Reactive Support, Notification, Suggestion, and Active Support), and then tested with 8 students to examine how the AI tutor proactively intervenes in real online learning scenarios. Using interaction log data, quiz scores, and survey and interview responses, students' learning ownership, engagement, and outcome were measured and compared across groups, each of which had selected one of the four proactivity levels for the AI tutor. The findings contribute to the knowledge base for optimizing AI tutor proactivity to foster meaningful online self-paced learning experiences, laying the groundwork for further exploration and dialogue.

2 Design and Context for Proactive AI Tutor System

The proactive AI tutor was built using the GPT-3.5-turbo model and integrated into a web-based platform using HTML, CSS, JavaScript, and Flask API. Students were exposed to a 15-minute lecture on reinforcement learning, followed by a quiz, post-survey, and interview. The AI tutor took three key tasks at different points of learning: (1) generating educational materials before learning (e.g., learning objectives, quiz questions, and intervention messages); (2) responding to students' real-time prompt messages; (3) delivering the prepared materials from step 1 at predetermined times. Before a student started watching the lecture video, the AI tutor was used to create the educational materials and embed them into the learning platform (see Figure 1). The AI tutor extracted learning objectives from the lecture text (e.g., "*Explain the roles of agents and rewards in reinforcement learning*") and created O/X quiz questions, setting them up before learning began. Then, the AI tutor generated a set of appropriate intervention messages at specific points (e.g., *"At this point, understanding the roles of agents and rewards is crucial. If needed, I can create a simple quiz to check your understanding."*) During the 15-minute online lecture, the AI tutor intervened in the student's learning process by providing tailored prompts, according to each student's selected level of proactivity among four options, which were selected based on the literature (Deng et al., 2024; Oh et al., 2024; see Figure 1). At the **Reactive Support** level, the tutor only responded to student-initiated queries. The **Notification** level automatically informed students of key points during the lecture, while the **Suggestion** level offered optional supplementary materials on key topics. At the **Active Support** level, the tutor paused the lecture to present critical content, requiring student review before proceeding.

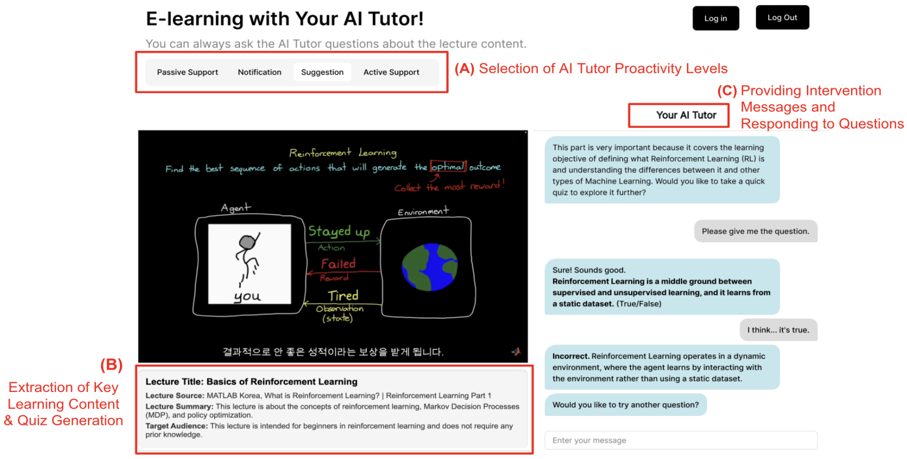


Figure 1: The proactive AI tutor embedded in a web-based self-paced learning platform

2 In-Situ Examination of the Role of AI Tutor Proactivity on Learning

The study involved eight graduate students (ages 24–34) enrolled in a Deep Learning course at a large private university in South Korea, all of whom lacked prior experience in reinforcement learning. Each participant engaged in a 30-minute intervention within an online self-paced learning environment using a between-subject design. After the session, we collected multiple data types: (1) interaction log data with the AI tutor (including the number of questions asked to measure *behavioral engagement*), (2) quiz scores on a 20-item test to measure *learning outcomes*, and (3) survey responses assessing *ownership* (the level of control and satisfaction a student feel over their learning while using AI tutor) and *cognitive engagement* (the extent to which a student maintains their focus on learning through constant interaction), with two items for each on a 7-point Likert scale, alongside interview responses to explore the rationale behind their survey ratings. We then analyzed differences in reported ownership, engagement, and learning outcomes (quiz scores) across groups of two students, each of which had selected one of the four proactivity levels for the AI tutor.

The results revealed varying impacts of AI tutor proactivity levels on student ownership, engagement, and learning outcomes. For *learning ownership*, the Reactive Support level received the highest score (M = 7), followed by Notification (M = 6), Suggestion (M = 3), and Active Support (M = 4.5). These results show that higher proactivity diminished students' perceived control over their learning. *Behavioral engagement*, measured by the number of student-initiated questions, was highest with Reactive Support (M = 5) and Notification (M = 4), while engagement dropped significantly with the Suggestion (M = 1.5) and Active Support levels (M = 1), indicating a decline in voluntary participation as proactivity increased. However, in *cognitive engagement*, measured by the extent to which they maintain their focus through continuing interaction, the different proactivity levels had minimal effect, indicating that proactive interventions did not disrupt the learning process. Finally, *learning outcomes*, measured by a 20-question quiz, were highest at the Active Support level (M = 18.75), followed by the Suggestion (M = 18.5), Notification (M = 16.5), and Reactive Support levels (M = 15.5), suggesting that increased proactivity showed higher learning outcomes.

3 Discussion

Our research explored the role of AI tutor proactiveness in shaping students' online self-paced learning experiences. Specifically, we designed the AI tutor to operate at different levels of proactivity and implemented it in real learning environments, conducting an in-situ examination of its meaningful affordance in practice. The findings reveal a dual role of AI tutor proactivity in online self-paced learning. While higher levels of proactivity might enhance learning outcomes and stimulate additional activities, they may inadvertently reduce learners' sense of ownership. This suggests that AI tutors should be carefully calibrated to balance proactive support with the need for learners to maintain control and autonomy over their learning experience. However, the findings also showed that groups with higher proactivity levels showed reduced perceptions of control over the learning process and lower behavioral engagement. The observed decrease in behavioral engagement among students with more proactive AI tutors might suggest that too much intervention may inhibit voluntary student participation. This decrease underscores the importance of designing AI tutor systems that encourage, rather than replace, active engagement. For effective online self-paced learning, AI tutor interventions might be adjusted to prompt student-initiated actions without overwhelming their learning process. These findings suggest several pathways for refining AI tutor design in self-paced learning contexts; integrating real-time monitoring and adaptive proactivity levels could enable AI tutors to dynamically adjust to learners' evolving needs, ensuring that support is neither overly intrusive nor overly passive (Meurisch et al., 2020; Deng et al., 2024). Further research might explore the development of AI tutor systems that promote both high engagement and autonomous learning by leveraging insights from these results. Together, this study highlights the importance of a balanced approach to AI tutor proactivity, where tailored, adaptive interventions can enhance learning outcomes without compromising learner autonomy and engagement in self-paced online environments.

REFERENCES

Baillifard, A., Gabella, M., Lavenex, P. B., & Martarelli, C. S. (2024). Effective learning with a personal AI tutor: A case study. *Education and Information Technologies*, 1-16.

Deng, Y., Liao, L., Zheng, Z., Yang, G. H., & Chua, T. S. (2024, July). Towards human-centered proactive conversational agents. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 807-818).

Meurisch, C., Mihale-Wilson, C. A., Hawlitschek, A., Giger, F., Müller, F., Hinz, O., & Mühlhäuser, M. (2020). Exploring user expectations of proactive AI systems. Proceedings of *the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(4), 1-22.

Oh, J., Kim, S., Lee, H., & Park, Y. (2024). Better to ask than assume: Proactive voice assistants' communication strategies that respect user agency in a smart home environment. *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

Modeling Class Intervention Impact on Absenteeism

Naruhiko Shiratori
Tokyo City University
nshirato@tcu.ac.jp

ABSTRACT: This study investigates the impact of attendance in a foundational intervention seminar on reducing absences in other courses the following week. Using weekly attendance data from 181 students and implementing a hierarchical Bayesian model with a Poisson distribution, we analyzed the effects of seminar attendance versus absence. Experimental results indicate that seminar attendance reduces absences in other classes by an average of 1.3 fewer absences in the early weeks, with the difference in absences most pronounced during Weeks 1 to 5. These findings underscore the effectiveness of early-semester attendance as an intervention strategy.

Keywords: Class Intervention Impact, Hierarchical Bayesian Modeling, Absenteeism

1 BACKGROUND AND PURPOSE OF THIS STUDY

While dropout rates at Japanese universities may appear low, they represent a serious issue for students, families, and universities. According to MEXT (2023), the dropout rate in 2024 is 2.10%, but this varies widely by university and faculty. For instance, dropout rates are higher at universities with lower academic rankings and can reach several tens of percent in some social science faculties. Studies have shown that first-semester GPA is closely linked to dropout risk, emphasizing the need for early interventions before final grades are available. Research by Ortiz Lozano et al. (2020) and Shiratori et al. (2020) suggests that while dropout prediction models are accurate after semester's end, the end of the semester may be too late for effective intervention, underscoring the importance of mid-semester actions.

However, assessing the effectiveness of such interventions is challenging, as it requires tracking students weekly. This study aims to evaluate intervention effectiveness using readily available attendance data. Specifically, we examine how attendance in a foundational seminar affects absences in other courses. This model enables weekly assessments of seminar attendance's impact on other class absences, allowing for targeted interventions during high-risk weeks and supporting improvements in early-semester attendance and educational outcomes.

2 DATA, VARIABLES, AND MODEL

The data used in this study consists of information on students (n=181) who entered the Faculty of Humanities at University A in Tokyo in 2018. The main variable is the number of weekly absences during the first semester of the freshman year. Classes are divided into two categories: the mandatory foundational seminar (seminar), taken by all students, and other courses (OT). The seminar is a course led by an advisor to support students' transition to university and facilitate early interventions. Other courses include language, ICT, and introductory specialized courses. We use data from all 15 weeks of the foundational seminar, with an average absence count of 1.72, a median of 1, and a standard deviation of 2.01. For other courses, the weekly absence counts across all classes are aggregated for each week from Week 1 to Week 15. Table 1 shows descriptive statistics (mean, median, and standard deviation) for weekly absences in these other courses. Notably, the mean absence count increases gradually from Week 2 to Week 14, exceeding 1.0 from Week 10 onward, indicating that students, on average, miss at least one OT class per week in the latter half of the semester.

Table 1: Descriptive Statistics of Weekly Absences in Other Courses.

	Week_1	Week_2	Week_3	Week_4	Week_5	Week_6	Week_7	Week_8	Week_9	Week_10	Week_11	Week_12	Week_13	Week_14	Week_15
Mean	0.38	0.26	0.44	0.66	0.75	0.72	0.80	0.91	0.94	1.01	1.50	1.61	1.81	1.98	1.86
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
Std.	0.80	0.64	0.93	1.12	1.22	1.20	1.36	1.43	1.51	1.65	1.76	1.81	1.84	1.89	1.82

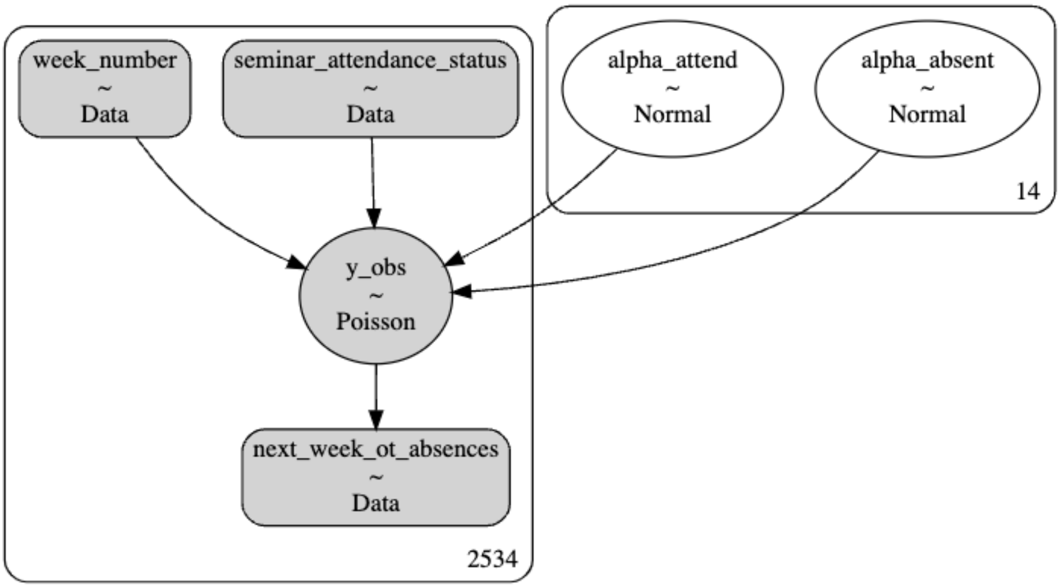


Figure 1: Weekly Impact of Seminar Attendance on OT Class Absences.

Figure 1 illustrates the model constructed for this study. The observed data include three variables: week_number, which represents the week; seminar_attendance_status, which indicates the attendance status in the seminar; and next_week_ot_absences, which represents the number of absences in OT classes in the following week. As priors, we define two parameters: alpha_attend for seminar attendance and alpha_absent for seminar absence. These parameters hierarchically influence the observed count of absences in OT classes (y_obs). The absence count in OT classes follows a Poisson distribution, where the mean (λ) is controlled by parameters such as alpha_attend and alpha_absent, reflecting the impact of seminar attendance or absence.

3 EXPERIMENTAL RESULTS AND DISCUSSION

For implementing this model, we used the programming language Python and its library PyMC, with the NUTS (No-U-Turn Sampler) algorithm as the learning method. We set the number of samples to 2,000, the burn-in period to 1,000, and the number of Markov chains to 2. The estimation results

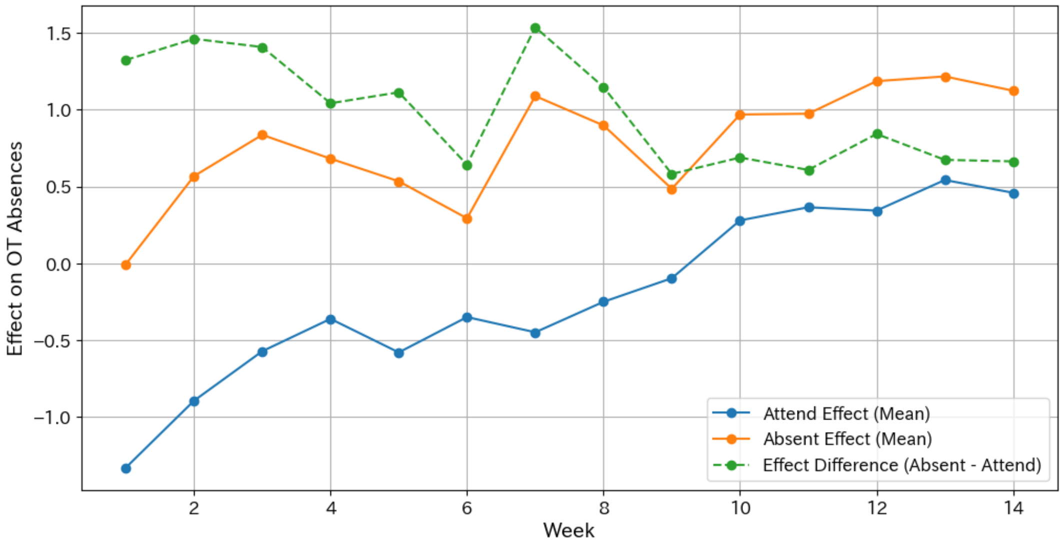


Figure 2: Weekly Impact of Seminar Attendance on OT Class Absences.

ensuring the reliability of the estimated values. showed that the Gelman-Rubin statistic (R-hat) for all parameters was below 1.1, confirming and Figure 2 shows the weekly impact of attendance and absence in the foundational seminar on absences in OT classes the following week. The vertical axis represents the average effect on absences in OT classes, where negative values indicate a reduction in absences and positive values indicate an increase. The blue solid line (Attend Effect (Mean)) demonstrates that attending the seminar contributes to reducing absences in OT classes, with particularly strong effects in Weeks 1 to 5. The orange solid line (Absent Effect (Mean)) indicates that absence from the seminar increases the risk of absences in OT classes, especially between Weeks 3 and 7. The green dotted line (Effect Difference) represents the difference between the effects of attendance and absence. Notably, during Weeks 1 to 8, the Effect Difference exceeds 1.0 in several weeks, highlighting the significant contribution of seminar attendance to reducing absences in other classes, particularly in the early part of the semester.

This study clarified the effect of attendance in the foundational seminar, an intervention class, in contributing to the reduction of absences in OT classes the following week. The results suggest that promoting attendance, particularly in the early part of the semester, is effective in reducing absences. Future research should examine interactions with other factors and assess the long-term

REFERENCES

Ministry of Education, Culture, Sports, Science and Technology in Japan. (2024). [Results of the 2023 Survey on the Number of Students Who Have Left School or Are on Leave] “Reiwa 5nendo Gakusei no Chutotaigakusha Kyugakusha Su no Chosakekka ni Tsuite (in Japanese)”.

José María Ortiz-Lozano, Antonio Rua-Vieites, Paloma Bilbao-Calabuig, and Martí Casadesús-Fa. (2020). University student retention: Best time and data to identify undergraduate students at risk of dropout. Innov. Educ. Teach. Int. 57(1), 74–85.

Naruhiko Shiratori, Tetsuya Oishi, Shintaro Tajiri, Masao Mori, and Masao Murota, (2020) [Making Dropout Patterns Using Transition of Dropout Probabil-ity] ”Chutaikakuritsu no Seni wo motiita chutaigakusei no ruikeika (in Japanese),” Nihon Kyoikukougakkai Ronbunshi (Japan journal of educational technology), 44(1), 11-22. https://doi.org/10.15077/jjet.43072

Analyzing AI-mediated Collaborative problem-solving: students' interactions and evaluations

Anni Silvola
Utrecht University, University of Oulu
a.e.silvola@uu.nl
Anouschka van Leeuwen
Utrecht University
a.vanleeuwen@uu.nl
Joonas Merikko
Annie Advisor Ltd.
joonas.merikko@annieadvisor.com
Hanni Muukkonen
University of Oulu
Hanni.muukkonen@oulu.fi
Jeroen Janssen
Utrecht University
j.j.h.m.janssen@uu.nl

ABSTRACT: Emerging use of conversational agents (CA) in different study contexts raises the need for empirical insights on the productive and ethical use of CAs in different study context. This poster presents preliminary insights on our ongoing study focusing on students' AI-mediated interactions, AI capabilities and sociocultural boundaries of CA use in the context of collaborative problem-solving. In the current study, higher education students (N=88) work in small groups of 3 to 4 students in scenario-based collaborative problem-solving situations to evaluate and experiment the use of conversational agents in study contexts, and to resolve four different types of scenarios, mediated by individual student-CA interactions. Survey, video, and chat-log data were collected to analyze students' AI-mediated interactions on individual and group-level. Preliminary insights of group-level qualitative analysis indicate that by students used AI inputs for collaboration by comparing, elaborating, contrasting, questioning both their own ideas and inputs of AI. Across the contexts, students' willingness of AI use and boundaries varied. Further findings will be presented by the time of the conference.

Keywords: conversational agents, AI capabilities, collaborative problem-solving, multichannel data

1 INTRODUCTION

Conversational agents (CAs) add new social and agentic elements on the interactions between human and technology, providing potential on more open and holistic support on the learning processes (Carolus et al., 2023). To gain better understanding about the productive and ethical ways of using CAs, more understanding is needed on the ways students interact with AI, on productive use of CAs, and students' boundaries that direct CA use. In our current research project, we investigate students' collaborative and individual AI-mediated interactions, in the context of collaborative problem-solving. We asked higher education students to work in small groups of 3 to 4 students and resolve four different types of scenario-based situations challenge situation. Scenarios included metacognitive challenges, focusing on the resource management and learning strategy use, and socioemotional challenges focusing on the situations of loneliness and insomnia. Different types of scenarios were presented for students to capture the boundaries in their thinking of the use of AI in different contexts, and to identify differences in students' AI interactions across different challenge types. Each student had access to discuss individually with CA that was prompted to support students' help-seeking processes and provide them support in challenge situations (Merikko & Silvola, 2023). The CA was a GPT-4 based prototype with the Support Bot Engine that interacts with a student through WhatsApp (Merikko & Silvola, 2023). The ongoing study provides empirical insights on the ethical, behavioral and social aspects of increasing AI use, thus aiming to inform both theory and practice about the productive uses of genAI in educational contexts.

2 RESEARCH AIMS AND OBJECTIVES

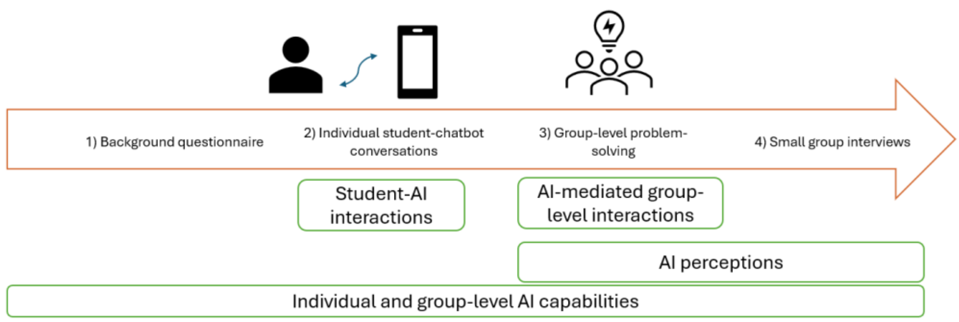
In this poster, we present preliminary findings on the group-level evaluations and observations of CA during collaborative problem-solving, and the interactions through which students translate their thinking and individual AI interactions towards collective problem-solving in the group. Focusing on the emerging collaborative interactional practices including AI-mediation (Barron, 2003), we address following research question:

1) What kinds of interactions emerge within small-groups that indicate efforts to translate individual-level student-CA inputs for groups' co-construction of knowledge?

3 METHODS

The participants of the study were bachelor- and master students from the field of educational sciences and psychology (11 male, 77 female). Students worked collaboratively with 2 to 3 scenarios they selected together. Instructors ensured that students understand instructions and monitored collaborative processes. Two parallel goals guided their collaborative learning: 1) to discuss different study-related challenges with the help of scenarios and generate collective solutions that any student could use to overcome the challenges, and 2) to experiment and evaluate CAs throughout their collaboration. Students discussed each scenario challenge first with CA. After that the group discussed scenario challenges, possible solutions, CA suggestions and constructed a solution as a group (Figure 1).

Figure 1. An overview of data collection



Survey-, video-, and chat-log data were collected. Video-data informed us about 1) students' ways of working with CAs on a group-level, 2) students' collective evaluations and experimentations of CA during collaborative problem-solving, and 3) students' AI perceptions and boundaries of AI use. chat-log data is being analyzed to identify different individual-level interactions and their connection on the group-level use of CAs. In the first phase of data-analysis, video-data were divided in the content-based episodes, each including one scenario and reflective group conversations. Three analytical layers are adopted to analyze the selected episodes from the video-data: 1) *Experimenting and evaluating CAs* that include collaborative efforts and utterances indicating how students' evaluate CA use and inputs in terms of helping their collaboration, 2) *AI-mediated knowledge co-construction* that includes such interactions through which CA use elaborates or informs their knowledge co-construction, 3) *social and ethical boundaries* that include utterances informing students' AI perceptions, possibilities and limitations of AI use, and students' reasoning on the suitable AI use. The unit of analysis is one student utterance which gives a meaningful level to identify what kinds of actions or initiations are being made in the group-level to utilize AI as a resource for learning. In the second analysis phase, epistemic network analysis is conducted to identify connection between different group-level AI-mediated contributions.

4 RESULTS

AI-mediated interactional practices include multilayered activities, with social, content-focused and ethical dimensions. Students were elaborating on and questioning their AI perceptions, address different affective reactions on the chatbot interactions, and collaboratively take different approaches to continue conversations with CA. Students' comparisons and sharing of their chatbot interactions on the group level make their varying perceptions and mixed concerns of AI use visible. In collaborative problem-solving, students use CA to elaborate on and reflect their thinking, to compare CAs' suggestions on their own experiences and idea, adding to or rejecting the ideas provided by the CA. Students' boundaries of CA use highlight the need for transparency of data use and concerns of privacy that are highlighted in cases where students are not sure of how CA works. The study addresses the novelty of CA technology for students by identifying multiple contradicting reactions, attitudes and motivations of using CA across different contexts. The data-analysis is currently being progressed.

REFERENCES

Merikko, J., & Silvola, A. (2024). An AI Agent Facilitating Student Help-Seeking: Producing Data on Student Support Needs. In M. Hlosta , I. Moser , & B. Flanagan, et al. (Eds.), *Joint Proceedings of LAK 2024 Workshops, co-located with 14th International Conference on Learning Analytics and Knowledge (LAK 2024)* (pp. 185-194). (CEUR Workshop Proceedings; Vol. 3667). CEUR-WS.org.

Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Carolus, A., Augustin, Y., Markus, A., & Wienrich, C. (2023). Digital interaction literacy model – Conceptualizing competencies for literate interactions with voice-based AI systems. *Computers and Education: Artificial Intelligence*, *4*, 100114. https://doi.org/10.1016/j.caeai.2022.100114

Epistemic Network Analysis of EFL Learner-AI Interaction in English Writing

Yunshu Xie, Suyeon Kim, Young Hoan Cho
Seoul National University
xieyunshu@snu.ac.kr, kismm@snu.ac.kr, yhcho95@snu.ac.kr

ABSTRACT: The advancement of artificial intelligence (AI) has drawn educators' attention to its educational potential. However, not all learners fully benefit from AI, and not every learner-AI interaction is equally effective. The efficacy of AI for learning depends largely on how learners interact with it. Despite the significance of learner-AI interaction, research on interaction patterns remains limited. This study investigates behavioral patterns in the learner-AI interaction during an English as a Foreign Language (EFL) writing task. Through an experiment involving 29 EFL undergraduates, three distinct interaction patterns emerged, exhibiting significant differences in their engagement with the AI tool. Findings suggest that even with the same AI tool, different learners engage in diverse ways to complete tasks. To foster productive learner-AI interactions, instructors should teach how to collaboratively participate in co-constructing knowledge with AI. Data-driven adaptive systems should be developed to enable monitoring of students' behavioral patterns, allowing instructors to provide personalized support.

Keywords: Artificial intelligence, Interaction, English writing, EFL

1 INTRODUCTION

Technological advancements are positioning AI as a crucial tool in education, particularly in language education. AI tools enhance learners' writing skills by providing systematic feedback on grammar, spelling and related aspects (Liu et al., 2021). However, it should be noted that not all learner-AI interactions are uniformly effective; the success of these interactions depends on how they are conducted (Wang et al., 2023; Kim & Cho, 2023). While AI-mediated learning holds significant promise, the underlying mechanisms of learner-AI interaction remain underexplored (Kim et al., 2024). Simply measuring changes in language skills can obscure important intermediate processes. For example, how learners accept or reject AI feedback and decide which suggestions to incorporate into their learning is unclear. Without understanding these processes, AI-supported learning resembles a "black box" where inputs and outputs are observable, but the critical learning mechanisms remain hidden. This study aims to address these gaps by examining interaction patterns between EFL learners and AI tools, providing insights into improving AI-driven educational solutions. This study seeks to answer the following research questions: (1) How many distinct clusters are identified in learner-AI interaction processes during the EFL writing task? (2) What are the differences in learner-AI interaction patterns between the clusters?

2 METHOD

This study involved 29 EFL undergraduates from diverse academic majors in South Korea. QuillBot (https://quillbot.com), a user-friendly AI tool, was used to support English writing. Participants had no prior experience with QuillBot and received instructions on the AI tool. After a practice session, participants engaged in 30 minutes of English essay writing under AI-supported conditions. The TOEFL writing prompt "Compare and contrast knowledge gained from experience with knowledge gained from books." was selected as the writing topic. All writing tasks were conducted in a laboratory, with the entire process recorded on video. To address the research questions, learners' behavioral data were segmented into idea units using ATLAS.ti™ software and coded according to coding schemes (Kim et al., 2023). The coding scheme included six categories: planning, individual writing, AI recommendation revision, AI recommendation acceptance, AI recommendation rejection, and monitoring. Three researchers independently analyzed video recordings using the coding scheme and the inter-rater reliability was high (Cohen's Kappa at .96). All disagreements were resolved through discussions.

3 RESULTS

3.1 Clusters of learner-AI interaction

This study explored learner-AI interaction patterns by investigating the number of clusters with hierarchical cluster analysis and carrying out k-means cluster analysis. Three learner-AI interaction patterns were identified: learner directed (Cluster 1), AI-dependent (Cluster 2) and collaborative (Cluster 3) Interaction. Kruskal-Wallis H test results indicated significant differences between mean ranks of the three clusters in individual writing ($H = 22.47$, $p < .01$), AI recommendation revision ($H = 10.09$, $p < .01$), AI recommendation acceptance ($H = 14.84$, $p < .01$), and monitoring ($H = 12.74$, $p < .01$). Specifically, cluster 1 (C1, N=15), characterized by learner directed interaction patterns, showed high independence with minimal AI usage (M=73.18, SD=7.42). Cluster 2 (C2, N=5), AI-dependent interaction, showed the highest levels of AI recommendation acceptance (M=24.94, SD=6.72) and monitoring (M=27.88, SD=6.33). Cluster 3 (C3, N=9), collaborative interaction patterns, exhibited a high proportion of individual writing behaviors (M=51.34, SD=6.59) and the highest AI recommendation revision (M=8.45, SD=7.56) among the three clusters. C3 also had a higher proportion of AI recommendation acceptance (M=22.99, SD=8.10) than C1.

3.2 Differences of learner-AI interaction patterns

Epistemic Network Analysis (ENA) was conducted using the ENA Web Tool to explore behavioral relationships across clusters. As shown in Figure 1, one significant distinction between C1 and both C2 and C3 is the absence of clear connections between acceptance-oriented interaction behaviors and other elements, suggesting that C1 learners lead the writing process independently, showing less reliance on AI recommendations. In contrast, C2 shows a significant distinction from C1 and C3 in that AI recommendation revision are not notably linked with other behaviors. However, for C2 learners, monitoring, acceptance, individual writing, and rejection behaviors often occurred around the same time, indicating a tendency among them to either readily accept or reject the AI's recommendations without revisions. They also actively monitor the appropriateness of their writing and AI usage, adjusting their strategies as needed. This pattern, characterized by one-way interactions, involves learners passively accepting AI recommendations. Lastly, C3 demonstrates that acceptance behaviors frequently co-occur with monitoring, rejection, and revision-oriented interactions. Compared to C2, C3 also demonstrates a tendency for individual writing and AI recommendation revision to appear together. This implies that learners in C3 engage in more mutual communication with the AI, valuing its recommendations but not following them mechanically. Instead, they critically evaluate each recommendation and choose to accept, revise, or reject it based on their needs.

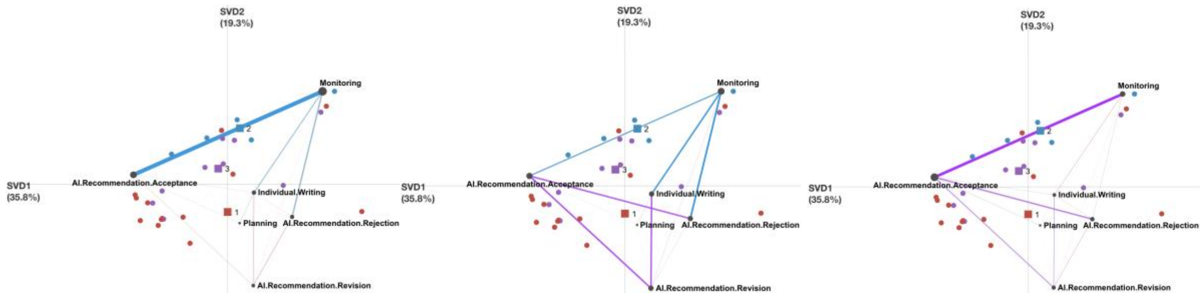


Figure 1: Network comparison of C1 (red), C2 (blue), and C3 (purple), shown in the order of C1-C2, C2-C3, and C1-C3. Edge width indicates the frequency of co-occurrences between codes.

4 DISCUSSION & CONCLUSION

This study found three distinct learner-AI interaction patterns (learner-directed, AI-dependent, collaborative Interaction), each demonstrating a unique approach to using AI tool for EFL writing. These findings are consistent with Kim et al. (2023) but expand upon its results in that the AI tool in this study provided continuous, input-based feedback, enabling more flexible, personalized interactions. The learner-AI interaction, a knowledge co-construction process, requires appropriate instructor guidance on handling AI recommendations to facilitate meaningful learning. For learners who either underutilize or overly depend on AI recommendations, data-driven adaptive systems are essential to monitor students' cognitive and emotional states, learning strategies, and engagement in real-time, while tracking their interaction and behavior patterns. Such systems can enable instructors to provide personalized support at critical moments in specific tasks. Additionally, engineers should consider these diverse learner-AI interaction characteristics when developing AI tools for EFL education. While these findings provide valuable insights, their generalizability is limited by our small sample size and use of a single AI tool. Future research should examine these patterns across larger populations and multiple AI platforms.

REFERENCES

Kim, H., Cho, Y. H., & Park, S. (2023). Exploring the interaction patterns between learners and AI translator in English writing. *The Journal of Educational Information and Media, 29*(1), 201-228. https://doi.org/10.15833/KAFEIAM.29.1.201

Kim, J., & Cho, Y. H. (2023). My teammate is AI: Understanding students’ perceptions of student-AI collaboration in drawing tasks. *Asia Pacific Journal of Education,* 1–15. https://doi.org/10.1080/02188791.2023.2286206

Liu, C., Hou, J., Tu, Y. F., Wang, Y., & Hwang, G. J. (2021). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 1-19. https://doi.org/10.1080/10494820.2021.2012812

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194*, 104703. https://doi.org/10.1016/j.compedu.2022.104703.

Tell Me Why I Should "Not" Follow Your Recommendation: On the Role of Explainable AI in Collaborative Human-AI Decision Making

Hasan Abu-Rasheed, Christian Weber, Madjid Fathi
University of Siegen
hasan.abu.rasheed@uni-siegen.de

ABSTRACT: Learning personalization has gained an ever-growing interest in recent years, due to its potential for offering tailored recommendations for the learner's needs and goals. However, the decision to adopt the learning recommendation is not solely that of the AI-based recommender, but involves the learners themselves to a great extent, since they are investing the time and effort to follow the recommendation. In this research, we investigated the dynamics of the decision-making process between the human learner and the AI system. We introduce an explainable AI approach to support the human understanding of the intelligent recommendations, and thus their ability to collaboratively modify it, and make a final decision on adopting it. We implement a multimodal explanation approach, combining information from domain experts and large language models (LLMs) to provide a comprehensive overview of the recommendation reasoning. We test our system in a complex scenario within nursing training. Our findings point out the high acceptance rate of explained recommendations, and the role that our explanations played in supporting the agency of the learners over making an informed decision about the recommendation.

Keywords: Explainable AI (XAI), Human-AI collaboration, Multimodal explainability, Recommender systems, Knowledge graphs.

1 INTRODUCTION

Personalized learning has become an important part of modern educational settings, especially with the rise of adaptive learning systems that aim to meet each learner's unique needs and abilities. These AI-supported systems provide recommendations based on learners' prior knowledge, engagement levels, learning goals, etc., enhancing the learning experience. However, recent research highlights that there should be a balance between AI-based recommendations and human input (Molenaar, 2022), ensuring that learners feel empowered to actively participate in making decisions about their learning pathways, rather than passively accepting AI suggestions. This active involvement allows learners to retain agency and utilize critical thinking while engaging with the recommendation.

In educational contexts, a collaborative approach to human-AI interaction is increasingly emphasized. This approach leverages the complementary strengths of AI systems and human facilitators, such as teachers or the learners themselves, for a balanced, hybrid model of decision-making (Holstein et al., 2019). Rather than relying exclusively on AI recommendations, collaborative systems invite human input to adjust, interpret, and sometimes override the AI's recommendations. This interplay is especially valuable when the AI system provides transparent explanations for its recommendations, allowing learners and teachers to understand and, if necessary, refine the suggestions (Abu-Rasheed, Abdulsalam, et al., 2024). Such hybrid models are not only conducive to better educational outcomes but also build trust in AI systems by fostering a sense of shared control (Ooge et al., 2022), aligning AI recommendations with pedagogical goals, and adapting to the broader educational context.

2 METHODOLOGY

In this study, we propose an approach based on explainable AI (XAI) for supporting collaborative decision-making about learning-path recommendations, particularly for nursing vocational education and training (VET), where training demands both cognitive and practical decision-making skills to support solving complex, real-world problems. Our approach has been developed within a 3-year project, focusing on elderly home nurses and the complex challenges they encounter daily, particularly in emergency scenarios such as evacuating elderly individuals with dementia during a fire. To that end, we generate a learning path recommendation for the nurses, based on a problem scenario. Then, we utilize a combination of multimodal explanations, visual and textual, which are supported by LLMs and expert input, to clarify the reasoning for the learning path selection for each learner. We emphasize the integration of domain experts in the process of defining the explanation goals, content, and generation process, to offer a controlled and high-quality rationale of explaining the AI-based recommendations. To generate the learning-path recommendation and its explanations, we utilize a knowledge graph (KG) as a network data structure, which is used to search for all learning paths, and rank them through a reinforcement learning approach, for which the learning environment is the KG itself. An explanation module is then developed to explain the reasoning for the graph-based recommender system. The KG structure allows creating relations between learning materials from different domains on a semantic level. Those relations enable the path-finding and ranking algorithm to include learning content from different domains that are required to solve the complex problem.

Our explanation approach is designed based on TExKG framework (Abu-Rasheed et al., 2024) to extract a comprehensive overview of the recommended learning path from the KG nodes and relation. Textual explanations are constructed as a template, which contains pre-defined information slots that domain experts and a controlled-LLM pipeline fill, see Figure 1. Visual explanations are generated from the KG environment, and they show the connections among the recommended learning materials, as well as their connections to the surrounding learning content in the KG, resulting in positioning the recommended path in a certain context that the graph relations and graph communities reveal.

3 EVALUATION AND RESULTS

We evaluated our explainability approach for human-AI collaboration with 24 staff members in two elderly homes, divided into control and treatment groups of an A/B test. A complex problem was identified for all participants and a personalized recommendation was generated for treatment group members, along with its explanation. Additionally, qualitative feedback was collected from the participants about their experiences. We measure the learner's irritation with the recommended path to reflect the change of their recommendation acceptance when the explanation is presented. Our results show that treatment group has 37.5% more acceptance of the recommendations, in comparison to the control group. Participants also reported that they were able to "skip" parts of the recommended path because the explanation allowed them to understand that they had already learned similar content, which they did not include in their profiles. This highlights the potential for participants to take informed actions, such as modifying their profile, to adjust the recommendation and adopt it fully or partially. Our results also show that the combination between expert and LLM textual explanations is necessary to provide explanations on contextual and reflection levels, because LLMs were not solely able to reach reflection-level explanations of the learning path. This points out, in turn, the role of human-AI collaboration in the task of generating the explanation itself, not only the task of making a decision about the recommendation.

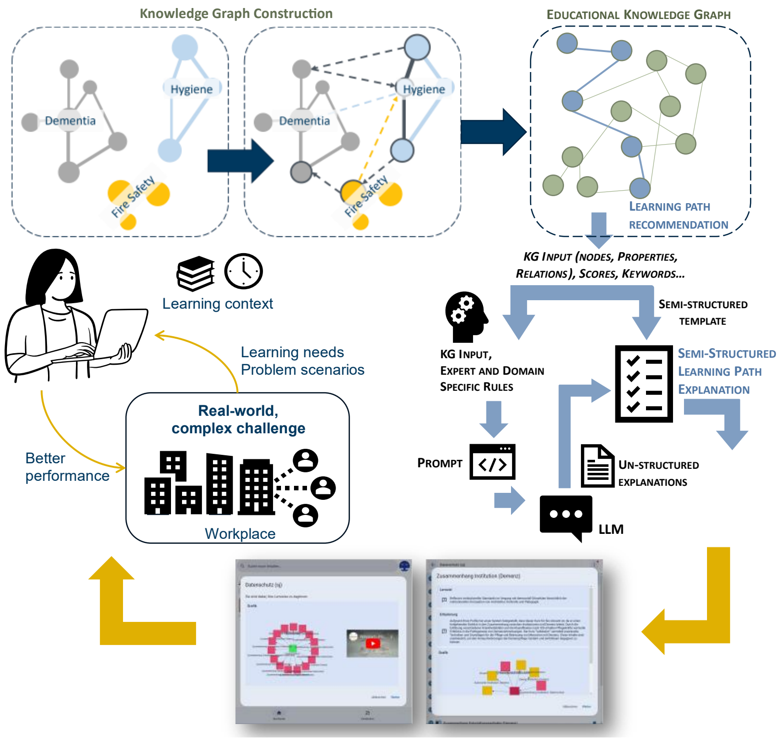


Figure 1. KG construction and explainability approach

4 CONCLUSION

In this paper, we presented an XAI approach for supporting human-AI collaboration in the decision-making process for learning recommendations. Our system utilizes a KG structure and domain expert input to generate textual and visual explanations of learning-path recommendations. Evaluation results of the proposed explainability solution demonstrate improved acceptance and increased decision-making ability among learners, regarding the adoption of the recommendations and understanding how to modify the system output, supporting their agency over the AI-system predictions and their own learning process.

REFERENCES

Abu-Rasheed, H., Abdulsalam, M. H., Weber, C., & Fathi, M. (2024). *Supporting Student Decisions on Learning Recommendations: An LLM-Based Chatbot with Knowledge Graph Contextualization for Conversational Explainability and Mentoring*. https://doi.org/10.48550/ARXIV.2401.08517

Abu-Rasheed, H., Nadeem, M., Dornhöfer, M., Zenkert, J., Weber, C., & Fathi, M. (2024). TExKG in Health Domain: The Application of Knowledge Graph Based Framework for Explainable Recommendations in the Contexts of Elderly Care, Mental Health, and Emergency Responses. In M.-R. Alam & M. Fathi (Eds.), *Integrated Systems: Data Driven Engineering* (pp. 265–285). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53652-6_16

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (Vol. 11625, pp. 157–171). Springer International Publishing. https://doi.org/10.1007/978-3-030-23204-7_14

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. https://doi.org/10.1016/j.caeai.2022.100070

Ooge, J., Kato, S., & Verbert, K. (2022). Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. *27th International Conference on Intelligent User Interfaces*, 93–105. https://doi.org/10.1145/3490099.3511140

The 'Promise' of LMS Dashboards

Danielle Hagood
University of Copenhagen
dh@ind.ku.dk

Diana Meda
University of Copenhagen
dihi@di.ku.dk

ABSTRACT: This poster aims to spark discussion around what learning management system (LMS) analytic dashboards promise to reveal. Our analysis evaluated four major LMS learning analytic dashboards (those offered by Instructure, Blackboard, D2L, and Moodle) at two timepoints from before (2018) and after (2024) the pandemic. Our results describe the claims and concepts related to learning-analytic dashboards. To best utilize the poster format, we present conceptual data visualizations to summarize the promises of LMS dashboards at each timepoint and to compare and contrast the differences between 2018 and 2024 around thematic discussion points from scholarly discourse and our mixed methods analysis.

Keywords: LMS, dashboards, higher education, mixed-methods

1 CONTEXT AND MOTIVATION

Today nearly all universities rely on learning management systems (LMSs) that also offer an analytic dashboard. This trend excellerated in response to the global COVID-19 pandemic, which pushed instructors to further rely such platforms (Turnbull et al., 2021). More recently, the explosion of generatiev AI, such as ChatGPT, has impacted instructors expectations and awareness of analytics (Pischetola et al., 2024). In essence LMS analytic dashboards are now a major part of teaching in the university and a primary tool which instructors consult to learn about what students do (e.g., Macfadyen & Dawson, 2010). Since LMSs dashboards and analytics can frame instructors' thinking and teaching, we consider it important to study the promises and claims imbedded in such analaytics in instructor-facing dashboards.

This poster aims to spark a discussion around the ways such analytic dashboards an impact instructors' consequential perceptions and decisions. Our analysis evaluates four major LMS learning analytic dashboards (those offered by Instructure, Blackboard, D2L, and Moodle) at two timepoints from before (2018) and after (2024) the pandemic. This analysis is framed in the context of three trends driving LMS adoption: increasing preference and demand for constructivist instruction (Wise & Quealy, 2006), accountability culture in higher education (Prinsloo et al., 2014), and how datafication transforms of educational activities (Williamson, 2017). In sum, we explore both the types of analytics are presented in the LMS system as well as what types of claims and rhetoric are made by the companies about LMS.

2 METHODS

Our analysis includes two sources of data: public descriptions of LMS and LMS dashboard technical specifications. Regarding the first, we assembled a text corpus by scrapping the websites of the LMS providers and downloading relevant grey-literature (e.g., API manuals) in both 2018 and 2024. We processed the text to identify relevant segments for analysis. We did this by retaining HTML element that included keywords we developed based on studying the websites (for example, “dashboard”, “analytic”, “intelligence”, “predict”, “AI”). Finally, we conducted a mixed-methods analysis of the resulting text subset including qualitatively reading and coding the text (Merriam, 2009) and quantitatively using common text summary analyses, including TF-IDF, concept networks, and sentiment analysis (Lindgren et al., 2020). Regarding the second, we accessed technical descriptions of the LMS dashboards (e.g., API dictionaries) where available.

3 RESULTS & DISCUSSION POINTS

To best utilize the poster format, we present conceptual data visualizations to summarize the promises of LMS dashboards at each timepoint and to compare and contrast the differences between 2018 and 2024 around thematic discussion points from scholarly discourse and our mixed methods analysis. (While we discuss differences between LMS, our presentation of findings does not emphasize differences between specific LMS.) Here we briefly describe key results linked to the discussion points we hope to promote with this poster, space restrictions prevent the visual presentation of the full results here.

3.1 Misalignments in how student engagement is defined

Overall, in both 2018 and 2024, there are major differences between the log-based measures in LMS dashboards and scholarly, validated measures (e.g., psychometric assessments) of motivational constructs, such as engagement. Focusing on engagement specifically, LMS dashboards conflate behavioral engagement and academic performance metrics in a way that doesn’t align with theory. Also, platforms vary in their definitions and representations of such engagement metrics. For instance, in 2018, Moodle represented activity completion with attained badges and competencies and an activity completion checklist. Canvas and Blackboard represented interactions with the platform and compare usage between students. Blackboard also represents activity with a scatter plot between activity and course grade. And in general, most LMS dashboards presented metrics that are related to behavioral engagement, in which LMS dashboards simply refer to such metrics as measures of engagement without differentiating various forms of engagement.

3.2 The shift from “at risk prediction” to “AI optimization”

In 2018, one of the major features promoted in LMS dashboards was the ability to detect at risk students. (We have debated the validity of “at risk” status in prior work, see Hagood, 2021) The majority of LMS, namely Blackboard, Canvas, and D2L, flagged students as at-risk based on based on low engagement and defined risk in terms of failing a course. Moodle was slightly different, defining at-risk is explicitly as “dropping out" or "no student activity in the last quarter of the course” rather than receiving a failing grade—likely because Moodle supports many online courses.

In both 2018 and 2024, the LMS dashboards highlighted identifying at-risk students using both descriptive and predictive methods. Adopting the descriptive approach, Blackboard and Canvas dashboards presented user engagement and performance statistics, which allow instructors’ “see which students are at-risk and need help” (Canvas Doc Team, 2018). Adopting a predictive approach, Moodle used a machine-learning model to identify at risk students using engagement metrics, while Blackboard and D2L both predict risk status using external data (e.g., Blackboard Predict and Performance Plus). Thematically, we found that the discussion of these tools focused on detection of factors contribute to at-risk status. This resulted in student-centered use of machine learning.

In contrast, in 2024, there is much more emphasis on AI features compared to prediction. This is unsurprising given the perceived cutting-edge: machine learning in 2018; generative AI in 2024. In 2024, nearly all LMS promote a use of AI, with a strong emphasis on how “assistants” might support teachers. We found that this foregrounded a more more teacher-centered perspective than in 2018, due to an emphasis on the way such tools can, for example, automate non-essential aspects of teaching and make instructional design more efficient.

4 CONCLUSION

In conclusion, with this poster we describe the framing and claims related to learning-analytic dashboards in LMS. We take look at the current (2024) and past (2018) rhetoric to contextualize the development of such dashboards in response to the pandemic and emergence of AI. We highlight key points for discussion informed by scholarly discourse and a mixed methods analysis. These discussion points most closely relate to the conference themes of expanding learning analytics methodological toolbox and rethinking learning analytics practice.

REFERENCES

Canvas Doc Team. (2018, August 25). What are analytics? Retrieved from Canvas website: https://community.canvaslms.com/docs/DOC-10742-67952724559

Hagood, D. E. (2021). *How University Instructors See Student Engagement and Risk Status: Constructing Definitions with Information from Instructional Practices and Learning Management Systems*. University of California, Davis.

Lindgren, B. M., Lundman, B., & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. *International journal of nursing studies*, *108*, 103632.

Macfadyen, L. P. & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education, 54*(2), 588-599.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Pischetola, M., Stenalt, M. H., Nøhr, L., Hagood, D. E., & Misfeldt, M. (2024). Desirable and realistic futures of the university: a mixed-methods study with teachers in Denmark. *International Journal of Educational Technology in Higher Education*, *21*(1), 29.

Prinsloo, P., Khalil, M., & Slade, S. (2024). Learning analytics as data ecology: A tentative proposal. *Journal of Computing in Higher Education*, *36*(1), 154-182.

Turnbull, D., Chugh, R., & Luck, J. (2021). Transitioning to E-Learning during the COVID-19 pandemic: How have Higher Education Institutions responded to the challenge?. *Education and Information Technologies*, *26*(5), 6401-6419.

Williamson, B. (2017). Big data in education: The digital future of learning, policy and practice. London, UK: SAGE

Wise, L., & Quealy, J. (2006). At the limits of social constructivism : Moving beyond LMS to re-integrate scholarship. *Proceedings of the 23rd Annual Ascilite Conference: Who’s Learning? Whose Technology?*, 899–907.

Examining the relationship between Repetitive Reflective Writing, Self-Regulated Learning Strategies, and Metacognitive Monitoring

Nazanin Reza zadeh Mottaghi [a], Srecko Joksimovic [a], Negin Mirriahi [a], Jelena Jovanovic [b], Wimukthi Thommadurage [a], Sithara Walpita Gamage [a], Shane Dawson [a]

[a]University of South Australia, Centre for Change and Complexity in Learning
[b]University of Belgrade, Faculty of Organizational Sciences

ABSTRACT: University students encounter various academic challenges that necessitate strong self-regulated learning (SRL) skills. SRL is essential for effectively managing the learning process, which involves setting goals, monitoring progress, and adapting strategies with minimal external assistance. A crucial aspect of SRL is metacognitive monitoring, which enables students to evaluate their learning and make informed adjustments. While research has predominantly focused on various aspects of metacognition, there is limited research on calibrating metacognitive monitoring using repetitive reflective writing practices. This study aims to address this gap by using a case study methodology to explore the relationship between repetitive reflective writing, metacognitive monitoring, and SRL strategies among undergraduate students through statistical and content analysis. The results indicated that repetitive reflective writing did not consistently improve metacognitive monitoring accuracy. The study also identified that SRL strategies such as goal setting and planning, environmental structuring, and seeking social assistance were associated with changes in metacognitive monitoring accuracy in the studied undergraduate programs.

Keywords: self-regulated learning, reflective writing, metacognitive monitoring, metacognitive judgments, large language models

1 INTRODUCTION AND BACKGROUND

University students must develop strong self-regulated learning (SRL) skills to meet various academic demands, including setting goals, monitoring progress, and adjusting strategies with minimal external support. These skills are especially important when transitioning from structured learning environments, such as secondary education, to more independent university settings. However, many students struggle with independent planning, monitoring, and evaluating their learning, often leading to a misalignment between their expectations and actual performance (Morphew, 2021). This study uses a case study approach to investigate how repetitive reflective writing and SRL strategies impact metacognitive monitoring among undergraduate engineering and early childhood pre-service education students, aiming to determine whether reflective practices can enhance metacognitive accuracy, and identify which specific SRL strategies predict metacognitive accuracy.

2 METHODS

This case study explored reflective writing, metacognitive monitoring, and SRL strategies in a third-year undergraduate Engineering course and a first-year Education course at an Australian university. By comparing disciplinary contexts, it examined how reflections and SRL strategies relate to metacognitive monitoring, aiming to enhance students' metacognitive awareness and SRL through distinct course designs:

1. Are repetitive reflections associated with the accuracy of metacognitive monitoring of undergraduate students?

2. To what extent do SRL strategies that students adopt on a particular assessment task predict the students' metacognitive monitoring accuracy on that task?

The assessment tasks incorporated reflective prompts encouraging students to evaluate the strategies they used while preparing and their anticipated grades. Tailored reflective questions, provided post-assessment, aimed to guide students in systematically analyzing their performance, fostering effective reflective practices, and enhancing their learning processes for improved academic outcomes.

3 DATA ANALYSIS

In response to RQ1, descriptive statistics were performed to analyze student data on anticipated and actual assessment grades, calculating the discrepancy by subtracting actual grades from expected grades without considering direction. To assess changes in metacognitive monitoring resulting from reflection within the two cohorts, linear mixed-effects models were applied.

model.1 <- lmer (Metacognitive monitoring ~ Reflection. Order + (1 | Student_ID), data = My.data)

To address RQ2, we conducted content analysis using GPT-4o, a prominent large language model (LLM) developed by OpenAI. We first adapted a coding scheme for SRL strategies based on Zimmerman and Pons' framework of SRL strategies (Zimmerman & Pons, 1986). The strategies included in this study were: Organizing and transforming, Goal setting and planning, Seeking information, Keeping records and monitoring, Environmental structuring, Self-consequences, Rehearsing and memorizing, Seeking social assistance, and Reviewing records. To train the LLM for the coding task, we used a subset of human-coded reflections through four stages, beginning with an initial set of 10 reflections. Two researchers collaboratively coded ten reflections to establish a baseline. The same ten reflections were then coded by the LLM to develop an appropriate prompt and initiate the coding procedure. Each reflection underwent five separate coding rounds to ensure the reliability of the LLM's codes. The final code was determined by taking the majority vote from these rounds. This five-round coding approach was implemented to minimize variations in the interpretation of the responses. Subsequently, the two researchers coded another 60 reflections (three sets of 20) independently, with any discrepancies resolved through discussion. The human coding agreement for these three sets of reflections reached 0.6 to 0.85 range at all stages. Following this, the same subsets of reflections were coded by the LLM with the final version of the prompt (the prompt was modified several times according to the researchers' discussion) and using the same five-round approach. The agreement between human coders and ChatGPT coding ranged from 0.65 to 0.8 based on Kappa statistics, indicating the model's reliability in qualitative coding. To ensure data privacy and security, the OpenAI API key was hosted by the university, with access restricted exclusively to the research team.

After identifying SRL strategies in students' reflections, we calculated the occurrence of each strategy within each reflection. Mixed-effects models were then applied to assess the impact of these SRL strategies on metacognitive monitoring accuracy**.**

model.1 <- lmer (Metacognitive monitoring ~ organizing + planning + seeking information + keeping records + environmental structuring + Rehearsing and memorizing + seeking social assistance + reviewing records + self-consciousness + (1 | Student_ID), data = My.data)

4 CONCLUSION, LIMITATIONS, AND IMPLICATIONS FOR FURTHER RESEARCH:

This study explored the effects of repetitive reflective writing and SRL strategies on undergraduate students' metacognitive monitoring. The results indicated that repetitive reflective writing was not associated with a consistent improvement of accuracy in metacognitive monitoring in either cohort. However, an initial improvement was present in the Engineering cohort. In terms of SRL strategies, our study revealed that *goal setting and planning* were associated with metacognitive monitoring in the Education cohort, whereas *seeking social assistance* and *environmental structuring* were associated with metacognitive monitoring in the Engineering cohort. While this study presented an innovative method for detecting SRL strategies using an LLM, it is essential to recognize the limitations of employing LLMs for content analysis tasks, since LLMs have not yet shown consistent performance and reliability in this area (Hou et al., 2024). Consequently, further research is necessary to investigate validity of SRL strategy detection through LLM-based content analysis.

In the Education cohort, *goal setting and planning* were positively associated with metacognitive monitoring. On the other hand, in the Engineering cohort, *environmental structuring* had a positive association with metacognitive monitoring, while *seeking social assistance* had a negative association. These findings suggest that the effectiveness of specific strategies for improving metacognitive monitoring depends on the field of study and the types of assessment tasks. However, further research is needed to explore ways to understand the relationship between different learning strategies and the accuracy of metacognitive monitoring, particularly in varied learning contexts. In conclusion, while our research provides valuable insights, these limitations highlight important areas for future exploration. Addressing these factors could enhance our understanding of how reflective practices and SRL strategies may influence the students' metacognitive monitoring.

5 REFERENCES

Hou, C., Zhu, G., Zheng, J., Zhang, L., Huang, X., Zhong, T., Li, S., Du, H., & Ker, C. L. (2024). Prompt-based and Fine-tuned GPT Models for Context-Dependent and -Independent Deductive Coding in Social Annotation. *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 518–528. https://doi.org/10.1145/3636555.3636910

Morphew, J. W. (2021). Changes in metacognitive monitoring accuracy in an introductory physics course. *Metacognition and Learning*, *16*(1), 89–111. https://doi.org/10.1007/s11409-020-09239-3

Zimmerman, B. J., & Pons, M. M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal*, *23*(4), 614–628. https://doi.org/10.3102/00028312023004614

PolyFeed: A Student-Facing Feedback Analytics Tool to Facilitate Feedback Processes

Bhagya Maheshi, Flora Jin, Mikaela Milesi, Hiruni Palihena, Ahmad Aldino, Thomas Ng, Shaveen Singh, Caleb Ooi, Minhua Zhou, Andrew Pham, Danijela Gasevic, Nicola Charwat, Philip Wing Keung Chan, Guanliang Chen, Wenhua Lai, Roberto Martinez-Maldonado, Dragan Gašević, Yi-Shan Tsai
Monash University

[Bhagya.GangodaGamachchige, Flora.Jin, mikaela.milesi, hiruni.palihena, ahmad.aldino, thoas.ng, Shaveen.Singh, cooi0004, minhua.zhou, andrew.pham, Danijela.Gasevic, nicola.charwat, Philip.K.Chan, Guanliang.Chen, wenhua.lai, Roberto.MartinezMaldonado, Dragan.Gasevic, yi-shan.tsai]@monash.edu

ABSTRACT: PolyFeed is an innovative feedback analytics (FA) tool designed to enhance students' engagement with feedback across various assessments and courses. Grounded in feedback theories and learning analytics, PolyFeed addresses the challenge of understanding and supporting students' interactions with feedback. Its key features include a browser extension and a dashboard. The browser extension enables students to highlight and label feedback, create reflective notes or action plans, and seek further explanations from GenAI, while the dashboard visualises insights, helping students track their strengths, weaknesses, and action plans across multiple assessments. A pilot study involving 18 higher education students in authentic learning settings revealed PolyFeed's potential to enhance learning through FA. Participants interacted with over 6 pieces of feedback each, resulting in 600 annotations, 457 notes, and 167 action plans. Students highly appreciated PolyFeed's ability to consolidate all feedback in one place and motivate engagement through analysis and visualisation. This demo will showcase PolyFeed's functionalities and their potential to transform student learning by improving their ability to engage with, understand and utilise feedback. Through a centralised platform for feedback management and analysis, PolyFeed empowers students to take ownership of their feedback process, fostering a more reflective and proactive approach to academic growth.

Keywords: Feedback analytics, feedback management, learning analytics, feedback interaction

DEMO VIDEO

Link: https://drive.google.com/file/d/1IEAVZkLpR62LwCAByHq-pRWAUQy5xnNL/view?usp=sharing

Enhance Your Presentation Skills with Presentable

Daniele Di Mitri[1], Jan Schneider[1], Nina Mouhammad[1], Stefan Hummel[1], Mohamad Alomari[2], Mahmoud Alhag Ali[1], Md Hafizur Rahman Masum[1], Hunain Arif[1], Melina Rose[3], and Roland Klemke[3]

[1]DIPF | Leibniz Institute for Research and Information in Education, Germany,
[2]Maastricht University, The Netherlands
[3]Cologne Game Lab, TH Köln, Germany

corresponding email: d.dimitri@dipf.de

ABSTRACT: Presentable is a research software designed to enhance users' presentation skills through Artificial Intelligence (AI)-driven feedback and guidance. Building on previous Oral Presentation for Automated Feedback systems (OPAFs) like the Presentation Trainer, Presentable offers a comprehensive approach to message composition and rehearsal. During the message composition, users are guided in creating a draft for their presentation, where they learn best practices and tailor their presentation to their audience. This includes identifying key topics of the presentation, prior knowledge of the audience, and desired objectives, ensuring a coherent flow from introduction to conclusion. The rehearsal phase focuses on effective speech delivery. First, users practice their presentation by focusing on their voice while Presentable records the audio of it. Then, while replaying the recording, users are guided through a series of self-reflection questions. Next, users are asked to practice their presentation by focusing on body language. Presentable employs a camera-based approach for body pose recognition, detecting and providing immediate corrective feedback on common presentation mistakes such as closed posture or improper hand display. This feedback helps users correct these errors, enhancing muscle memory and retention. After practising the presentation, users can look at the video recording and are guided through a second self-reflection phase. Presentable securely stores data in an online database, allowing users to access and review their sessions anytime. Designed for integration with Learning Management Systems, it ensures data remains on the educational institution's server. Teachers can use the Presentable dashboard to monitor student progress, correct mistakes, and address misleading feedback. Developed by the HyTea project, Presentable is funded by the Federal Ministry of Education in Germany. Link to the video demo: https://www.youtube.com/watch?v=_px-MHfkb6c.

Keywords: Oral Presentation Skills, AI Feedback, Presentation Trainer, Educational Technology, Multimodal Learning Analytics

.

SimVision: Supporting Reflection in Team Healthcare Simulations with AI-powered Analytics and GenAI

Riordan Alfredo[1], Vanessa Echeverria[1,3], Linxuan Zhao[1], Jie Xiang Fan[1], Mikaela Milesi[1], Yueqiao Jin[1], Lixiang Yan[1], Xinyu Li[1], Zachari Swiecki[1], Gloria Fernandez-Nieto[1], Samantha Dix[1], Rosie Wotherspoon[1], Rebecca Winter[1], Hollie Jaggard[1], Abra Osborne[1], Dragan Gašević[1], Simon Buckingham Shum[2], Roberto Martinez-Maldonado[1]

[1]Monash University, [2]University of Technology Sydney, [3]RMIT University

Corresponding authors: riordan.alfredo@monash.edu, vanessa.echeverria@monash.edu, roberto.martinezmaldonado@monash.edu

ABSTRACT: Simulation training is critical for future nursing professionals in healthcare, where students have the opportunity to apply their knowledge of patient management, teamwork, and communication skills into practice. However, short and stressful scenarios often leave students needing help to reflect on their performance after the simulation (e.g., teacher-led debriefing and individual reflection). Traditional reflection methods may not provide evidence to guide teams' performance discussions. To address this challenge, we present **SimVision**, an interactive LA system co-designed with educators and students to support reflection in nursing simulation. SimVision consists of three parts: teacher-facing tagging tool, debriefing dashboard, and student-facing individual reflection dashboard. SimVision leverages the power of AI and multimodal data, capturing teacher's observations, and students' data from position tracking, audio, and heart rate sensors. The collected data is presented as visualisations [1-2]. During the debrief, teachers can access the analytics to select visualisation and discuss the team's performance. After the debrief, students have access to a reflection dashboard in two versions: 1) AI summaries and data comics [3], and 2) a conversational agent (VizChat)[4]. We have conducted evaluations involving five nursing educators assessing the teacher-facing tools since 2023, as well as 42 nursing students evaluating the student-facing dashboards after completing their simulations in 2024. Demo link: https://youtu.be/gy2ZqCnYtJ4

Keywords: Learning Analytics Dashboard, Large Language Model, GenAI, Multimodal Learning Analytics, Teamwork, Reflection

References

[1] R. Alfredo et al., "Designing a Human-centred Learning Analytics Dashboard In-use," Journal of Learning Analytics, pp. 1–20, Oct. 2024, doi: 10.18608/jla.2024.8487.

[2] V. Echeverria et al., "TeamSlides: a Multimodal Teamwork Analytics Dashboard for Teacher-guided Reflection in a Physical Learning Space," in Proceedings of the 14th Learning Analytics and Knowledge Conference, in LAK '24. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 112–122. doi: 10.1145/3636555.3636857.

[3] M. E. Milesi *et al.*, "'It's Really Enjoyable to See Me Solve the Problem like a Hero': GenAI-enhanced Data Comics as a Learning Analytics Tool," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, in CHI EA '24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–7. doi: 10.1145/3613905.3651111.

[4] L. Yan et al., "VizChat: Enhancing Learning Analytics Dashboards with Contextualised Explanations Using Multimodal Generative AI Chatbots," in Artificial Intelligence in Education, A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, and I. I. Bittencourt, Eds., Cham: Springer Nature Switzerland, 2024, pp. 180–193. doi: 10.1007/978-3-031-64299-9_13.

TeamTeachingViz: A Multimodal Teaching Analytics Dashboard to Support Team Teaching Reflection

Riordan Alfredo[1]*, Paola Mejia-Domenzain[2]*, Vanessa Echeverria[1,3], Dwi Rahayu[1], Linxuan Zhao[1], Haya Alajlan[1], Zachari Swiecki[1], Tanja Käser[2], Dragan Gašević[1], Roberto Martinez-Maldonado[1]

[1]Monash University, Australia; [2]EPFL, Switzerland; [3]Escuela Superior Politécnica del Litoral, Ecuador

**Both authors contributed equally to this research.*

riordan.alfredo@monash.edu, paola.mejia@epfl.ch, vanessa.echeverria@monash.edu, dwi.rahayu@monash.edu, linxuan.zhao@monash.edu, haya.alajlan@monash.edu, zach.swiecki@monash.edu, tanja.kaeser@epfl.ch, dragan.gasevic@monash.edu, roberto-martinezmaldonado@monash.edu

ABSTRACT: TeamTeachingViz is a multimodal teaching analytics dashboard designed to support reflection on team teaching practices in higher education. Grounded on spatial pedagogy and co-teaching literature, the tool allows educators to filter data by specific time intervals or class topics and visualise three aspects: 1) a dynamic classroom map that shows educators' space usage and audio activity; 2) spatial pedagogy indicators illustrating the distribution of the observed activities and 3) co-teaching strategies employed by each pair of educations. Multimodal data was collected from 12 instructors during 36 in-the-wild STEM university-level team teaching sessions. The data included indoor positioning and voice activity (captured via sensors), as well as spatial pedagogy behaviors (documented through observations). Semi-structured interviews were conducted with the educators to review and discuss their own data using **TeamTeachingViz**. Findings suggest that TeamTeachingViz can promote self-awareness by helping educators recognize patterns in their interactions with colleagues and their use of classroom space. This, in turn, can guide the development of actionable goals for improving collaboration in their teaching practice. This demo will showcase TeamTeachingViz's functionalities and present insights on its benefits, challenges, and ethical considerations in incorporating multimodal data into teaching analytics (https://youtu.be/awem5viCawY).

Keywords: teaching analytics, multimodal learning analytics, co-teaching, teaching reflection, spatial pedagogy, in-the-wild, LA dashboard.

StREAM Insights: Using real-time student engagement data to tackle student inequalities and remove structural barriers to learning

Rachel Maxwell
Kortext
Rachel.Maxwell@kortext.com

Andrew Bates
Kortext
andrewb@kortext.com

ABSTRACT: Launching early in 2025, StREAM Insights provides executive staff within universities with cost-effective insights and visualisations around a proven engagement algorithm that codifies student participation in their learning and enables the targeting of outreach initiatives to improve student support and success. Insights capitalises on the affordances of Kortext Fusion – a SaaS data platform providing flexible, scalable and adaptable end-to-end data analytics capability through accelerated data onboarding and management with advanced tooling that supports data aggregation best practices. Insights empowers digital transformation by unlocking siloed application data with advanced data management, enabling a consistent and more engaging student experience through personalized learning.

Student engagement data is used to provide holistic overviews of student participation with their learning to enable real-time reporting against institutional metrics designed to meet strategic and sector-wide objectives e.g. to support student success and continuation, or work to reduce awarding gaps for groups of students typically under-represented within UK higher education.

The executive leadership **Dashboard**, created for the Beta program, provides a high-level overview and health check of student engagement across the institution to support proactive management of student success initiatives through year-on-year comparisons of enrolment, demographic and student engagement data. The dashboard is a result of collaboration across the education sector, raising the prominence of key insights for senior leaders and providing increased flexibility to support multiple use cases at the individual user level.

Client feedback in the early design phases is also informing the ongoing design of the Reports and Insights screens. **Reports** will offer focused visualisations to address specific use cases that track retention rates or explore student success metrics, creating customisable and shareable reports reflecting institutional goals and personal areas of focus. The powerful analytics tools within the **Insights** screen will identify actionable opportunities based on internal targets and goals, reviewing learning over time and based on institutional thresholds and algorithms.

As part of the Insights **demonstration**, we will share the findings of our Beta program, undertaken with 6 English universities as part of a collaborative and iterative product development cycle that ensures the Insights visualisations provide maximum usefulness to address the latest sector and institutional reporting requirements.

Keywords: access and participation, data visualization, insights, institutional oversight, real-time reporting, student disadvantage, student engagement, student success, structural barriers

Learning Analytics Data Visualization in a Virtual Reality Teacher Training Simulation

Birte Heinemann, Sergej Görzen, Ana Dragoljić, Sven Hansen, Jasmin Hartanto, Lars Florian Meiendresch, Jona Recker, Maurice Schwarze, Marc Troll, Ulrik Schroeder
RWTH Aachen University, Germany
heinemann@cs.rwth-aachen.de

ABSTRACT: This Demo presents a Learning Analytics integration in Teach-R, a virtual reality teacher training application, which offers teacher training students the possibility to train skills in a simulated classroom with the help of a coach who controls the students' behavior. A central part of the training sessions is feedback and discussions with the coach and peers, which could be enriched with learning analytics. To make the feedback particularly illustrative, we experimented with visualizations directly in VR that use different data. Firstly, gaze data can show how the teachers' visual attention was distributed among the virtual students. For this, the students are displayed in different transparency levels; individuals who have never been looked at remain invisible in the feedback visualization, while those who have been looked at a lot are solidly colored. Secondly, position data is used to generate different visualizations, e.g. in group work situations, it is interesting to see whether you have only ever been in one area of the room. For this purpose, we use a heat map on the classroom floor (see Figure 1) and various 3D visualizations to give teachers a feeling for their movements in the classroom.

Keywords: Learning Analytics, Virtual Reality, Classroom Management, Feedback, Teacher Training, Spatial Pedagogy

LINK

https://youtu.be/sJgrIpg2oXs

Figure 1: Teaser showing an exemplary visualization for position data. A heatmap shows how the teacher has moved in the classroom. Left: VR View. Right: Concept Draft.

Demo of ouladFormat: an R Package for Loading and Formatting the Open University Learning Analytics Dataset

Emma Howard
Trinity College Dublin
emhoward@tcd.ie

ABSTRACT: A core criterion of learning analytics research is that it uses data from learners engaged in education systems. However, educational data sets can involve time-consuming preprocessing. The Open University Learning Analytics Dataset (OULAD) features data from 32,593 students from 22 presentations of 7 modules (Kuzilek et al., 2017). Apart from module information, the OULAD includes student assessment, registration, virtual learning environment (VLE) and demographic data. The OULAD is available online under a database schema. The ouladFormat R package (Howard, 2024) loads and formats the OULAD for data analysis. The main function, combined_dataset(), draws on the other functions in the package to return a single formatted data set for analysis. The function is flexible as the user can specify different aspects of the returned data set including: the type of student data to be included (assessment, registration, VLE and demographics), the module of interest, the specific semester weeks of VLE to include, whether the VLE data are returned as views per day, per week or per a predefined activity classification etc. The returned data can then be used for analysis e.g., for investigating similar groups of learners using cluster analysis. This demo demonstrates how to use the ouladFormat R package.

Keywords: OULAD, learning analytics, educational tools, reproducibility, MOOCs

Demonstration Link:
https://drive.google.com/file/d/1r7tOPr8pvkyvk1J24LmvtpJrPDRsXRKw/view?usp=sharing

REFERENCES

Howard, E. (2024). *ouladFormat: Loads and formats the Open University learning analytics dataset for data analysis* (Version 1.1.2) [R package]. CRAN. https://cran.r-project.org/web/packages/ouladFormat/index.html

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. *Scientific Data, 4*, 1-8. https://doi.org/10.1038/sdata.2017.171

Guiding student dialogue using Clair, a collaborative learning agent for interactive reasoning

Adelson de Araujo, Pantelis M. Papadopoulos, Susan McKenney, Ton de Jong
Department of Learning, Data Analytics, and Technology, University of Twente
{a.dearaujo, p.m.papadopoulos, susan.mckenney, a.j.m.dejong}@utwente.nl

ABSTRACT: While teachers can guide small group student dialogue, scaling timely guidance across multiple simultaneous groups is impractical. Collaborative Conversational Agents (CCAs) emerge as a viable solution, for example, by utilizing learning analytics to identify patterns in student dialogue to trigger interventions grounded in dialogic instructional theory. Yet, off-the-shelf CCA implementations are currently limited. For this reason, we recently developed Clair (de Araujo et al., 2024a; de Araujo et al., 2024b), a learning analytics-based CCA integrated in the Go-Lab platform (de Jong et al., 2021). Clair intervenes in student dialogue using reflective prompts (aka 'talk moves'), deeply rooted in learning sciences. Clair's content-independent intervention strategy, based on the Academically Productive Talk (APT) framework, also known as Accountable Talk (Michaels et al., 2016), is easily customizable to various collaboration settings, from secondary to higher education, and can facilitate student dialogue in various learning contexts with little configuration effort. Unlike single-user conversational agents, Clair does not react to all messages, which would distract group interaction. Instead, Clair reacts to targeted situations in which intervening is theorized as fostering productive interactions. Clair is capable of handling student dialogues in various languages, and it has been evaluated so far in schools in Brazil, the Netherlands, Germany, Finland, and Taiwan.

Keywords: collaborative learning, conversational agents, academically productive talk

VIDEO

https://youtu.be/5IAy9UdcgsQ

REFERENCES

de Araujo, A., Papadopoulos, P. M., McKenney, S., & de Jong, T. (2024a). A learning analytics-based collaborative conversational agent to foster productive dialogue in inquiry learning. Journal of Computer Assisted Learning, 40(6). https://doi.org/https://doi.org/10.1111/jcal.13007

de Araujo, A., Martens, J., & Papadopoulos, P. M. (2024b). Enhancing student dialogue productivity with learning analytics and fuzzy rules. In International Conference on Artificial Intelligence in Education (AIED'24). Springer, Cham. https://doi.org/10.1007/978-3-031-64299-9_36

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., …, & Zacharia, Z. C. (2021). Understanding teacher design practices for digital inquiry–based science learning: the case of Go-Lab. Educational Technology Research and Development, 69(2), 417–444.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B. (2016). Accountable Talk: Classroom conversation that works. University of Pittsburgh.

From One-Way Text to Socratic Dialogue: Designing an Artificial Intelligence Chatbot to Support Student Learning in Higher Education

Isabel Hilliger, Carlos González
Pontificia Universidad Católica de Chile
{ihillige, cgonzalu}@uc.cl

Mar Pérez-Sanagustín, Gabriel Ferrettini, Esteban Villalobos
Université Toulouse III - Institut de Recherche en Informatique de Toulouse
{mar.perez-sanagustin, Gabriel.Ferrettini, esteban.villalobos}@irit.fr

Andre Nascimento
Universidade Federal Rural de Pernambuco
andre.camara@ufrpe.br

Sergio Celis
Universidad de Chile
scelis@uchile.cl

ABSTRACT: Artificial intelligence (AI) chatbots, as a form of generative AI, hold significant promise for higher education. However, there is a notable gap between what these chatbots can offer and what instructors need to foster deeper learning in their subjects. To bridge this gap, we developed an AI chatbot designed to function as a Socratic tutor, supporting student learning through guided questioning. Initially conceived as a text messaging tool, this chatbot evolved through a design-based research approach into a Streamlit web-based application. Unlike ChatGPT, where students start from an empty prompt, our Socratic tutor begins with a learning scenario defined by the instructor. Additionally, the platform provides real-time analytics of student interactions, enabling instructors to monitor and adjust their teaching strategies promptly. So far, this chatbot has undergone four iterations and has been implemented in 10 subjects, involving a total of 277 students. To support various learning activities, it leverages GPT-4 Turbo and course materials, such as the course syllabus and readings. Feedback from different stakeholders has been instrumental in refining the tool. Insights gathered from three workshops highlight its potential to integrate interactive exercises into various classroom settings. This demo aims to present the design journey and current applications of our Socratic tutor, contributing to the broader discourse on the integration of generative AI in educational contexts.

Keywords: Generative Artificial Intelligence, Artificial Intelligence chatbot, Higher Education, Socratic tutoring, Critical thinking

Video:https://www.canva.com/design/DAGVKZDFp2c/YsEUdnWu9N-bjj6SB1BeqQ/watch?utm_content=DAGVKZDFp2c&utm_campaign=designshare&utm_medium=link2&utm_source=uniquelinks&utlId=h59a231a471

eDoer: AI-driven Rapid Curriculum Design and Learning Recommendations

Authors: Dr. Gábor Kismihók
TIB Leibniz Information Centre for Science and Technology
Gabor.Kismihok@tib.eu

Dr. Mohammadreza Tavakoli
TIB Leibniz Information Centre for Science and Technology
Reza.Tavakoli@tib.eu

Abdolali Faraji
TIB Leibniz Information Centre for Science and Technology
Abdolali.Faraji@tib.eu

Mohammadreza Molavi
TIB Leibniz Information Centre for Science and Technology
Mohammadreza.Molavi @tib.eu

Mohammad Moein
TIB Leibniz Information Centre for Science and Technology
Mohammad.Moein@tib.eu

ABSTRACT: In this demonstration, we showcase the AI-driven learning platform eDoer (https://edoer.eu/). eDoer supports open education by empowering learners through individual learning recommendations, on the basis of quality assured curricula. These curricula are built on openly available learning resources, as follows: 1) eDoer utilizes a Large Language Model (chatGPT) based curriculum development module to help teachers to create and update their curricula. 2) Teachers populate their curricula with openly available, quality controlled learning content. They can use eDoer's AI based recommender, which supplies them with relevant learning content based on their topic and learning objectives. 3) Teachers can generate and validate assessments with AI for their courses in their curricula. 4) Teachers publish their curricula for their learners, including open learning content and assessments. 5) Learners use these curricula to learn and develop themselves towards the learning objectives. They can receive feedback on their progress and learning content recommendations based on their learning preferences and previous learning history. 6) Teachers can create or tailor existing and suitable didactical methods, which exploit the benefits eDoer. These methods include means of reflection, problem oriented discussions, and personalised feedback. eDoer is multilingual, and it is available openly and without any restrictions as Web, Android and IoS applications. Video link: https://www.youtube.com/watch?v=DOKu7tGlzS8

Keywords: Artificial Intelligence, learning recommendation, recommender system, open education

edX-LIMS: System for Learning Intervention and its Monitoring for edX MOOCs

Ruth Cobos
Universidad Autónoma de Madrid
Ruth.Cobos@uam.es

Álvaro Becerra
Universidad Autónoma de Madrid
Alvaro.Becerra@uam.es

ABSTRACT: edX-LIMS (System for Learning Interventions and Its Monitoring for edX MOOCs) is a Learning Analytics system developed by the GHIA (Group for Advanced Interactive Tools) research group at the Universidad Autónoma de Madrid (UAM) in Spain. This system has been used for five years with a MOOC from UAM on edX. The system provides instructors with valuable insights through a dashboard, including real-time feedback on learner engagement, progress, and performance. Additionally, it can detect self-regulation challenges or difficulties encountered by learners, offering instructors real-time predictions on the likelihood of learners dropping out or successfully completing the MOOC. This information enables instructors to make data-driven decisions and intervene when necessary. Moreover, the system empowers learners by allowing them to visualize their own analytics within the course through a dashboard. edX-LIMS implements an intervention strategy that is activated weekly. The evaluation with learners and teaching staff yielded several interesting findings. First, both MOOC learners and instructors reported feeling more connected to one another. Second, instructors successfully implemented intervention strategies tailored to the specific needs of their learners. Finally, learners expressed increased motivation to continue with the course, which resulted in a decrease in dropout rates.

Keywords: Dashboard, Data-Driven Intervention, Learning Analytics, Feedback, MOOC, Self-Regulated Learning, Prediction.

DEMO VIDEO LINK: HTTPS://YOUTU.BE/PUAVRMTXDKK

REFERENCES

Topali, P., Cobos, R., Agirre-Uribarren, U., Martínez-Monés, A., & Villagrá-Sobrino, S. (2024). 'Instructor in action': Co-design and evaluation of human-centred LA-informed feedback in MOOCs. *Journal of Computer Assisted Learning*. 40(6), 3149–3166. https://doi.org/10.1111/jcal.13057

Cobos, R. (2023). Self-Regulated Learning and Active Feedback of MOOC Learners Supported by the Intervention Strategy of a Learning Analytics System. *Electronics*, 12(15), 3368. https://doi.org/10.3390/electronics12153368

Pascual, I., & Cobos, R. (2022). A proposal for predicting and intervening on MOOC learners' performance in real time. En A. Vázquez-Ingelmo, Y. Dimitriadis, A. Martínez-Monés, & F. J. García-Peñalvo (Eds.), *CEUR Conference Proceedings* (Vol. 3238, pp. 26–38). LASI-Spain 2022, Salamanca, Spain, June 20-21, 2022. CEUR-WS. https://ceur-ws.org/Vol-3238/paper4.pdf

ACKNOWLEDGEMENTS

This work has been supported by projects: HumanCAIC (TED2021-131787B-I00 MICINN) and SNOLA (RED2022-134284-T).

19squared: A Collaborative Authoring Tool with Learning Analytics for Immersive 360° Educational Experiences

Matthias Ehlenz
RWTH Aachen University
ehlenz@lbz.rwth-aachen.de

Frederic Maquet
RWTH Aachen University
maquet@lbz.rwth-aachen.de

Birte Heinemann
RWTH Aachen University
heinemann@cs.rwth-aachen.de

ABSTRACT: In this demonstration, we present an authoring tool called 19squared for immersive educational experiences based on 360° media. This web-based tool, unlike many existing ones, is implemented with education as the primary focus and comes with several built-in learning analytics features. 19squared is intended not only to be used by teachers to create content but also to focus on the constructive process, allowing students to collaborate on (own) projects. We have implemented appropriate user-management, real-time collaboration features, an easy-to-use web interface, and xAPI data collection for user interactions to achieve this. In the long term, we hope to gain insights e.g. into learning processes involved in creating immersive experiences or the cognitive processes in exploring 360° environments. Beyond the research perspective, our data collection can provide added value to educators by monitoring students' progress and gaining a deeper understanding of collaborative design processes, especially when individual feedback (and grading) is required. Several dashboards and visualizations were implemented and bundled for this purpose.

19squared has been used in teacher training, school workshops and thesis projects, ranging from the humanities to STE(A)M subjects and vocational training. We are committed to open source and open science and release our work as such.

Keywords: Immersive learning, learning analytics, collaborative learning, authoring tool

1 LINKS TO ONLINE RESOURCES

The video is to be found at https://www.youtube.com/watch?v=ZCX-NqcSRrA

A public instance is hosted at https://19squared.de/

The source code is available at https://git.rwth-aachen.de/medialab/interactive360vr

PeerGrader: A Peer Assessment System to Enhance Student Agency and Self-regulated Learning

Author: Zexuan Chen
Institution: School of Foreign Studies, Southern Medical University (China)
Email: SerlinaChen@163.com

Author: Bart Rienties
Institution: Institute of Educational Technology, The Open University (UK)
Email: bart.rienties@open.ac.uk

Author: Simon Cross
Institution: Institute of Educational Technology, The Open University (UK)
Email: simon.j.cross@open.ac.uk

ABSTRACT: Based on learning analytics in peer assessment, PeerGrader was designed and developed to enhance student agency and self-regulated learning in peer assessment. This demonstration will showcase its basic functions, customized features, and iterations. PeerGrader serves two key stakeholders: students and instructors. It facilitates a seamless workflow for students, allowing them to submit initial drafts, engage in peer assessments, refine their drafts based on peer feedback. In addition, PeerGrader features customized tools to enrich the peer assessment experience. Specifically, student assessors enjoy GenAI-facilitated feedback giving, learning analytics dashboard, self-decided workloads, self-selected proficiency pairing, dynamic tracking of assessing record, scoring moderation, multi-edit capabilities for qualitative and quantitative feedback; meanwhile, student assessors can switch to the role of assessees to review feedback from peers at any time, enhancing their learning experience. For instructors, PeerGrader streamlines task setting, peer assessment configuration, task release, and monitoring of activity progress. Furthermore, the instructor view provides detailed pairing information between assessors and assessees, along with both qualitative and quantitative feedback given and received. PeerGrader has undergone three rounds of iteration over three semesters, involving 636 undergraduates from three universities, refining its functionality based on user feedback. We would invite discussion on learning analytics of self-regulated learning in peer assessment based on the data collected by PeerGrader.

Keywords: peer assessment; PeerGrader; student agency; self-regulated learning

A Multimodal Teamwork Analytics Dashboard for Supporting Awareness and Reflection in Small Group Collaboration

Rogers Kaliisa, Jan Arild Dolonen, Crina Damsa, Rachelle Esterhazy, Daniel Sanchez, Krishna Siddharth Rajasekaran

rogers.kaliisa@iped.uio.no; j.a.dolonen@iped.uio.no; crina.damsa@iped.uio.no
University of Oslo, Norway

ABSTRACT

During small group collaboration, students work with their peers to solve problems that may be too complex to tackle individually. However, students engaged in collaborative learning, especially in extended group projects, may encounter challenges related to the management of their collaboration, motivational and emotional involvement. Therefore, they often need support to identify and manage these challenges productively. Leveraging advances in generative artificial intelligence, our demonstration presents an AI-supported Multimodal Learning Analytics (MMLA) dashboard designed to enhance students' awareness and reflection on group dynamics in collaborative, authentic learning environments. The dashboard features three core analytics components: (1) **Meeting Analytics**—analyzes content from group meetings recorded on Microsoft Teams, with insights into action items, speaking time, and turns of talk; (2) **Collaboration Experience**—offers scores based on students' self-reported experiences of group dynamics, providing a group and self-reflection tool; and (3) **Document Analytics**—summarizes and reviews collaborative documents to track group documents' progress and content alignment with course requirements. This dashboard was developed through a design-based research approach and multiple iterations, involving work by researchers and technology developers and feedback from students and teachers. It has been field-tested in a master's level course in Legal Education with 28 students across 8 groups. Data-guided reflection sessions and interviews with students indicate that students find multimodal data visualizations beneficial for raising awareness about group dynamics and with potential to reveal issues within the group that the individual students cannot easily express. The Demo can be found here.

Keywords: Multimodal learning analytics tool, small group collaboration

ACKNOWLEDGMENTS

"The development of this dashboard was supported by the Norwegian Research Council under the TEAMLEARN Project, at the University of Oslo, Norway."

SmartQuiz: Automated Generation of Multiple-Choice Questions from Educational Videos

Giacomo Cassano
Politecnico di Milano
giacomo.cassano@polimi.it

Daniel Ebbert
University of South Australia
daniel.ebbert@mymail.unisa.edu.au

Nicoletta Di Blas
Politecnico di Milano
nicoletta.diblas@polimi.it

Negin Mirriahi
University of South Australia
negin.mirriahi@unisa.edu.au

ABSTRACT: Creating high-quality multiple-choice questions (MCQs) for educational videos is a time-consuming task for educators, limiting the time they can dedicate to student support. This demo introduces *SmartQuiz*, an AI-powered framework that automatically generates pedagogically relevant MCQs from video transcripts. The SmartQuiz pipeline segments the video transcript into topics and identifies the concepts explained within each topic. These concepts are classified into three categories: essential (core ideas critical for understanding the main message), supportive (e.g., examples), and organizational (e.g., outlines). For each essential concept, the corresponding transcript text is refined to be self-contained, and an MCQ is generated. SmartQuiz ensures the quality of its MCQs by adhering to established item-writing guidelines, thereby mitigating common flaws associated with automated question generation. The demo illustrates the generation of 12 distinct MCQs from a 9-minute educational video on *Introduction to Machine Learning*. Two ongoing studies are evaluating the impact of these MCQs on student learning and collect lecturer feedback on question quality. While the current implementation focuses exclusively on audio transcripts, excluding visual elements, SmartQuiz demonstrates the potential to significantly reduce the workload of MCQ creation while maintaining pedagogical integrity.

Keywords: Question Generation, Large Language Models, Artificial Intelligence, Educational Videos, Automated Assessment, Video-Based Learning

1 VIDEO

https://www.youtube.com/watch?v=owEtKGQxTm4

Interactive Long-Term Study Planning for Individual Student Support, Leveraging Process Mining and AI

Johannes Wagner-Schirrmeister [a], René Röpke [b], Tobias Johnen [c], Rike Carpantier [a], Ulrik Schroeder [c], Maren Scheffel [a]

[a] Ruhr University Bochum
{johannes.wagner-schirrmeister, rike.carpantier, maren.scheffel}@rub.de

[b] TU Wien
rene.roepke@tuwien.ac.at

[c] RWTH Aachen University
{johnen, schroeder}@cs.rwth-aachen.de

ABSTRACT: This demo showcases an AI- and Process Mining-powered interactive study planning tool designed to assist students in higher education with complex personalized, data-driven feedback. The tool integrates rule-based artificial intelligence (AI) and process mining to provide real-time, context-aware guidance, enabling students to visualize study program structures and receive automated feedback on plan validity. By leveraging historical student data and curriculum rules, the tool offers personalized recommendations, empowering students to make informed, autonomous decisions about their study paths. Developed through an iterative human-centered design process, the tool addresses the complexities of long-term academic planning while balancing guidance with student autonomy. Through ongoing user testing and stakeholder engagement, the tool is continually adapted to the requirements, feedback, and insights of users, stakeholders, and researchers. The demo will focus on the tool's interface and feedback mechanisms, highlighting its potential to support dynamic long-term study planning, especially regarding deviations from recommended study plans. With our tool, we address the complexities of study planning and support, giving students the tools to make informed decisions and offering promising and more successful study paths. In doing so, we aim to support students in managing the challenge of comprehensive long-term study planning successfully, leading to better student outcomes as preliminary results indicate.

Keywords: Study Planning, Feedback, Artificial Intelligence, Process Mining, Evaluation

1 DEMO VIDEO

https://doi.org/10.13154/294-12108

2 ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research.

Participatory Causal Modelling of Learning Systems

Ben Hicks
University of Technology Sydney
Ben.Hicks@student.uts.edu.au

ABSTRACT: Learning Analytics aims to improve the learning process. This necessitates a *causal* interpretation of observational data. One way to model causal structure is by using causal Directed Acyclic Graphs. The visual formalism of the model requires little technical knowledge to engage with, providing an opportunity for non-technical experts to remain engaged deep into the crafting of critical statistical assumptions about the learning system, including the importance of latent variables. My research will apply these models to several potential cases, including equitable learning outcomes and student support. The models will be co-constructed between stakeholders with a wide range of expertise, using the visual formalism to help foster a shared understanding. Key decisions I am considering concern the evaluation of how this process influences participants' thinking about the system as well as how best to engage with the range of stakeholders required to facilitate the modelling.

Keywords: causal models, DAGs, participatory design, human centered LA

1 BACKGROUND

Causal claims in education have traditionally been warranted using the 'gold standard' of Randomized Control Trials (RCTs), however this is not always feasible, ethical, nor possibly desirable if we want to understand causal effects outside of a controlled experiment (Sullivan, 2011). As such, researchers and practitioners in Learning Analytics (LA) are generally hesitant to make causal claims from observational data (Viberg et al., 2018). This presents a challenge: LA products are built upon observational data (e.g. dashboards, automated feedback systems) and for them to be actionable a causal claim from this observational data is required.

LA inherits this wariness towards causal claims from a general historical reluctance to making causal claims outside of an RCT (Pearl & Mackenzie, 2018). Making a causal claim requires assumptions that come from outside the data, from our contextual knowledge of how the data came to be (Robins & Wasserman, 1999). Strong theory provides one pathway to providing this knowledge (Borsboom et al., 2021). There have been concerted efforts in the field of LA to incorporate learning theories but, in part due to the many viewpoints that offer perspectives on learning, the field lacks a single definition of what constitutes a learning theory (Khalil et al., 2022). Another way to incorporate contextual knowledge is through domain experts. A variety of techniques, under the umbrella of Participatory Design and Co-Design methods, attempt to address this through engaging stakeholders at various stages of the design process. These techniques have been growing in popularity in LA (Sarmiento & Wise, 2022) as have calls for their use (Dollinger & Lodge, 2018).

These approaches all face a translation problem, as the contextual knowledge needs to be formalized and structured enough to be executable in code. As it stands, existing participatory methods in LA do not directly inform the abstract data models, leaving considerable work for the data expert to do. These issues combine to hinder power balance and value alignment between participants (Dollinger & Lodge, 2018), and the accessibility of the process to all participants (Vezzoli et al., 2020). To truly participate all parties must be able to interrogate the system being designed (Kitto et al., 2020), a problem that is exacerbated in contexts that require technical knowledge (Dollinger & Lodge, 2018).

2 CAUSAL MODELS

One way that other fields have addressed the problem of making causal claims from observational data is through Structural Causal Models (Pearl, 2009). These models can be represented using a Directed Acyclic Graph and are often referred to simply as DAGs. The causal DAG consists of nodes, representing variables, and directed edges (arrows), representing the flow of causation. We represent the relationship "A causes B" as A → B. These graphical models can then be used to identify an appropriate statistical method to make causal claim from observational data.

In collaboration with others, I have published initial work in exploring the possible affordances of causal DAGs in the field of education. Hicks et al. (2022) used the modelling of a student at-risk system to examine how we think about a system and introduced causal DAGs to the LA community. Weidlich et al. (2023) demonstrated how these graphical tools address bias within a system and compared these methods with Structural Equation Models, a similar approach that traditionally avoids making causal claims. In Kitto et al. (2023) we examined how causal DAGs, and the underlying mathematics that relates the causal DAG to conditional independence relationships, might help bridge the gap between big data and the theories of the learning sciences. This closely aligns with how an individual might develop their own understanding of a system. In Hicks et al. (2023) we present causal DAGs as a possible way to think clearly about the interplay between learning and learning outcomes, and the effects of intervening in learning systems.

There is an opportunity for these tools to solve two problems at once: (i) making causal claims from observational data (either in estimating causal effects or evaluating theories), and (ii) allowing non-technical stakeholders to have a greater say in the crafting of key statistical assumptions in the building of LA products and highlighting the data that is required. There has been little work to date in the use of causal models in education to address the first problem. Boerebach et al. (2013) provide an example of using multiple DAGs to compare competing theories. Another line of work, beginning with a master's thesis by Brokenshire (2007), used causal discovery methods to formalize Self-Regulated Learning theories and argued that causal models could help LA practitioners intervene in learning systems (Kumar et al., 2015). Historically, the broader adoption of these methods has likely been hindered by a lack of expertise, a lack of data, and lack of tools – issues that are less problematic in the field today (Kitto et al., 2023).

3 RESEARCH QUESTIONS

The goal of my research project is to help bring stakeholders traditionally excluded from the creation of important analytical assumptions closer to the mathematical machinery underpinning LA products, enabling them to be more meaningfully involved in their creation. This will be done by developing an approach to co-construct causal DAGs with non-technical experts. This research aims to answer:

1. How effectively can non-technical experts co-create a causal model of a learning system?

2. Does the process of co-creating a causal model help develop thinking about a learning system?

Research Question 1 (RQ1) relates to the affordances of the *causal model as a product*, whereas Research Question 2 (RQ2) relates to the potential value inherent in the *process of causal modelling*.

4 PILOT STUDY

Four collaborative modeling sessions have been run, where an expert in the relevant domain attempted to represent their system in a causal DAG, facilitated by a causal modeler. Two of the sessions included different researchers modeling student belonging, one a researcher modeling at-risk students, and the last an expert modelling equitable learning outcomes. The dialogue from the at-risk modeling sessions was synthesized into Hicks et al. (2022). The two models created on Belonging are shown below in Figure 1, and these preliminary models are feeding into further research.

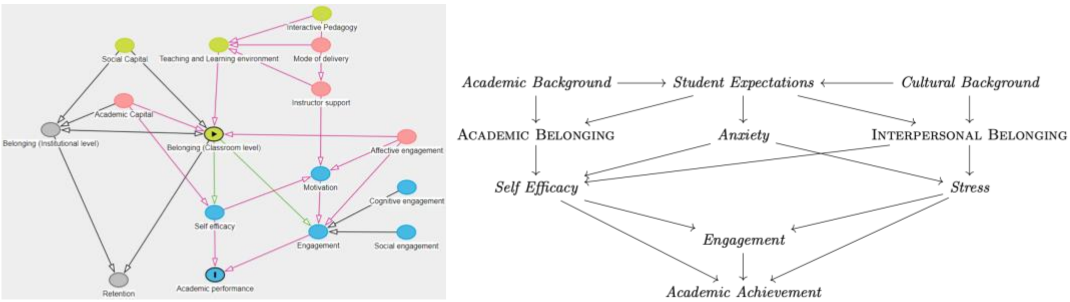


Figure 1: Two causal DAGs from participatory modelling sessions with two researchers about the claims they see in the literature on student sense of belonging.

From these sessions a series of dialogue 'prompts' was generated to help facilitate future sessions and to articulate what kind of thinking moves are made during this process. These prompts articulate potential modeling moves given the current state of the DAG, matching parts of the DAG with questions for the modeler to ask of the model. One such example is below, in Figure 2.

Beginning with any two causally connected nodes in the graph: $X \longrightarrow Y$

*Does X influence Y **directly**, or is there some other variable (M) in between?* $X \longrightarrow M \longrightarrow Y$

(M is called a *mediator* which we are inserting into the path $X \to Y$)

Are there other things (M) that X changes that in turn change Y?

(Adds a new path with a mediator in it)

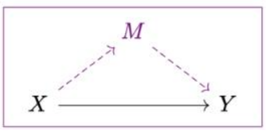


Figure 2: A prompt for 'adding a causal chain' to a causal DAG, to help translate from the visual formalism to a dialogue between the participant and the causal modelling facilitator. Each prompt includes an initial subgraph, questions to interrogate the DAG, and details on changes to the model based on possible answers to the question.

These prompts are being used to help facilitate future sessions and codify the kinds of thinking moves that happen during the collaborative design of a causal DAG. It was noted by participants (both the domain expert and the modeler) that it seemed that there was something of value in the *process* of drawing a causal DAG. It is worth noting that in each modelling session latent variables were highlighted by the experts as important to the system.

5 RESEARCH METHODOLOGY & QUESTIONS FOR THE DC

Causal models of a variety of learning systems will be sketched out in collaborative design sessions with a range of expertise within each system. The sessions will involve a causal modelling facilitator (myself) and at least one expert in the context of the system. Prompts (such as Figure 2 above) will be designed to support the facilitator, based on prior sessions and guided by the structure of *the powerful questions* from Culmsee and Awati (2014). Three potential cases have been selected so far: student support systems, evaluation of key influences and measurement of finishing a program of education 'well', and understanding the key drivers of equitable learning outcomes. The resulting models have potential applications informing LA design of predictive models, qualitative analysis and dashboards. For each case study several design sessions will be run independently with stakeholders with varying levels of expertise or different backgrounds in how they understand the system. I will identify experts on the system to participate by considering their knowledge of the system, awareness of issues, and availability to participate. A participant's knowledge of the system may be tacit, such as the case of a student understanding a student support system, or more theoretical, as in the case of a researcher. Throughout the session the facilitator will ask the participants to explain their thinking and strategies during the model development. After the session a short semi-structured interview will record the participants' reflections on the process. The models will then be compared and contrasted, and records (images taken of the models) of the sessions analyzed for key moments in the co-creation of the model.

Finding participants with a range of system perspectives for the case studies is something I would like guidance on at the DC. Another aspect I would like to discuss is in the **evaluation of how the process of co-creating causal models influences thinking about the learning system**, so I will now detail that component of the research. Currently I am proposing coding of the narrative self-report at the end of the session alongside the images of the model development. I am contemplating coding the participants' understandings of the system against the Theory Construction Methodology (TCM) framework (Borsboom et al. 2021). TCM outlines a formalization process from 'proto-theory' to 'formal model'. This formalization might manifest graphically, as a semi-formal causal graph (only the 'G' from DAG) that increasingly adheres to the rules of a causal DAG. It might also manifest through dialogue and questioning, as the participant comes up with newly formed questions to check explainable adequacy of their model (step four of TCM) or ideas to evaluate the worth of the theory (step five). These instances of possible formalization in thinking will be examined from the graphical and audio transcripts.

5.1 Coding of the participatory modelling sessions

Each session the evolution of the causal DAG construction (either video or photos / screenshots) will be recorded, and in small groups the audio as well. For smaller groups, where the audio transcript is available, these will be matched to the evolving DAG and examined for instances of:

- Dialogue between the participants and the modeler influencing the DAG, or indicating surprise, or a more abstract level of thinking about the system.

- Structure in the DAG (such as paths, absences of edges) influencing dialogue. This may be directly noticed by the participant or prompted by the facilitator (see Figure 2 for an example).

5.2 Narrative self-report

At the end of the session (for small groups) or in a follow up interview (for larger sessions) participants will be asked to reflect on the process itself, and their own thinking. This will be open-ended, but I plan that it should include two questions: *(1) Did the process itself offer new insights, or was it merely a (potentially new) way to synthesize your current knowledge?; and (2) Did the process in principle fail to capture something important about how you understand the system (such as due to the visual medium or the structural constraints)?* Participants will also be asked to reflect on the challenging points and decisions made during the model development. Images of the causal graph development, at various stages, will be used as a prompt for this. **The challenge here is in developing a plan to link the data from these reflective questions to session transcript data (showing the model construction through dialogue and diagrams) in a way that can capture the evolution of the participants thinking about the system.**

6 CONCLUSION

Participatory and Co-Design methods seldom bring the non-technical expert so close to the critical statistical assumptions as collaborating on drawing a causal DAG seems to. This will hopefully keep non-technical stakeholders at the LA design table for longer and help highlight what data might be missing and how important it is (Wise et al., 2022). LA products designed with the assumptions built in for making causal claims from observational data will be on more rigorous ground for making that data actionable. Additionally, the causal DAG may then be utilized for further analysis, such as making stronger causal claims (Weidlich et al., 2022) or testing theories (Kitto et al., 2023). There may also be a secondary benefit in the causal modeling process in how it pushes participants to think in a more structured way along with a visual aid. I believe this last potential benefit of the research is the most challenging aspect to understand and evaluate and what I propose to focus on at the Doctoral Consortium.

REFERENCES

Boerebach, B. C., Lombarts, K. M., Scherpbier, A. J., & Arah, O. A. (2013). The teacher, the physician and the person: Exploring causal connections between teaching performance and role model types using directed acyclic graphs. *PLoS One*, **8**(7), e69449. https://doi.org/10.1371/journal.pone.0069449

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science,* **16**(4), 756–766. https://doi.org/10.1177/1745691620969647

Brokenshire, D. (2007). *Discovering causal models of self-regulated learning* (Masters thesis). Simon Fraser University. https://central.bac-lac.gc.ca/.item?id=MR41042&op=pdf&app=Library&oclc_number=667804408

Culmsee, P., & Awati, K. (2014). The map and the territory: A practitioner perspective on knowledge cartography. In *Knowledge Cartography: Software Tools and Mapping Techniques* (pp. 261-292). London: Springer London.

Dollinger, M., & Lodge, J. M. (2018, March). Co-creation strategies for learning analytics. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 97-101).

Hicks, B., Kitto, K., Payne, L., & Buckingham Shum, S. (2022). Thinking with causal models: A visual formalism for collaboratively crafting assumptions. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 250–259). New York, NY: Association for Computing Machinery (ACM). https://doi.org/10.1145/3506860.3506899

Hicks, B., Weidlich, J., Kitto, K., Buckingham Shum, S., Drachsler, H. (2023). Causation and the Interplay Between Learning Outcomes and Learning Interventions. In "Aligning the Goals of Learning Analytics With Its Research Scholarship: An Open Peer Commentary Approach". *Journal of Learning Analytics* 10 (2):14-50. https://doi.org/10.18608/jla.2023.8197

Kitto, K., Gulson, K., Thompson, G., & Payne, L. (2020, July). Technical democracy in educational decision making. In *Fairness, Accountability, and Transparency in Educational Data (FATED): A 1-day workshop co-located with the EDM 2020 conference.*

Kitto, K., Hicks, B., & Buckingham Shum, S. (2023). Using causal models to bridge the divide between big data and educational theory. *British Journal of Educational Technology.*

Kumar, V. S., Clemens, C., & Harris, S. (2015). Causal models and big data learning analytics. In *Ubiquitous learning environments and technologies* (pp. 31–53). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44659-1_3

Pearl, J. (2009). *Causality*. Cambridge University Press. https://doi.org/10.1017/cbo9780511803161

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Robins, J. M., & Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. *Computation, causation, and discovery*, 1999, 305-21.

Sarmiento, J. P., & Wise, A. F. (2022, March). Participatory and co-design of learning analytics: An initial review of the literature. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 535-541).

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education*, **3**(3), 285–289. https://doi.org/10.4300/JGME-D-11-00147.1

Weidlich, J., Gašević, D., & Drachsler, H. (2022). Causal inference and bias in learning analytics: A primer on pitfalls using directed acyclic graphs. *Journal of Learning Analytics*, **9**(3), 183–199. https://doi.org/10.18608/jla.2022.7577

Weidlich, J., Hicks, B., & Drachsler, H. (2023). Causal reasoning with causal graphs in educational technology research. *Educational technology research and development*, 1-19.

Wise, A. F., Sarmiento, J. P., & Boothe Jr, M. (2021, April). Subversive learning analytics. In *LAK21: 11th international learning analytics and knowledge conference* (pp. 639-645).

Vezzoli, Y., Mavrikis, M., & Vasalou, A. (2020, March). Inspiration cards workshops with primary teachers in the early co-design stages of learning analytics. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 73-82).

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. The current landscape of learning analytics in higher education. *Computers in human behavior*, 89:98–110, 2018.

Adaptive Feedback in Learning Environments: A Multimodal Approach to Enhancing Feedback Sensitivity and Learner Engagement

AboulHassane CISSE
sn22869p@st.omu.ac.jp
Osaka Metropolitan University
Sakai, Osaka, Japan

Abstract

This research explores the application of multimodal data, including eye-tracking, heart rate variability, and emotion recognition, to deliver real-time adaptive feedback in learning environments. Specifically, the study investigates how this feedback influences learner engagement, task performance, and persistence during English language conversations with a conversational agent in a simulated restaurant scenario. By integrating Affective Backchannels (AB), Conversational Strategies (CS), and their combination (AB+CS), the system provides personalized feedback tailored to both emotional and cognitive states. Preliminary results from 9 participants reveal that real-time feedback effectively reduces frustration—evidenced by lower heart rates and more positive emotional expressions while significantly improving task accuracy and Willingness to Communicate (WtC). This research contributes to learning analytics and adaptive learning technologies by demonstrating how multimodal data can enhance cognitive and emotional learning outcomes. Future work will focus on expanding the dataset, refining individual physiological baselines, and exploring scalability across diverse educational settings, including more emotionally complex scenarios.

Keywords

Adaptive learning, multimodal feedback, learner engagement, emotion recognition, personalized feedback.

ACM Reference Format:
AboulHassane CISSE. 2025. Adaptive Feedback in Learning Environments: A Multimodal Approach to Enhancing Feedback Sensitivity and Learner Engagement. In *Proceedings of (Companion Proceedings 15th International Conference on Learning Analytics & Knowledge (LAK25)).* ACM, Dublin, DUB, Ireland, 6 pages.

1 INTRODUCTION

Adaptive learning environments are designed to personalize instruction to meet individual learner needs, thereby enhancing engagement and improving outcomes. Feedback plays a pivotal role in boosting motivation and performance, particularly when delivered effectively ([Hattie and Timperley(2007)]; [Shute(2008)]). However, learners often experience cognitive overload or disengagement during complex tasks ([Sweller et al.(2011)]). Adaptive feedback, informed by both emotional and cognitive cues, has the potential to address these challenges by offering personalized, real-time support ([Azevedo and Aleven(2013)]).

While most adaptive systems emphasize cognitive performance, they frequently neglect the influence of emotional states on learning ([D'Mello and Graesser(2012)]). This study seeks to bridge that gap by leveraging multimodal data such as eye-tracking, heart rate, and emotion recognition to deliver adaptive feedback that enhances learner engagement and persistence in complex tasks. The novelty of this research lies in integrating real-time cognitive and emotional data, extending prior work on conversational dynamics with the inclusion of emotional and physiological cues ([Ayedoun et al.(2016)]; [Picard(1997)]).

In the domain of second language acquisition (SLA), Willingness to Communicate (WtC) is a critical factor in determining learners' ability to use the language effectively in real-world scenarios. Higher WtC has been linked to increased confidence, improved social interaction skills, and expanded professional opportunities in a globalized workforce ([MacIntyre et al.(1998)]). However, many learners face persistent barriers to communication due to anxiety, insufficient practice, or cultural differences.

Traditional language learning methods often fail to address these challenges, as they lack real-time, personalized feedback and opportunities for realistic dialogue. To fill this gap, advanced conversational agents have emerged as promising tools, simulating real-world interactions and providing immediate, context-sensitive feedback to help learners build confidence and communication skills.

This study explores the use of a conversational agent in a low-pressure restaurant scenario, which serves as an ideal starting point for enhancing WtC ([CISSE et al.(2024)]). Restaurant conversations are practical, familiar, and allow learners to practice structured dialogue in a controlled yet realistic setting. This environment minimizes anxiety and facilitates the development of conversational fluency. Additionally, the predictable nature of restaurant interactions makes them well-suited for personalized feedback and skill-building.

While the restaurant scenario provides an effective foundation, it has its limitations. It may be less applicable in high-stakes or emotionally charged settings, such as job interviews or public speaking events, where interactions often involve complex cultural norms

Companion Proceedings 15th International Conference on Learning Analytics & Knowledge (LAK25),
,

and emotional intelligence that are difficult to replicate computationally. Future iterations of this system could address these limitations by incorporating advanced affective computing capabilities, expanding the system's adaptability to a wider range of contexts.

2 LITERATURE REVIEW

2.1 Adaptive Learning Technologies

Adaptive learning technologies aim to personalize instruction by tailoring strategies to individual learner behavior, thereby enhancing engagement and improving outcomes. [Kulik and Fletcher(2016)] found that adaptive systems improve learning efficiency by personalizing the pace of instruction and providing targeted support. Similarly, [Desmarais and Baker(2012)] emphasized the importance of intelligent tutoring systems (ITS) in dynamically adjusting learning paths based on performance predictions. Despite their success in improving cognitive outcomes, many adaptive systems fail to leverage emotional data, underscoring the need to integrate multimodal inputs to address both cognitive and affective needs ([Azevedo and Aleven(2013)]).

2.2 Multimodal Interaction Data

The use of multimodal interaction data, such as eye-tracking, heart rate variability (HRV), and emotion recognition, provides deeper insights into learner engagement and emotional states.
[D'Mello and Graesser(2012)] demonstrated how systems like AutoTutor utilize emotional and cognitive engagement metrics to adapt feedback. [Jaques et al.(2014)] further highlighted the predictive power of HRV and emotion data in assessing engagement and task difficulty. However, few systems effectively leverage these data streams for real-time feedback adaptation, leaving a gap in the practical application of biometric inputs to personalize learning experiences.

2.3 Multimodal Learning Analytics (MMLA) and Affective-sensitive Adaptive Feedback Systems

Advancements in multimodal learning analytics (MMLA) have led to the development of systems that integrate diverse data streams to enhance learning outcomes.
For instance, [Schneider et al.(2017)] introduced the Presentation Trainer, a system that provides real-time feedback on nonverbal communication skills using multimodal data. The system's immediate and actionable feedback supports skill development during practice sessions. Building on this, [Schneider et al.(2018)] proposed the Multimodal Learning Hub (MLH), which captures and integrates customizable multimodal data configurations to support ubiquitous learning scenarios.

[Kim et al.(2018)] explored emotionally aware AI-driven smart classrooms, capable of monitoring presenters' emotional states and adjusting feedback to optimize engagement and memorability. Similarly, [Deeva et al.(2021)] reviewed automated feedback systems, highlighting the need for personalized, data-driven solutions tailored to learners' individual needs. These studies emphasize the importance of incorporating multimodal data to improve adaptive learning technologies.

Earlier systems, such as MACH (My Automated Conversation Coach) by [Hoque et al.(2013)], demonstrated the potential of leveraging multimodal data to enhance social interaction skills through real-time conversational feedback.
More recently, [Schneider et al.(2019)] extended the Presentation Trainer with an immersive Virtual Reality (VR) module, bridging the gap between practice and performance by simulating real-world scenarios.
Additionally, [Worsley(2012)] emphasized the importance of time-series analysis in MMLA, showing how temporal patterns in multimodal data can reveal learners' cognitive and emotional states, enabling more dynamic and responsive educational systems.

2.4 Feedback Sensitivity

Effective feedback plays a critical role in enhancing learning outcomes and motivation. [Shute(2008)] demonstrated that immediate feedback benefits novices, while delayed feedback fosters reflective learning. [Lipnevich and Smith(2009)] emphasized the impact of feedback tone, noting that overly critical feedback can demotivate learners. Despite these findings, many systems overlook the potential of integrating emotional and cognitive data to optimize feedback sensitivity. Incorporating biometric signals, such as HRV and emotion recognition, into feedback systems can enhance their effectiveness by making them more adaptive to individual learner needs ([D'Mello and Graesser(2012)]).

2.5 Conversational Strategies (AB, CS, and AB+CS)

Affective Backchannels (AB), Conversational Strategies (CS), and their combination (AB+CS) have proven effective in fostering engagement and communication in intelligent tutoring systems. AB includes non-verbal cues like nodding or affirmations, which convey empathy, while CS comprises verbal prompts such as open-ended questions or clarifications ([Cassell et al.(2000)]).
[Ayedoun et al.(2016)] demonstrated that AB and CS, when combined AB+CS, significantly improve Willingness to Communicate (WtC), especially in language learning contexts. This study builds on these concepts by integrating biometric data into conversational strategies, enabling real-time, personalized feedback to enhance learner engagement and communication skills.

3 RESEARCH QUESTIONS AND HYPOTHESES

- **Research Question (RQ1):** *What are the correlations between multimodal data features (e.g., eye-tracking, heart rate, emotion) and adaptive feedback effectiveness in terms of engagement and task performance?*
  Hypothesis (H1): Multimodal data features will positively correlate with engagement and task performance, where higher emotional and cognitive cues (e.g., stable heart rate, focused eye-tracking) will indicate increased effectiveness of adaptive feedback [Shute(2008), Azevedo and Aleven(2013)].
- **Research Question (RQ2):** *How does feedback timing influence learner engagement during conversational tasks?*
  Hypothesis (H2): Real-time adaptive feedback will significantly enhance learner engagement during conversational tasks compared to delayed feedback by maintaining a steady

interaction flow and reducing frustration. [Lipnevich and Smith(2009), Sweller et al.(2011)].

- **Research Question (RQ3):** *What is the effect of feedback timing on learners' willingness to communicate (WtC) in complex scenarios?*
  Hypothesis (H3): Real-time feedback will lead to higher willingness to communicate (WtC) by improving learner confidence and persistence during complex tasks compared to delayed feedback [MacIntyre et al.(1998)].

4 METHODOLOGY

4.1 Participants

The study will involve 18–30 university students aged 18 years or older who will participate in English language conversations with a conversational agent. Participants will have basic English proficiency and primarily speak Japanese as their first language. This demographic was selected due to the well-documented challenges faced by Japanese learners in oral communication, including cultural hesitation to speak and high levels of language anxiety. These factors make this group ideal for exploring interventions aimed at improving Willingness to Communicate (WtC).

4.2 Technologies

To collect and analyze multimodal data, the experiment employed advanced technologies that provided real-time insights into participants' cognitive and emotional states. These tools ensured the adaptive feedback system was both responsive and personalized. The following technologies were utilized during the study:

- **Tobii Eye-Tracker**: This device tracked participants' gaze patterns, enabling the system to monitor attention levels and focus during conversational tasks. By analyzing fixation points and saccades, the eye-tracker identified moments of distraction or hesitation, allowing the conversational agent to provide timely corrective feedback.
- **OpenFace Software**: OpenFace, an open-source tool, was used to analyze facial expressions and detect emotional states such as frustration, confusion, or engagement. By evaluating subtle facial muscle movements, such as eyebrow raises or smiles, OpenFace captured emotional cues indicating participants' levels of comfort or difficulty during interactions. This data was essential for tailoring the agent's responses to participants' emotional needs, fostering a supportive learning environment.
- **RookMotion Device**: The RookMotion wearable measured heart rate variability (HRV), a physiological indicator of stress and cognitive load. By tracking fluctuations in HRV, the system assessed how participants responded to challenging tasks or feedback. High stress levels, indicated by reduced HRV, triggered the agent to provide simpler instructions or empathetic encouragement, ensuring participants remained engaged without feeling overwhelmed.

These technologies worked in tandem to provide a comprehensive view of participants' cognitive and emotional states during the study. The integration of eye-tracking, facial expression analysis, and HRV data ensured that feedback was context-sensitive and adaptive, enabling the conversational agent to dynamically address participants' needs. This multimodal approach allowed the system to respond effectively to real-time challenges, making the feedback more impactful and tailored to individual experiences.

4.3 Experiment Design

This experiment evaluated the effectiveness of adaptive feedback in enhancing learner engagement, task performance, and persistence. Participants engaged in simulated restaurant conversations with a conversational agent acting as a waitress, practicing conversational skills in English through tasks such as:

- Ordering food or drinks.
- Asking about menu preferences.
- Handling follow-up questions (e.g., clarifying an order).

The restaurant scenario was chosen for its practical relevance and low-stakes nature, minimizing anxiety while promoting conversational fluency. This structured context aligns with real-world scenarios and offers learners opportunities for consistent skill development, making it especially suitable for beginner and intermediate learners.

Objectives and Workflow: The study aimed to:

- Assess the impact of real-time feedback on engagement, task accuracy, and persistence.
- Explore whether the scenario promotes Willingness to Communicate (WtC).

Participants were divided into three groups based on feedback conditions. Tasks were designed to progressively evaluate participants' ability to adapt and persist through interactions.

4.4 Data Collection

4.4.1 *Quantitative Data.*

Quantitative data will include:

- **Biometric Data**:
  - Gaze patterns from the Tobii eye-tracker to assess focus.
  - Emotional states from OpenFace to measure engagement or frustration.
  - HRV from the RookMotion device to track physiological stress.
- **Task Performance Metrics**: Completion rates, error rates, and conversation metrics (e.g., turns, pauses).
- **Engagement Metrics**: Time spent on tasks and frequency of feedback interactions.

To ensure accuracy, individual physiological baselines will be established for each participant before the experiment. Self-reports collected via pre- and post-surveys will triangulate biometric data and account for cultural and personal variability in emotional expression. This triangulation provides a comprehensive understanding of participants' engagement and emotional states, reducing potential misinterpretation.

4.4.2 *Qualitative Data.*

Qualitative data will include:

- **Pre-Survey**:
  - Collects demographic information (e.g., English proficiency).
  - Assesses participants' confidence in using English in real-life scenarios (e.g., restaurant interactions).
- **Post-Survey**:

  - Gathers feedback on participants' experiences during the conversation tasks.
  - Captures perceived changes in confidence, engagement, and task difficulty.

5 DATA ANALYSIS

5.1 Quantitative Analysis

5.1.1 ***Correlation Analysis.*** Correlation analysis will be conducted to assess the relationships between biometric signals (e.g., eye-tracking, heart rate variability, emotion recognition) and task performance indicators such as completion rates and error rates. This analysis aims to determine how physiological and emotional responses influence participants' communication effectiveness.

5.1.2 ***Analysis of Variance (ANOVA).*** ANOVA will be used to compare engagement levels, task persistence, and communication performance across the three experimental conditions:

- **Real-time adaptive feedback**
- **Delayed feedback**
- **Control (non-adaptive feedback)**

This analysis will identify significant differences between conditions to evaluate the effectiveness of adaptive feedback strategies.

5.2 Qualitative Analysis

Qualitative data from pre-surveys and post-surveys will be analyzed to identify trends and changes in participants' confidence levels, engagement, and perceived usefulness of the feedback. Comparisons between pre- and post-survey responses will reveal whether the feedback influenced participants' Willingness to Communicate (WtC).

To control for multiple hypothesis testing, Bonferroni corrections will be applied to maintain robust statistical significance thresholds. This approach reduces the likelihood of false positives when examining correlations across a large number of biometric and task-related features.

5.3 Temporal Analytics

Temporal analytics will be incorporated to track how engagement metrics and emotional states evolve during each session. Inspired by [Worsley(2012)] 's work on time-series analysis in multimodal learning, the study will explore changes in:

- Heart rate variability (HRV)
- Gaze fixation patterns
- Emotional expressions

These temporal trends will provide insights into how participants adapt to feedback in real-time and how their persistence develops over successive conversational turns.

6 RESULTS

6.1 Engagement

Real-time adaptive feedback had a significant impact on learner engagement among the 9 participants. Those receiving Affective Backchannels (AB) combined with Conversational Strategies (CS) exhibited higher gaze fixation (average fixation: 0.45 for both eyes), indicating sustained attention during tasks. Additionally, participants experienced a reduction in heart rate, with the average heart rate decreasing from 81 bpm to 69 bpm during adaptive feedback sessions. This physiological change suggests that real-time feedback not only maintained engagement but also reduced stress, helping participants feel more comfortable during interactions.

Temporal analysis revealed a gradual decline in physiological stress indicators, such as heart rate, over the course of the tasks. This trend highlights increasing participant comfort with the conversational agent. Notably, Japanese learners showed significant gains in Willingness to Communicate (WtC), particularly in later stages of the interaction, demonstrating the effectiveness of adaptive feedback in reducing initial hesitation.

6.2 Task Performance

Participants who received real-time adaptive feedback demonstrated significantly higher task accuracy compared to those receiving delayed or traditional feedback. The task completion rates clearly indicate the effectiveness of integrating Affective Backchannels (AB) with Conversational Strategies (CS) to provide personalized, real-time support during learning interactions [CISSE(2024)]:

- **AB+CS group**: 92%, reflecting the benefits of immediate, adaptive feedback in maintaining focus and reducing confusion during tasks.
- **Delayed feedback group**: 74%, showing moderate improvement, but lacking the immediate corrective support needed to sustain optimal performance.
- **Control group**: 61%, emphasizing the limitations of non-adaptive feedback in supporting learners during complex tasks.

Additionally, participants in the AB+CS group exhibited the lowest error rates, further underscoring the value of real-time feedback. This reduction in errors can be attributed to the system's ability to dynamically address participants' challenges by providing context-sensitive feedback tailored to both emotional and cognitive states.

Real-time adaptive feedback enabled learners to correct mistakes more effectively and stay on track, particularly in scenarios requiring complex decision-making or multitasking. For example, during tasks involving multiple conversational turns or nuanced menu preferences, participants in the AB+CS group outperformed those in other groups by quickly adapting to suggestions provided by the conversational agent [CISSE(2024)].

This finding highlights the role of immediate feedback in reinforcing task-related behaviors, sustaining attention, and building confidence. In contrast, delayed feedback, while somewhat beneficial, failed to provide the real-time scaffolding necessary to minimize errors promptly. The control group's lower performance demonstrates the limitations of static, non-adaptive feedback in addressing real-time learning challenges.

These results reinforce the importance of integrating real-time multimodal feedback mechanisms in learning environments to optimize task performance and reduce learner frustration. Future studies could explore how this approach generalizes to more complex scenarios or higher-stakes environments [CISSE(2024)].

6.3 Emotional Responses

Emotion recognition data indicated lower frustration levels among participants during real-time adaptive feedback sessions. For example:

- **Lip corner depression (AU15_R)**: Average of 0.24 in the AB+CS group, compared to 0.36 in the control group.
- **Smile intensity (AU06_R)**: Average of 1.32 in the real-time feedback group, reflecting greater engagement and satisfaction during tasks.

While physiological data demonstrated reduced frustration, self-reports validated these findings by confirming positive emotional experiences among participants. However, discrepancies in a subset of participants emphasize the importance of considering cultural and individual variability in interpreting biometric signals.

The restaurant setting was an effective experimental context due to its low-stakes nature, which mitigated anxiety and facilitated second-language communication practice. Participants rehearsed structured dialogues in a controlled environment, providing a solid foundation for building conversational skills. Future research could expand this approach to high-stakes scenarios, such as job interviews or public speaking, to explore its applicability in emotionally charged contexts.

6.4 Learner Persistence

Learner persistence was notably higher among participants receiving real-time adaptive feedback compared to those in the control group, highlighting the impact of personalized support on sustained engagement. Persistence rates were as follows:

- **AB+CS group**: 88%, indicating that immediate and tailored feedback effectively encouraged participants to stay engaged, even during complex and demanding tasks.
- **Control group**: 65%, reflecting the challenges faced by participants without adaptive feedback in maintaining focus and perseverance during interactions.

Real-time feedback enabled participants to overcome barriers in conversational tasks, such as navigating intricate menu options or responding to unexpected follow-up questions. The adaptive nature of the feedback, which dynamically responded to participants' cognitive and emotional states, provided the necessary scaffolding to help them persist in their efforts. For example, learners in the AB+CS group reported feeling more confident and supported when faced with conversational challenges, attributing this to the agent's empathetic and contextually relevant cues.

These results suggest that adaptive feedback strategies play a pivotal role in enhancing learners' Willingness to Communicate (WtC), especially in scenarios requiring persistence and problem-solving. Participants receiving real-time feedback demonstrated greater resilience, maintaining their willingness to engage with the conversational agent despite encountering complex or unfamiliar situations. This aligns with prior findings that emphasize the role of personalized feedback in fostering persistence by reducing cognitive load and mitigating frustration [Shute(2008), Sweller et al.(2011)].

In contrast, participants in the control group often struggled to maintain engagement during difficult tasks, as static, non-adaptive feedback lacked the flexibility to address their individual needs. This resulted in higher dropout rates or incomplete conversational exchanges, further underscoring the limitations of traditional feedback mechanisms.

The findings reinforce the potential of real-time adaptive feedback to support learners in maintaining focus and motivation during extended tasks, ultimately enhancing their WtC in complex, dynamic scenarios. Future research could explore how such strategies perform in high-stakes environments, such as professional interviews or academic presentations, to assess their broader applicability.

7 DISCUSSION

This study demonstrates that real-time adaptive feedback, particularly when combining Affective Backchannels (AB) and Conversational Strategies (CS), significantly enhances engagement, task performance, and emotional responses. Tailored, immediate feedback reduced frustration and stress, as evidenced by lower heart rates and more positive emotional expressions. In contrast, delayed feedback resulted in lower engagement and task completion rates, highlighting the importance of real-time feedback for maintaining flow and confidence during communication.

By integrating emotional and physiological data (e.g., eye-tracking, heart rate, emotion recognition), this study advances learning analytics and adaptive learning technologies. The findings emphasize the need to address both cognitive and emotional dimensions in multimodal learning environments to create more personalized and effective feedback systems.

7.1 Limitations

The small sample size of 9 participants limits the generalizability of the findings. Expanding the study to include a larger, more diverse participant pool is necessary to validate the results and assess scalability. Additionally, the restaurant scenario, while effective for fostering low-stakes conversational confidence, may not reflect the complexities of high-stakes or emotionally charged environments.

7.2 Implications and Future Research

The conversational agent demonstrated efficacy in enhancing learners' Willingness to Communicate (WtC) by addressing anxiety and providing structured, personalized feedback. This approach bridges the gap between classroom instruction and real-world communication challenges, offering learners a practical, controlled setting for skill development.

However, the system's utility in high-pressure contexts, such as job interviews or public speaking, remains limited. These scenarios demand advanced conversational strategies, greater cultural sensitivity, and the ability to handle dynamic emotional responses, which the current system does not fully replicate. Future research should explore:

(1) The adaptability of conversational agents in high-stakes environments.
(2) The integration of advanced affective computing to simulate nuanced emotional and cultural interactions.
(3) The system's effectiveness with advanced learners who require more diverse and spontaneous interactions.

While the restaurant scenario proved valuable for structured practice, future iterations could extend to emotionally complex settings to better reflect the challenges learners face in professional and social interactions.

8 CONCLUSION

This study's preliminary findings demonstrate that real-time adaptive feedback significantly enhances learner engagement, task performance, emotional responses, and persistence during conversational tasks. By integrating Affective Backchannels (AB) and Conversational Strategies (CS), the system created a supportive and personalized learning environment. Participants receiving feedback based on biometric data—such as eye-tracking, heart rate, and emotion recognition—showed greater engagement and reduced stress compared to those receiving delayed or traditional feedback.

The study underscores the potential of multimodal interaction data in personalizing feedback to address both cognitive and emotional dimensions of learning. While the results are promising, the small sample size of 9 participants limits generalizability. Future research will expand the sample and explore diverse educational contexts to provide deeper insights into the long-term effects of multimodal feedback on learning outcomes.

As this research evolves, the goal is to refine adaptive feedback technologies to ensure scalability and reliability in personalized learning environments. Future efforts will focus on:

(1) Enhancing the triangulation of biometric and qualitative data for more accurate emotional state interpretation.
(2) Refining methods for establishing individual physiological baselines.
(3) Expanding self-report measures to better capture nuanced cultural and individual differences.

These advancements will contribute to the development of adaptive learning systems capable of improving educational outcomes across various fields.

Acknowledgments

This work was supported by JST SPRING, Grant Number JPMJSP2138.

References

[Ayedoun et al.(2016)] Emmanuel Ayedoun, Yuki Hayashi, and Kazuhisa Seta. 2016. Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 1–28. https://doi.org/10.1007/s40593-018-0171-6

[Azevedo and Aleven(2013)] Roger Azevedo and Vincent Aleven (Eds.). 2013. *International Handbook of Metacognition and Learning Technologies*. Springer.

[Cassell et al.(2000)] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill (Eds.). 2000. *Embodied Conversational Agents*. MIT Press.

[CISSE(2024)] Aboul Hassane CISSE. 2024. Real-time Adaptive Learning Environments Using Gaze and Emotion Recognition Engagement and Learning Outcomes. *32nd International Conference on Computers in Education (ICCE 2024)* (2024).

[CISSE et al.(2024)] Aboul Hassane CISSE, Kazuhisa Seta, and Yuki Hayashi. 2024. Integrative Analysis of Multimodal Interaction Data: Predicting Communication Dynamics and Willingness to Communicate (WtC) in Human-Agent Interaction. *Learning Analytics Summer Institute Europe 2024 (LASI Europe 2024 DC)* (2024). https://ceur-ws.org/Vol-3738/paper4.pdf

[Deeva et al.(2021)] Galina Deeva, Daria Bogdanova, Estefanía Serral, Monique Snoeck, and Jochen De Weerdt. 2021. A review of automated feedback systems for learners: Classification framework, challenges, and opportunities. *Computers in Education* (2021).

[Desmarais and Baker(2012)] Michel C. Desmarais and Ryan S. J. D. Baker. 2012. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction* 22, 1 (2012), 9–38. https://doi.org/10.1007/s11257-011-9106-8

[D'Mello and Graesser(2012)] Sidney K. D'Mello and Arthur C. Graesser. 2012. AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems* 2, 4 (2012), 23. https://doi.org/10.1145/2395123.2395128

[Hattie and Timperley(2007)] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of Educational Research* 77, 1 (2007), 81–112. https://doi.org/10.3102/003465430298487

[Hoque et al.(2013)] Mohammed Ehsan Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: My automated conversation coach. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*.

[Jaques et al.(2014)] Natasha Jaques, Cristina Conati, Jason M. Harley, and Roger Azevedo. 2014. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In *Intelligent Tutoring Systems (Lecture Notes in Computer Science, Vol. 8474)*. Springer, 29–38. https://doi.org/10.1007/978-3-319-07221-0_4

[Kim et al.(2018)] Yelin Kim, Tolga Soyata, and Reza Feyzi Behnagh. 2018. Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education. *IEEE Access* (2018).

[Kulik and Fletcher(2016)] James A. Kulik and John D. Fletcher. 2016. Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research* 86, 1 (2016), 42–78. https://doi.org/10.3102/0034654315581420

[Lipnevich and Smith(2009)] Anastasiya A. Lipnevich and Jeffrey K. Smith. 2009. The effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied* 15, 4 (2009), 319–333. https://doi.org/10.1037/a0017841

[MacIntyre et al.(1998)] Peter D. MacIntyre, Richard Clément, Zoltán Dörnyei, and Kimberly A. Noels. 1998. Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal* 82, 4 (1998), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

[Picard(1997)] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press.

[Schneider et al.(2017)] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. 2017. Presentation Trainer: What experts and computers can tell about your nonverbal communication. *Journal of Computer Assisted Learning* 33, 2 (2017), 164–177. https://doi.org/10.1111/jcal.12175

[Schneider et al.(2019)] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. 2019. Beyond reality: Extending a presentation trainer with an immersive virtual reality module. *Sensors* (2019).

[Schneider et al.(2018)] Jan Schneider, Daniele Di Mitri, Bibeg Limbu, and Hendrik Drachsler. 2018. Multimodal Learning Hub: A tool for capturing customizable multimodal learning experiences. In *European Conference on Technology Enhanced Learning (EC-TEL)*.

[Shute(2008)] Valerie J. Shute. 2008. Focus on formative feedback. *Review of Educational Research* 78, 1 (2008), 153–189. https://doi.org/10.3102/0034654307313795

[Sweller et al.(2011)] John Sweller, Paul Ayres, and Slava Kalyuga. 2011. *Cognitive Load Theory*. Springer.

[Worsley(2012)] Marcelo Worsley. 2012. Multimodal learning analytics: Enabling the future of learning through multimodal data analysis and interfaces. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*. 353–356. https://doi.org/10.1145/2388676.2388755

Understanding and Supporting Students' Learning in Generative AI-assisted Writing Tasks

Author: Kaixun Yang
Monash University
Kaixun.Yang1@monash.edu

ABSTRACT: Students are increasingly relying on Generative AI (GAI) to support their writing, a key pedagogical practice in education. In GAI-assisted writing, students can delegate core cognitive tasks (e.g., generating ideas and turning them into sentences) to GAI while still producing high-quality essays. This presents new challenges for researchers and educators to develop methods that ensure students engage in meaningful cognitive processes during the GAI-assisted writing process. This PhD project aims to first explore the common behavioral patterns displayed by students in GAI-assisted writing tasks, how and to what extent these behavioral patterns are indicative of their cognitive processes, and how these behaviors influence the quality of their written work and learning. Based on these insights, the project will then develop effective technology-based scaffolds to support students' learning in this new GAI-assisted writing setting.

Keywords: Writing Process Analysis, AI-assisted Writing, Generative AI, Student Learning

1 INTRODUCTION

In education, writing is a prevalent pedagogical practice employed by teachers to enhance students' learning (Defazio et al., 2010). Writing not only helps students in expressing their thoughts and ideas but also significantly contributes to the cognitive process of understanding and internalizing knowledge (Lea et al., 1998). Despite the importance of writing, some students may find it a difficult and even daunting task, as they struggle to articulate their thoughts on paper while simultaneously mastering the conventions of writing (Odell and Swersey, 2003). As a result, various digital tools have been developed to support students in their composition processes (Schcolnik, 2018). Recent advancements in Generative AI (GAI) represent a significant development in enhancing AI-powered writing tools (Zhao, 2023), as these models demonstrate remarkable capabilities in understanding and generating human-like text (Chang et al., 2024).

As more and more higher education institutions embraced GAI to support teaching and learning, GAI-assisted writing has become increasingly common among students (Jin et al., 2024). However, some critics argue that relying too much on GAI for writing may hinder students' development of creativity and critical thinking skills (Campoverde-Quezada et al., 2024). For example, students can now delegate key rhetorical and cognitive tasks to GAI (Knowles, 2022), resulting in high-quality written work without engaging in meaningful learning during the writing process. Therefore, it is essential to provide structured support (e.g., scaffolding) that gradually enables students to develop their ability to effectively integrate GAI tools to produce high-quality work, while also ensuring they engage in meaningful learning (e.g., critical thinking). Despite the importance of this issue, few studies have explored how students use GAI in their writing, whether and to what extent meaningful learning occurs through this usage, and how their GAI usage patterns relate to the quality of their written work. Thus, there is still limited understanding of how to effectively design scaffolding that supports students' learning in GAI-assisted writing tasks. This PhD project aims to, first, uncover common patterns of GAI usage and their relationship to students' cognitive writing processes and the quality of their written work. Second, it seeks to design and evaluate effective scaffolding strategies to support student learning in GAI-assisted writing, grounded in the insights from the initial analysis and relevant theories in pedagogical writing practices.

2 RELATED WORK AND EXISTING SOLUTIONS

2.1 Writing Process Analysis

Since the past decade or two, analyzing the writing process was deemed challenging due to the difficulties in tracking activities in handwritten form (Sinharay et al., 2019). However, the advent of digital writing tools has made it more feasible to observe and reconstruct the writing process using keystroke logging. Keystroke logging is a prevalent method for examining the writing process in educational settings, involving the recording and timestamping of keystroke actions to reconstruct the writing process (Leijten and Van, 2013). Current research primarily centers on deriving writing behaviors (e.g., between-word pauses) from keystroke logging. These behaviors are then used to explore their relationships with other writing aspects such as essay quality (Vakkari et al., 2021), task complexity (Révész et al., 2017), and language proficiency (anak Engkamat and Nasri, 2012). Additionally, investigating the cognitive processes behind writing behaviors is a significant area of interest in existing research. Writing encompasses a set of recursive and intertwined cognitive processes (e.g., planning, translating, reviewing, and monitoring) (Koppenhaver and Williams, 2010). Therefore, it is essential to understand the patterns in writer's cognitive processes to better support them in a writing task (hang and Deane, 2015). (Conijn et al., 2019) proposed the method of mapping features from the keystroke logs to higher-level cognitive processes, such as planning and revising. (Baaijen et al., 2012) developed methods and measures for analyzing keystroke logging with the goal of enhancing the correlation between keystroke data and cognitive processes. Overall, existing studies show that writing process analysis using keystroke logging is widely applied in educational research, particularly in linking writing behaviors to cognitive processes. However, writing process analysis in the context of GAI-assisted writing still requires further exploration.

2.2 GAI-assisted Writing

Existing research on GAI-assisted writing remains relatively limited and can be broadly divided into three categories: (i) the development of GAI-assisted writing systems and the collection of relevant datasets; (ii) the analysis of writers' behaviors during GAI-assisted writing; and (iii) the evaluation of writers' performance in GAI-assisted writing tasks.

For developing GAI-assisted writing systems and gathering relevant datasets, (Coenen et al., 2021) introduced Wordcraft, an AI-assisted editor designed for collaborative story writing using few-shot learning and conversational affordances. Another notable work by (Lee et al., 2022) presented CoAuthor, a dataset capturing interactions between 63 writers and four instances of GPT-3 across 1,445 writing sessions. For analyzing writers' writing behaviors in GAI-assisted writing, (Cheng et al., 2024) proposed a methodology based on learning analytics to evaluate human writing processes in

GAI-assisted writing, comparing writing behaviors across different groups (e.g., creative vs. argumentative writing). For correlating writing behaviors with writing performance, (Shibani et al., 2023) introduced CoAuthorViz, a tool that visualizes keystroke logs from GAI-assisted writing and explored the correlations between various human writing behaviors and the quality of the final products. (Nguyen et al., 2024) used Hidden Markov Models combined with hierarchical sequence clustering to analyze human-AI interactions in academic writing, finding that doctoral students engaging in iterative, highly interactive writing processes with AI tools tend to perform better. However, these studies fall short in providing practical guidance for designing next-generation GAI writing systems that more effectively support student learning. They often fail to link GAI writing behaviors to the learning process and to establish causal relationships between these behaviors and writing quality.

3 RESEARCH QUESTIONS

This PhD project is divided into two phases. The first phase focuses on **understanding** writing behaviors in GAI-assisted writing by examining common GAI usage patterns, their relationship with cognitive processes, and their impact on writing quality through the analysis of GAI-assisted writing datasets. The second phase aims to **support** students' learning in GAI-assisted writing tasks by designing scaffolding based on the findings from the first phase and evaluating their effectiveness. Formally, in **Phase 1**, we aim to answer the following research questions: **(RQ1)** What are the common patterns of GAI usage in the setting of GAI-assisted writing, and how are they correlated with cognitive writing processes? **(RQ2)** What GAI-assisted writing behaviors contribute to the quality of written products? In **Phase 2**, we seek to address: **(RQ3)** What scaffolding can be designed to effectively support student' learning in GAI-assisted writing?

4 METHODOLOGY

4.1 RQ1

We will focus on a public GAI-assisted writing dataset consisting of 1,445 writing sessions (Lee et al., 2022). This dataset includes not only the final written products but also keystroke logging captured throughout the entire writing process. Several educational studies have already recognized its value and incorporated it into their research (Shibani et al., 2023; Cheng et al., 2024).

We plan to construct GAI-writing behaviors from keystroke logging. We argue that the temporal dynamics of GAI-writing behaviors reveal more nuanced insights than overall GAI usage patterns. For example, one writer might use AI extensively at the beginning of their writing process and then gradually reduce its use, while another might start without AI assistance and increasingly rely on it as they progress. Although these two writers may show similar overall AI usage when evaluated based on the final product (i.e., the total number of times AI suggestions are sought), their temporal AI usage patterns are distinctly different. To capture these differences, we plan to apply **time-series clustering** techniques. Writing sessions vary in length and contain different amounts of keystroke logs, making standard clustering methods like K-means unsuitable for handling such variability. Therefore, we will use **Dynamic Time Warping (DTW)** (Müller, 2007), a technique that measures the similarity between action sequences regardless of differences in length. Additionally, to uncover the potential impact of GAI usage on human writing behaviors that reflect cognitive processes, we will conduct **Epistemic Network Analysis (ENA)** (Shaffer and Ruis, 2017), a popular learning analytics method that utilizes network models to depict epistemic actions, on each cluster to evidence the cognitive behaviors inherent to co-writing with GAI.

4.2 RQ2

To address RQ2, we will use the same dataset as in RQ1. Our goal is to detect **causal relationships** revealing the GAI-assisted writing behaviors that significantly contribute to the quality of written products, which can be used to better inform the design of scaffoldings in RQ3. We will focus on the behavioral patterns identified in RQ1. To evaluate the quality of the written products, we plan to use measures (e.g., lexical sophistication) commonly employed in previous writing research. Given the challenges of designing randomized controlled experiments for our study—such as variations in how participants might respond to GAI suggestions based on their individual writing styles, making it difficult to control interactions with GAI consistently across participants—we plan to apply **causal modeling** (Feder et al., 2022). This statistical method is designed to uncover and understand cause-and-effect relationships using existing observational data, allowing us to detect how different GAI-assisted writing behaviors contribute to essay quality.

4.3 RQ3

To address RQ3, we will design **technology-based scaffolds** (Sharma and Hannafin, 2007) that provide targeted support and guidance to learners during the GAI-assisted writing process, informed by our findings in RQ1 and RQ2. The effectiveness of technology-based scaffolds has been demonstrated in various fields, including self-regulated learning (Lim et al., 2023) and problem-based learning (Simons et al., 2007). We propose that incorporating **real-time feedback** and **automated prompts or hints** within technology-based scaffolds can enhance student learning in GAI-assisted writing. These tools can guide students to actively engage in meaningful learning activities (e.g., self-review) rather than passively relying on GAI assistance. For instance, immediate feedback can be triggered when algorithms detect behaviors indicative of less meaningful engagement, while automated prompts or hints can be provided when students struggle to effectively use GAI tools in their writing (e.g., How to effectively adapt GAI-generated text to suit the writer's context). Additionally, GAI writing settings (e.g., GAI parameters) can be dynamically adjusted based on students' writing progress or teachers' specific requirements (e.g., desired writing genre).

5 ETHICAL CONSIDERATION

Currently, our study does not involve any ethical concerns. As we plan to conduct human studies in the future to evaluate the effectiveness of our proposed scaffolding, we will seek ethical approval from the Human Research Ethics Committee of Monash University.

6 CONTRIBUTION OF SUGGESTED SOLUTION

The solution proposed in this project differs from existing approaches in three keyways. First, while most studies categorized writers based on their overall AI usage during the entire writing session, they often overlooked the temporal dynamics of these behaviors as the writing session progressed. Additionally, these studies revealed limited insight into writers' writing behaviors and the corresponding cognitive processes and thus offered limited implications for supporting learning in these GAI-assisted writing tasks. Second, when examining the relationship between GAI usage and essay quality, most research focused on comparing two groups: those using GAI writing assistance and those without. This approach failed to explore how different GAI-assisted behaviors impact writing quality. Some research explores the relationship between writing behaviors and essay quality, but mainly identifies correlations rather than causal relationships. While this can contribute to understanding writing behaviors in this new context, it provides limited practical guidance for designing next-generation writing systems that can better support students' learning. Third, to our knowledge, no existing scaffolds have been developed to support students' learning specifically in GAI-assisted writing tasks.

7 ACHIEVED SO FAR

A paper addressing RQ1 has been submitted to the British Journal of Educational Technology. It identifies common patterns of GAI usage (e.g., a preference for independent writing) and explores their connections to cognitive processes involved in knowledge building (e.g., knowledge transformation). The paper is currently under review. Another paper addressing RQ2 has been accepted for presentation at the International Conference on Learning Analytics and Knowledge 2025 (LAK25), presenting the causal relationships between GAI-assisted writing behavioral patterns and essay quality.

REFERENCES

anak Engkamat, D. T., & Nasri, N. M. (2012). The relationship between English writing ability levels and EFL learners’ metacognitive behavior in the writing process. International Journal of Academic Research in Progressive Education and Development, 1(4).

Baaijen, V. M., Galbraith, D., & De Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. Written Communication, 29(3), 246-277.

Campoverde-Quezada, D. A., & Valdiviezo-Ramírez, E. A. (2024). The Double-Edged Sword: Benefits and Challenges that Artificial Intelligence Tools can Bring to EFL Teaching and Learning. Revista Metropolitana de Ciencias Aplicadas, 7(2), 304-316.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 15(3), 1-45.

Cheng, Y., Lyons, K., Chen, G., Gašević, D., & Swiecki, Z. (2024, March). Evidence-centered Assessment for Writing with Generative AI. In Proceedings of the 14th Learning Analytics and Knowledge Conference (pp. 178-188).

Coenen, A., Davis, L., Ippolito, D., Reif, E., & Yuan, A. (2021). Wordcraft: A human-AI collaborative editor for story writing. arXiv preprint arXiv:2107.07430.

Conijn, R., Roeser, J., & Van Zaanen, M. (2019). Understanding the keystroke log: the effect of writing task on keystroke features. Reading and Writing, 32(9), 2353-2374.

Defazio, J., Jones, J., Tennant, F., & Hook, S. A. (2010). Academic Literacy: The Importance and Impact of Writing across the Curriculum--A Case Study. Journal of the Scholarship of Teaching and Learning, 10(2), 34-47.

Feder, A., Keith, K. A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., ... & Yang, D. (2022). Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. Transactions of the Association for Computational Linguistics, 10, 1138-1158.

Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2024). Generative AI in Higher Education: A Global Perspective of Institutional Adoption Policies and Guidelines. arXiv preprint arXiv:2405.11800.

Knowles, A. M. (2022, July). Human-AI collaborative writing: Sharing the rhetorical task load. In 2022 IEEE International Professional Communication Conference (ProComm) (pp. 257-261). IEEE.

Koppenhaver, D., & Williams, A. (2010). A conceptual review of writing research in augmentative and alternative communication. Augmentative and Alternative Communication, 26(3), 158-176.

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. Studies in higher education, 23(2), 157-172.

Lee, M., Liang, P., & Yang, Q. (2022, April). Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In Proceedings of the 2022 CHI conference on human factors in computing systems (pp. 1-19).

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. Written Communication, 30(3), 358-392.

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., ... & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. Computers in Human Behavior, 139, 107547.

Müller, M. (2007). Dynamic time warping. Information retrieval for music and motion, 69-84.

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. Studies in Higher Education, 1-18.

Odell, L., & Swersey, B. (2003). Reinventing invention: Writing across the curriculum without WAC. Language and Learning across the Disciplines, 6(3), 39-54.

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. Language Learning, 67(1), 208-241.

Schcolnik, M. (2018). Digital Tools in Academic Writing?. Journal of Academic Writing, 8(1), 121-130.

Shaffer, D., & Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. Handbook of learning analytics.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. Interactive learning environments, 15(1), 27-46.

Shibani, A., Rajalakshmi, R., Mattins, F., Selvaraj, S., & Knight, S. (2023). Visual Representation of Co-Authorship with GPT-3: Studying Human-Machine Interaction for Effective Writing. International Educational Data Mining Society.

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. Instructional science, 35, 41-72.

Sinharay, S., Zhang, M., & Deane, P. (2019). Prediction of essay scores from writing process and product features using data mining methods. Applied Measurement in Education, 32(2), 116-137.

Vakkari, P., Völske, M., Potthast, M., Hagen, M., & Stein, B. (2021). Predicting essay quality from search and writing behavior. Journal of the Association for Information Science and Technology, 72(7), 839-852.

Zhang, M., & Deane, P. (2015). Process features in writing: Internal structure and incremental value over product features. ETS Research Report Series, 2015(2), 1-12.

Zhao, X. (2023). Leveraging artificial intelligence (AI) technology for English writing: Introducing wordtune as a digital writing assistant for EFL writers. RELC Journal, 54(3), 890-894.

Navigating the 21st Century Learning: Assessing Graduate Attribute Development through Learning Analytics

Vimukthini Jayalath
University of South Australia
balage.jayalath@mymail.unisa.edu.au

ABSTRACT: The complex demands of the modern workforce have increased pressure on universities to equip graduates with the necessary skills beyond technical expertise. While universities have taken steps to integrate graduate attributes such as problem-solving and teamwork in curricula for accreditation, evaluating these skills often involves subjective methods that, while insightful, face challenges in ensuring consistency and scalability. This study addresses the gap in measuring graduate attributes, leveraging psychometrics, machine learning, and generative AI. The research aims first to map course assessments to these graduate attributes and then use the developed fine-grained mapping for measurement. This innovative approach has the potential to improve curriculum alignment and better support students' skill development while also contributing a new methodology to the field. Developing a learner profile incorporating graduate attribute progressions throughout their degree program, can be valuable a resource for learners and contribute to learning analytics by helping universities better prepare graduates for the 21st-century workplace.

Keywords: Graduate attributes, educational assessment, curriculum mapping, generative AI

1 INTRODUCTION

Higher education plays an active role in enhancing the graduates' skills, capabilities, and disciplinary expertise. Higher Education Institutions (HEIs) prioritising producing work-ready graduates are more likely to meet accreditation standards and maintain a high academic reputation (Oraison et al., 2019). In the Australian tertiary education sector, "graduate attributes" (GAs) and "graduate qualities" (GQs) describe the skills learners are expected to acquire upon graduation (Barrie, 2006). These skills typically include communication, leadership, problem-solving, and collaboration. Universities have integrated GAs through industry placements, co-curricular and extra-curricular activities (Jackson & Bridgstock, 2021).

While HEIs have integrated GAs into their curricula, significant challenges remain with evaluating their attainment. The commonly used methods of evaluating these broad competencies can be biased and not always scalable (Barthakur et al., 2024), with significant methodological challenges. Those include the lack of a common framework for implementing GA teaching (Hammer et al., 2021), the skills' definitions tend to be broad, vague and general, and the lack of a standard method to assess GAs at scale (Sanil et al., 2019). University courses are also designed with a simple binary mapping between GAs and learning activities and/or assessments, which is usually an oversimplification of the relationship between GA and learning activities (Barthakur et al., 2024). However, achieving higher granularity at scale presents considerable challenges. As a result, studies evaluating the development of GAs through assessment scores are limited.

There has been a growing number of attempts to assess skills similar to GAs within the field of Learning Analytics (LA), such as using Log stream data and assessment data (Barthakur et al., 2024; Milligan, 2015). Similarly, recent advancements in Generative Artificial Intelligence (GenAI; Vaswani et al., 2017) present an opportunity to enhance the mapping between GAs and learning activities. In this regard, GenAI and LA can improve curriculum mapping by reducing subjectivity and providing a more nuanced, and scalable approach, thereby supporting GA development evaluation in HEIs (Zamecnik et al., 2024).

This doctoral research aims to refine the curriculum mapping process initially and then assess the learners' GA development by analysing their assessment grades employing techniques from psychometrics and educational assessment (Mislevy, 2017). Finally, the research will use LA-based approaches to analyse the progression of learners' GA profiles over time, identifying developmental trajectories to compare learner profiles across different fields. The study will offer valuable insights into GA development by integrating longitudinal assessment data from multiple courses with the detailed mapping provided by GenAI. This approach will contribute to LA by presenting learner profiles and enhancing the understanding of GA growth and progression.

2 BACKGROUND

2.1 Graduate Attributes Measurement

Graduates with strong interpersonal skills and disciplinary knowledge meet modern employers' broader expectations. These competencies, often called Graduate Attributes, encompass skills and knowledge expected upon graduation (Oliver, 2011). In Australia, GAs are recognised as essential for employment and lifelong learning, requiring universities to provide evidence of GA attainment (Oliver, 2011). As part of government accreditation, Australian universities are required to map GAs into their curriculum. This is usually done by mapping GAs to the Course Learning Objectives (CLOs), which are, in turn, mapped to different learning activities and course assessments. Commonly identified GAs in Australian Universities include 1) Written and oral communication, 2) Critical, analytical, creative, and reflective thinking, 3) Problem-solving, 4) Information literacy, 5) Learning and working independently, 6) Learning and working collaboratively, and 7) Ethical and inclusive engagement with communities, cultures, and nations (Oliver, 2011).

Traditional methods of evaluating GAs are limited to curriculum vitae, transcripts (Ajjawi & Boud, 2023), follow-up interviews, reference letters, quizzes, games (Sutil-Martín & Otamendi, 2021) and self and peer-rating scales (Kyllonen, 2013), which are subjective and not scalable. Hence, the need for an objective approach is presented using assessment data to evaluate GAs and their development (Barthakur et al., 2024). While the literature contains cross-sectional studies evaluating GAs, the number of longitudinal studies covering from enrolment to graduation is limited.

Measurement science and educational assessment aim to reliably assess learners' attributes, laying a foundation for accurate educational assessments (Mislevy, 2017). The mappings connecting GAs with CLOs and CLOs with learning activities and assessments can be used as input to the measurement models (Bergner, 2017; Milligan, 2018), such as Item Response Theory (IRT; Baker, 2001), Cognitive Diagnostic Models (CDMs; (de la Torre & Minchen, 2014), and Mixed Membership Models (Blei, 2015), that can provide insights into the attainment of GAs. CDMs are probabilistic models used to assess learners' skills based on their mastery of different skills, providing more detailed and accurate insights into learners' abilities. However, the binary nature of such mappings presents a significant oversimplification and limitation for a more accurate assessment of GA attainment. Redefining the curriculum mapping to incorporate the more nuanced relationships between GAs, CLOs and learning activities is highly labour-intensive and subjective. In this regard, GenAI has shown significant potential in overcoming the challenges of curriculum mapping (McLaren et al., 2024). For example, Zamecnik and colleagues (2024) explored the use of GenAI to map GAs to 26 courses in a higher education program using assessment data and achieved an accuracy of 71%, demonstrating the potential of GenAI to improve the efficiency and accuracy of curriculum mapping. Collectively, these methods can provide a deeper understanding of GA development. This research seeks to build on these studies to explore AI's role in mapping GAs and validating findings across different university courses to improve the understanding of GA development.

2.2 Learner Profiles and Learning Analytics

Learner profiles (Kaffenberger, 2019), in the context of HEIs, offer a representation of a graduate's journey and acquired skills, including GAs, to capture the full scope of a student's growth. While lacking a universally agreed definition, learner profiles have been used in various educational contexts, including personalising learning, profiling online learners in MOOCs (Barthakur et al., 2023), analysing Learning Management System (LMS) usage (Zamecnik et al., 2022), categorising and visualising learners by attributes (Kaffenberger, 2019). Learner profiles developed through LA methods provide a holistic view of learners' skills, overcoming the limitations of grade-based academic transcripts. However, integrating GA development into learner profiles remains a challenge. Barthakur and colleagues (2024) introduced an LA approach to track GA development within Initial Teacher Education programs, and this doctoral research aims to expand upon that method by analysing multiple courses and programs. By incorporating GAs into learner profiles, the research aims to support goal setting, personalised feedback, and data-driven decision-making, ultimately promoting holistic learner development and success.

3 RESEARCH QUESTIONS

Based on the above review and the need to assess GA development longitudinally, using an integrated data-driven approach using GenAI, this doctoral research aims to answer the following research questions. A significant challenge in assessing GA development is the unclear degree of the relationship between the assessments and CLOs in tracking GA progression. This study investigates the shift from binary to weighted mapping (Appendix 1), focusing on how GenAI can support this transition. Accordingly, our first research question is shaped as follows:

RQ1*: How can GenAI help understand the relationship between GA development and curriculum?*

Subsequent questions focus on using LA approaches to identify learner profiles and compare transitions throughout the degree programs based on newly derived scores from weighted mappings.

RQ2a*: What are the various learner profiles of GA development?*

RQ2b*: Are there any associations in developing different GAs in different degree programs?*

RQ3*: How are the learner profiles transitioning across a degree program over time? Are there any specific patterns of this transition across other degree programs?*

4 METHODOLOGY

This research will use anonymised learner assessment grades and curriculum mapping data from several degree programs at a large public Australian university. The university aligns its courses and study programs to develop seven GAs in its graduates.

A pilot study: A pilot study of selected courses from *one degree program* is conducted initially to develop a data pipeline for scaling. Then, a larger dataset covering several degree programs and multiple cohorts will be examined for a broader generalisation of the findings. (Appendix 2)

Study 1: Formulating around RQ1, in the study's first phase, a ground truth is established through a manual weighting process using the course outline documents and marking rubrics, conducted by domain experts (course coordinators/markers) with demonstrated reliability in curriculum and assessments. Agreement between two independent experts will be measured using Krippendorff's alpha (Krippendorff, 2011), and conflicts are resolved to create a unified ground truth for comparison with GenAI-generated mappings. In the second phase, weighted mappings between Graduate Attributes (GAs) and assessments will be developed using GenAI, with efforts to enhance reproducibility through deterministic settings, though challenges remain due to the evolving nature of LLMs. These mappings will be validated, refined for accuracy, and extended to other courses.

Study 2: Building on the first study, the obtained weighted mapping will be used to derive Graduate Attribute (GA) scores for learners using CDMs, focusing on one year of the degree program. This phase will analyse GA scores across courses and assessments within a selected academic year, clustering learners into homogeneous profiles based on their GA performance. Each learner will be assigned to one profile type for that year, providing insights into how they demonstrate GAs over time. The clustering methods will be selected based on their ability to provide meaningful educational insights, such as uncovering probabilistic variations with Gaussian Mixture Models or offering interpretability with k-means clustering. The focus is on using clustering to understand learner behavior and GA attainment, enabling targeted interventions and personalized support. This approach will be applied across all program years to track learner profiles over time, emphasizing the integration of machine learning with educational outcomes to enhance teaching, learning, and curriculum design.

Study 3: The third study will analyse the progression of learners' GA profiles over time to identify developmental trajectories. It will build on the year-by-year analysis of GA development by combining individual GA profiles to uncover overarching patterns, drawing from methodologies outlined by Barthakur and colleagues (2024). Analysing the transition of profiles about GA would initially be conducted using data from one undergraduate degree program and then replicated to more than 10 programs in different fields to compare the progression of profiles across degree programs.

The presented studies will address the gap in the literature on deriving a score for GAs through weighted mapping and assessment data, providing a holistic profile of the learners' progression of GAs. This approach will offer insights for university stakeholders to enhance courses and degree programs and to align better with accreditation standards.

5 CURRENT PROGRESS

Study 1 of the pilot study is in progress, with the GenAI-weighted mapping completed and the expert-weighted mapping set for the upcoming month. After the pilot study for Study 1 is completed, it will be replicated across multiple courses. This phase is critical before progressing to the next two studies. Study 1 is expected to conclude by March 2025 with an ideal timeline to receive expert feedback before starting the subsequent studies.

6 CONTRIBUTION

This doctoral thesis aims to integrate methodologies from various fields, such as psychometrics and LA, to explore strategies to evaluate GA development in university learners. Theoretically, it explores how GAs evolve across different academic disciplines, enhancing current assessment practices. Methodologically, it introduces a novel approach using GenAI for more detailed mapping of CLOs to GAs, improving over binary mappings. For example: In a first-year accounting course, assessments may involve multiple GAs such as problem-solving, teamwork, and communication. Using nonbinary weighting, each assessment is evaluated for its relative emphasis on these GAs, providing an understanding of how students engage with them. For instance, a group project may weight teamwork at 50%, communication at 30%, and problem-solving at 20%. Clustering learners based on their GA performance across assessments could reveal distinct profiles, such as "Collaborative Communicators" (strong in teamwork and communication) or "Analytical Problem-Solvers" (strong in problem-solving but needing support in teamwork). These learner profiles can provide actionable insights for educators to tailor feedback and teaching strategies, curriculum designers to ensure balanced GA development, and academic advisors to guide students in course selection and co-curricular activities. This approach transforms machine learning outputs into practical educational strategies, highlighting the benefits of nonbinary weighting and clustering in supporting student development. Practically, the research aids universities in refining educational assessments and program improvements by analysing GA progression patterns, ultimately benefiting academic institutions in understanding students' strengths and weaknesses.

7 REFERENCE

Ajjawi, R., & Boud, D. (2023). Changing representations of student achievement: The need for innovation. *Innovations in Education and Teaching International*, *0*(0), 1–11.

Baker, F. B. (2001). *The Basics of Item Response Theory. Second Edition*.

Barrie, S. C. (2006). Understanding What We Mean by the Generic Attributes of Graduates. *Higher Education*, *51*(2), 215–241.

Barthakur, A., Dawson, S., & Kovanovic, V. (2023). Advancing leaner profiles with learning analytics: A scoping review of current trends and challenges. *Proceedings of the 13th International Conference on Learning Analytics & Knowledge (LAK'23)*, 606–612.

Barthakur, A., Jovanovic, J., Zamecnik, A., Kovanovic, V., Xu, G., & Dawson, S. (2024). *Towards Comprehensive Monitoring of Graduate Attribute Development: A Learning Analytics Approach in Higher Education*.

Bergner, Y. (2017). Measurement and its Uses in Learning Analytics. In Columbia University, USA, C. Lang, G. Siemens, University of Texas at Arlington, USA, A. Wise, New York University, USA, D. Gasevic, & University of Edinburgh, UK (Eds.), *Handbook of Learning Analytics* (First, pp. 35–48). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.003

Blei, D. M. (2015). *Mixed-membership Models (and an introduction to variational inference)*.

Botterill, M., White, C., & Steiner, T. (2010). *Developing professional skills: Introducing students to graduate attributes in first year engineering at RMIT*.

de la Torre, J., & Minchen, N. (2014). Cognitively Diagnostic Assessments and the Cognitive Diagnosis Model Framework. *Psicología Educativa*, *20*(2), 89–97.

Hammer, S., Ayriss, P., & McCubbin, A. (2021). Style or substance: How Australian universities contextualise their graduate attributes for the curriculum quality space. *Higher Education Research & Development*, *40*(3), 508–523.

Jackson, D., & Bridgstock, R. (2021). What actually works to enhance graduate employability? The relative value of curricular, co-curricular, and extra-curricular learning and paid work. *Higher Education*, *81*(4), 723–739.

Kaffenberger, M. (2019). *A Typology of Learning Profiles: Tools for Analysing the Dynamics of Learning*. Research on Improving Systems of Education (RISE).

Krippendorff, K. (2011). *Computing Krippendorff's Alpha-Reliability*.

Kyllonen, P. C. (2013). Soft Skills for the Workplace. *Change: The Magazine of Higher Learning*, *45*(6), 16–23.

McLaren, B. M., Herckis, L., Teffera, L., Branstetter, L., Rose, C. P., Kisow, M., Reis, R., Rinsem, M., Alenius, M., & Miller, L. (2024). *Community College Information Technology Education: Curriculum Mapping, a Learning Science Framework, and AI Learning Technologies*.

Milligan, S. (2015). *Crowd-Sourced Learning in MOOCs: Learning Analytics meets Measurement Theory*. https://dl.acm.org/doi/10.1145/2723576.2723596

Milligan, S. (2018). *Methodological Foundations for the Measurement of learning in learning analytics*.

Mislevy, R. J. (2017). On Measurement in Educational Assessment. In *Handbook on Measurement, Assessment, and Evaluation in Higher Education* (2nd ed.). Routledge.

Oliver, B. (2011). Assuring Graduate Outcomes. In *The Australian Learning and Teaching Council*.

Oraison, H., Konjarski, L., & Howe, S. (2019). Does university prepare students for employment?: Alignment between graduate attributes, accreditation requirements and industry employability criteria. *Journal of Teaching and Learning for Graduate Employability*, *10*(1), 173–194. https://doi.org/10.3316/informit.580981748647262

Sanil, A., Patwardan, A., Shah, H., & Sawant, P. (2019). *Enhancing Attainment of Graduate Attributes Using Data Science* (SSRN Scholarly Paper 3365528). https://doi.org/10.2139/ssrn.3365528

Sutil-Martín, D. L., & Otamendi, F. J. (2021). Soft Skills Training Program Based on Serious Games. *Sustainability*, *13*(15), Article 15. https://doi.org/10.3390/su13158582

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010.

Zamecnik, A., Barthakur, A., Wang, H., & Dawson, S. (2024). Mapping Employable Skills in Higher Education Curriculum Using LLMs. In R. Ferreira Mello, N. Rummel, I. Jivet, G. Pishtari, & J. A. Ruipérez Valiente (Eds.), *Technology Enhanced Learning for Inclusive and Equitable Quality Education* (pp. 18–32). Springer Nature Switzerland.

Zamecnik, A., Kovanović, V., Joksimović, S., & Liu, L. (2022). Exploring non-traditional learner motivations and characteristics in online learning: A learner profile study. *Computers and Education: Artificial Intelligence*, *3*, 100051. https://doi.org/10.1016/j.caeai.2022.100051

APPENDICES

Appendix 1

Table 1: Example of the binary mapping

CLO	Graduate Attributes						
	GA1	GA2	GA3	GA4	GA5	GA6	GA7
CLO1	1	0	0	0	1	0	1
CLO2	0	1	0	1	0	0	0
CLO3	1	1	0	0	0	1	0
CLO4	0	0	1	0	0	0	1
CLO5	1	1	0	0	0	0	0

Table 2: Example of the weighted mapping

CLO	Graduate Attributes						
	GA1	GA2	GA3	GA4	GA5	GA6	GA7
CLO1	0.5	0	0	0	0.25	0	0.25
CLO2	0	0.1	0	0.9	0	0	0
CLO3	0.3	0.3	0	0	0	0.4	0
CLO4	0	0	0.75	0	0	0	0.25
CLO5	0.7	0.3	0	0	0	0	0

Appendix 2

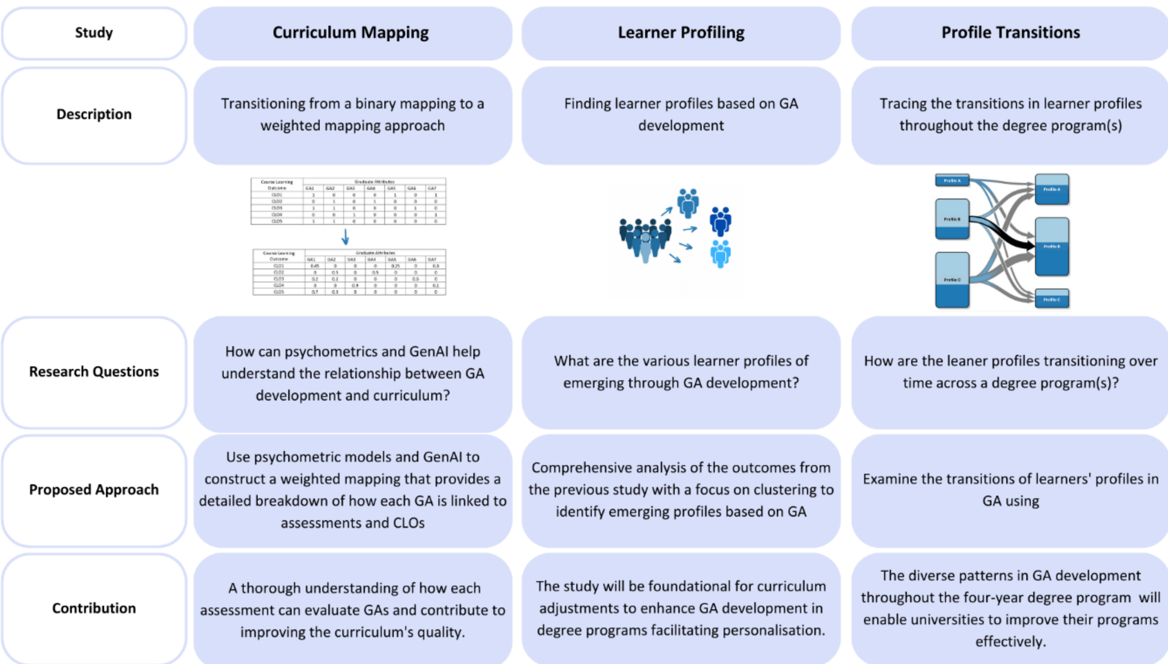


Figure 1: Summary of the doctoral research.

Table Dynamics: AI-Enhanced Analysis of Keyboard Access and Power Dynamics in Collaborative STEM Learning Groups

Hakeoung Hannah Lee
The University of Texas at Austin
hklee@utexas.edu

ABSTRACT: The complexity of collaborative learning necessitates examining multiple dynamics, particularly those involving power. Grounded in sociocultural perspectives of learning, this study uses a unique mixed-methods design, leveraging an artificial intelligence(AI)-based Activity Mapping approach alongside qualitative analysis to explore collaborative group dynamics among students in the Advancing Out-of-school Learning in Mathematics and Engineering program. This study examines how access to a shared keyboard within collaborative groups reflects and/or mediates learning dynamics, as understood through patterns of social and intellectual authority. Theoretically, this work advances the understanding of power dynamics around shared tools, uncovering nuanced ways in which power relations are constructed, maintained, and challenged in collaborative learning. Methodologically, the study introduces an approach that integrates qualitative frameworks with AI-enhanced analysis of large-scale classroom video data, offering a level of specificity unattainable by either approach alone. The study provides key insights to inform the design of more inclusive, equitable collaborative learning environments.

Keywords: Collaborative Learning, Power Dynamics, Artificial Intelligence, Social Interaction, Computer-Supported Collaborative Learning Environment, STEM Education, Activity Mapping

1 INTRODUCTION

The inherent complexity of collaborative learning necessitates a detailed examination of multiple dynamics (Vygotsky, 1987), particularly those involving power (Engle et al., 2014; Wertsch et al., 1993). Power dynamics play a pivotal role in learning (Esmonde & Booker, 2017; Lave & Wenger, 1991), as learning is inherently a social process (Lave & Wenger, 1991; Vygotsky, 1978). Understanding the distribution of authority in collaborative learning environments reveals much about how learning processes unfold and evolve (Esmonde & Booker, 2017; Lave & Wenger, 1991). It provides insights into how learners interact, negotiate, and co-construct knowledge, while also highlighting the dynamic nature of power relations and their impact on learning outcomes. Engle and colleagues (2014) propose a framework for understanding how undue influence develops in student discussions. Building on this work, Langer-Osuna and colleagues (2020) synthesize the influence framework into the constructs of social and intellectual authority. In educational settings, *social authority* is omnipresent, occurring whenever individuals interact, and *intellectual authority* arises during intellectual activities, typically recognized in educational contexts as participation in academic tasks. Despite its critical importance, this dimension of educational research remains underexplored (Langer-Osuna et al., 2020).

Power dynamics are mediated by the availability of historically contingent artifacts or tools (Engeström, 2015; Esmonde & Booker, 2017). While advanced technological tools and artificial intelligence (AI) have been extensively explored for understanding learning and human behavior (e.g., Computer-Supported Collaborative Learning, Multimodal Learning Analytics), the use of AI technology to investigate power dynamics among students in collaborative settings has been extremely limited. There has been relatively little focus on the learning processes through which inequity can emerge in collaborative learning (Shah & Lewis, 2019) and how authority is distributed as students engage in these settings (Hübscher-Younger & Narayanan, 2003; Langer-Osuna et al., 2020). This gap presents a significant opportunity, as AI's strengths in large-scale data analysis could reveal new insights into power dynamics and engagements in these settings. This study highlights the importance of examining how access to tools and objects relates to power dynamics and engagement in collaborative learning. While qualitative approaches have offered nuanced insights into micro-interactional group work (Jordan & Henderson, 1995), they require a complementary quantitative approach to systematically capture and analyze the temporal dynamics. To address this, the study uses an AI-based multimodal learning analytics (MMLA; Blikstein, 2013) tool named *Activity Map* (see Figure 1), which I co-created to quantitatively and longitudinally analyze group dynamics (Lee, Jatla, et al., under review).

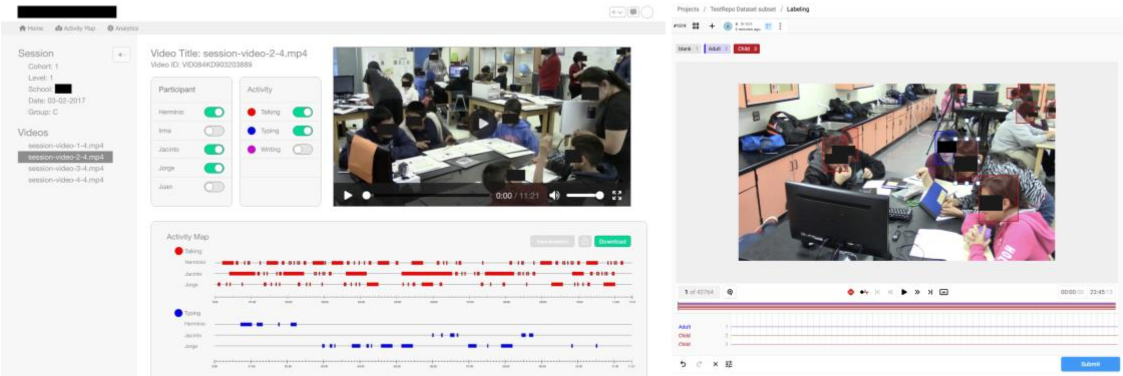


Figure 1: Activity Map Tool

This study explores collaborative group dynamics among students in the Advancing Out-of-School Learning in Mathematics and Engineering (AOLME) program (PIs: Dr. Celedón-Pattichis, Dr. Pattichis, & Dr. LópezLeiva; NSF grants #1949230, 1613637), using Activity Map. This tool tracks and visualizes students' multimodal participation—particularly focusing on talking, typing, and writing activities using AI. Grounded in sociocultural perspectives on learning, I examine learning in authentic contexts (Esmonde & Booker, 2017) and emphasize the importance of using real-world data (Cukurova et al., 2020). Specifically, I draw attention to the often-overlooked non-verbal interactions, focusing on one key object in the learning setting: the keyboard. In the program, one keyboard was shared among the members of each group, a deliberate constraint intended to enhance engagement and support students in sharing their multiple abilities (Cohen et al., 1999). The aim of this study is to improve our understanding of how shared tools such as the keyboard serve as pivotal resources and data points for analyzing power dynamics within groups and the intricate learning process. By examining these dynamics around the keyboard, we can gain insights into how authority is distributed and negotiated within groups. Authority is defined as "the probability that certain specific commands (or all commands) from a given source will be obeyed by a given group of persons" (Weber, 1947, p. 139). I conceptualize authority as a micro construct of power, examining how access to the keyboard, as detected using the Activity Map, reflects and/or mediates patterns in social and intellectual authority (Langer-Osuna et al., 2020). Understanding these dynamics can lead to better facilitation of environments that promote equitable participation, foster innovation, and support the transformative potential of collaborative learning. Therefore, the main research question guiding this study is: <u>How</u>

does keyboard access within collaborative student groups reflect and/or mediate the learning dynamics, as understood through patterns of social and intellectual authority?

2 CONCEPTUAL FRAMEWORK

This study draws on the framework for modeling the dynamics of influence (Engle et al., 2014), operationalizing influence through social and intellectual authority (Langer-Osuna et al., 2020), as well as Cultural-Historical Activity Theory (CHAT) (Cole, 1996; Engeström, 2015; Vygotsky, 1978), which together provide a comprehensive framework for understanding how individuals interact with tools and each other within a social and cultural context. An *activity system* is a complex web of interacting aspects (i.e., subject, object, tools, rules, community, division of labor) that work together in a social context to achieve a goal. Power dynamics can be explored within the historical contexts of mediational means, rules, labor divisions, and communities, and in the historical interactions between different activity systems (Engeström, 2015; Esmonde & Booker, 2017; Vygotsky, 1978). These dynamics are further mediated by the availability of historically contingent artifacts and discourses. Artifacts have material and ideal histories (Cole, 1996) and carry varying degrees of power (Esmonde & Booker, 2017; Wertsch, 1997). Paying attention to objects like a keyboard allows us to understand how they mediate human activity and connect to cultural, social, and individual aspects of learning. When tools are integrated into an activity, they create a new structure where cultural (mediated) and natural (unmediated) processes work together (Cole, 1996). The proposed method of mapping group activities over time (see Section 3.2) can be seen as an exploration of the local histories of participation. By tracing these histories, I examine how access to a shared keyboard reflects and/or mediates patterns of social and intellectual authority within collaborative educational settings (Engle et al., 2014; Esmonde & Booker, 2017; Langer-Osuna et al., 2020). Acknowledging CHAT's limitations in analyzing broader systems of power, I draw on the social and intellectual authority framework (Langer-Osuna et al., 2020) to discuss interactions between classroom practices and the ideological foundations of larger social structures.

3 METHODS

3.1 Study Context and Dataset

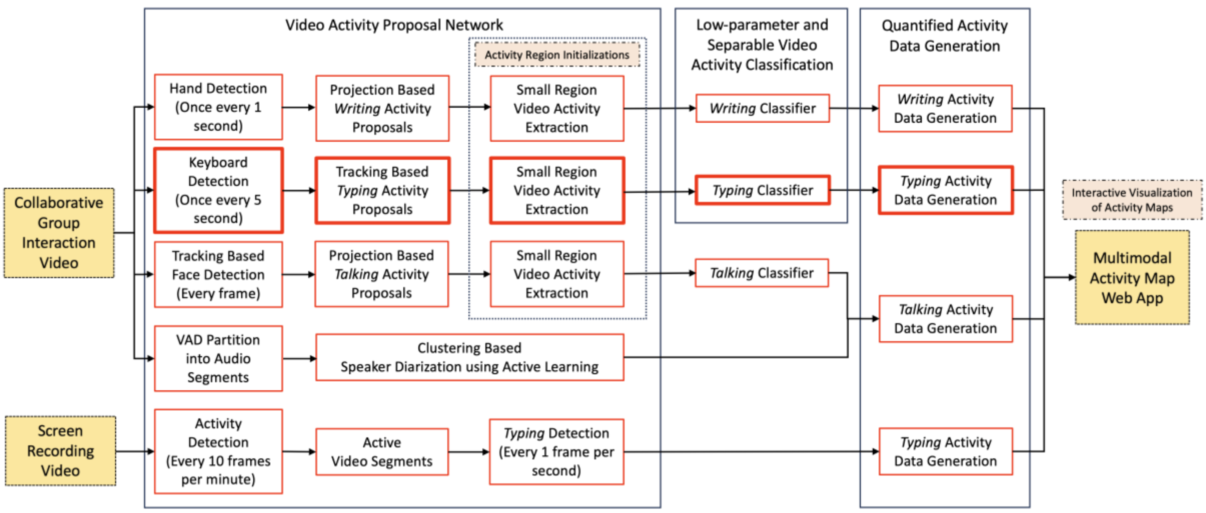


Figure 2: Workflow of the Multimodal Activity Tracking System for the Activity Map Tool *(Adapted from Lee, Jatla, et al., under review)*

The dataset, comprising approximately 2,218 hours of multimedia data collected over three years, was used to develop neural network models powering the Activity Map tool (Lee, Jatla, et al., under review; see Figures 1 & 2). It includes videos capturing collaborative group work among students in the AOLME program. This program offered a bilingual, integrated mathematics and computer programming curriculum designed to provide middle school students, especially those from underrepresented groups, with access to experiences related to STEM knowledge and practices. The program was held in two Title I middle schools in the Southwest region of the U.S., predominantly enrolling Latinx students. The curriculum is based on collaborative learning and project-based activities, with Level 1 focused on digital image and video creation, and Level 2 on object-oriented programming and robotics applications. The dataset is segmented into three yearly cohorts, each containing different levels of curriculum implementation, and further categorized by schools and student groups. Each group, consisting of 4-7 students, a facilitator, and a co-facilitator, participated in about 12 sessions per level, with each session lasting 1–3.5 hours. Students worked in teams where co-facilitators—middle school students who had previously participated in the program—co-taught with a facilitator who was an undergraduate or graduate student. This study focuses specifically on Cohort 2, Level 2 (C2L2), which entailed detailed examination of 180 hours of video recordings across 12 sessions, and analyzes activity maps created for six groups, totaling 39 individuals.

3.2 Research Design and Data Analysis

This study employs a unique mixed-methods research design (Creswell & Clark, 2017) that innovatively combines quantitative AI-based human activity data collection and analysis, using the Activity Map approach (Lee, Jatla, et al., under review), with qualitative manual coding and social interaction analysis (Jordan & Henderson, 1995). It allows for an over-time, in-depth, and nuanced exploration of how keyboard access within collaborative student groups reflects and/or mediates the learning dynamics, as understood through patterns of social and intellectual authority.

The method comprises four main phases. ***First***, I generated activity maps for the C2L2 videos using an AI tool, Activity Map, which I co-created in prior work (Lee, Jatla, et al., under review). This tool detects and quantifies typing activities (see Appendix, Figure 1) through four key stages (refer to the bolded row in Figure 2). First, the video activity segment proposal network generates candidate segments of possible typing activities. Second, optimized low-parameter dyadic 3D-CNN classifiers determine whether the activity is taking place. Third, the interactive visualization stage utilizes the quantified detection results to create an interactive visualization of typing activities in the form of an *activity map*. Fourth, the AI-human Activity Mapping cross-validation method (Lee, Jatla, et al., under review) is employed to enhance detection accuracy. These activity maps visually represent the dynamics of group interactions, particularly focusing on the distribution of keyboard access within groups. ***Second***, I synthesized activity maps by examining them over time to identify the patterns that exist within groups. Each activity map was reviewed and classified into one of three categories—*Equitable Access*, *Dominant Access*, or *Exclusive Access* (see Appendix, Table 1)—to reveal how group dynamics evolve across multiple sessions. I hypothesized that keyboard access might mirror or reflect authority within the groups, suggesting that those who primarily control the keyboard could either wield greater influence or occupy a more secretarial role. ***Third***, I refined my analysis using synthesized activity maps from the second phase as a form of data reduction. From each of the three categories identified earlier, I randomly selected two videos, totaling six sessions for detailed examination. Operationalizing social and intellectual authority (see Table 1 in Langer-Osuna et al., 2020), I manually coded the context of keyboard usage in these selected sessions to identify the types of social and intellectual work occurring in relation to the keyboard. Based on preliminary analysis of a single session, I hypothesized that the keyboard serves various functions in relation to power (see Table 2 in the Appendix for details). ***Finally,*** I synthesized the data from all previous stages to examine how social and intellectual authority are distributed within collaborative groups. Here, I explicate the potential relationships between keyboard access and social and intellectual authority in collaborative group work and analyze variations across different groups and sessions to discern patterns.

4 SCHOLARLY SIGNIFICANCE

Examining artifacts or tools in learning environments offers insights into how they mediate interaction, communication, and learning, as access to these often determines who holds authority and influence within a given context (Wertsch et al., 1993; Wertsch, 1997). In the case on one group, Keyboard access emerged as a microcosm of broader social interactions and power negotiations in the collaborative group setting. There was a marked disparity in keyboard access, with certain individuals dominating this resource. This uneven distribution correlates with these individuals' ability to steer group discussions and decision-making processes, indicating a direct link between keyboard access and social power dynamics. Notably, substantial contributions, such as coding or detailed analysis, were predominantly made by a subset of participants, reinforcing their socially negotiated degree of authority and influence within the group. This suggests that the keyboard's role extends beyond its physical utility to encompass significant sociotechnical implications in mediating access, controlling input, and shaping group interactions. From this analysis, I hypothesize that AI-identified keyboard access may be as a useful proxy for access to intellectual work, while qualitative coding can be used to identify the nature of that work related to the keyboard. I also emphasize the need for further exploration into how access to a key tool in collaborative settings reflects and/or mediates existing power structures, thereby shaping collaborative learning dynamics and outcomes.

This work contributes to the field of education both theoretically and methodologically, advancing the understanding of group interactions in ways that are relevant to learning analytics as well as sociocultural theory. The study enhances our understanding of the power dynamics surrounding the use of the keyboard, shedding light on who can author and share ideas, and, consequently, whose voices are heard and valued. Methodologically, the study introduces an innovative analytical method by integrating AI technologies with qualitative analysis, while addressing their affordances and limitations. While the qualitative framework identifies authority configurations at the group level by focusing on discursive practices, activity maps reveal which individuals consistently occupy positions of authority or exclusion, capturing nonverbal interactions. By highlighting what they illuminate about each other, this work highlights the potential of AI to contribute to the analysis of power dynamics. However, it should be approached with caution. Relying on AI-based data analysis methods, while innovative, may introduce biases or oversights if not carefully interpreted alongside qualitative analysis and cross-validation. The application of AI in this context is just one step toward a more nuanced understanding of complex power dynamics. Future research should focus on refining both the Activity Map tool and analytical methods, as well as deeply exploring the affordances of such tools to build an analytical framework that can reliably inform discussions about equity and power (Lee & Gargroetzi, 2023). I plan to leverage the tool's full capabilities, incorporating multimodal participation (see Appendix, Figure 2; Lee, Sung, et al., under review) and generative AI-based analysis (Nixon et al., 2024) in future work.

REFERENCES

Blikstein, P. (2013). Multimodal learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 102-106).

Cohen, E. G., Lotan, R. A., Scarloss, B. A., & Arellano, A. R. (1999). Complex instruction: Equity in cooperative learning classrooms. *Theory Into Practice, 38*(2), 80-86.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Harvard University Press.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.

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

Engeström, Y. (2015). *Learning by expanding: An activity-theoretical approach to developmental research.* Cambridge University Press.

Engle, R. A., Langer-Osuna, J. M., & McKinney de Royston, M. (2014). Toward a model of influence in persuasive discussions: Negotiating quality, authority, privilege, and access within a student-led argument. *Journal of the Learning Sciences, 23*, 245–268.

Esmonde, I., & Booker, A. N. (2017). *Power and privilege in the learning sciences*. Routledge.

Hübscher-Younger, T., & Narayanan, N. H. (2003). Authority and convergence in collaborative learning. *Computers & Education, 41*(4), 313-334.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, *4*(1), 39-103.

Langer-Osuna, J., Munson, J., Gargroetzi, E., Williams, I., & Chavez, R. (2020). "So what are we working on?": How student authority relations shift during collaborative mathematics activity. *Educational Studies in Mathematics, 104,* 333-349.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge university press.

Lee, H. H., & Gargroetzi, E. (2023). "It's like a double-edged sword": Mentor perspectives on ethics and responsibility in a learning analytics–supported virtual mentoring program. *Journal of Learning Analytics*, *10*(1), 85-100.

Lee, H. H., Jatla, V., Lujan, M. E., Shi, W., Egala, U., Pattichis, M. S., & Celedón-Pattichis, S. (under review). Visualizing Collaborative Learning through Activity Maps: Tracking Multimodal Participation Using an AI-Based Multimodal Activity Mapping Video Analysis System.

Lee, H. H., Sung, H., Celedón-Pattichis, S., & Pattichis, M. S. (under review). Toward an Inclusive Understanding of Collaborative Learning Using MMLA: Exploring Multimodal Participation Dynamics in Collaborative STEM Learning Context by Gender and Linguistic Diversity.

Nixon, N., Lin, Y., & Snow, L. (2024). Catalyzing Equity in STEM Teams: Harnessing Generative AI for Inclusion and Diversity. *Policy Insights from the Behavioral and Brain Sciences*, *11*(1), 85-92.

Shah, N., & Lewis, C. M. (2019). Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework. *Cognition and Instruction*, *37*(4), 423-452.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Vygotsky, L. S. (1987). Thinking and speech. In R. Rieber & A. Carton (Eds.), *The collected works of L. S. Vygotsky: Problems of general psychology* (Vol. 1, pp. 39-285) (N. Minick, Trans.). Plenum.

Weber, M. (1947). *The theory of social and economic organization*. Oxford University Press.

Wertsch, J., Tulviste, P., & Hagstrom, F. (1993). A sociocultural approach to agency. In E. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 336-356). Oxford University Press.

Wertsch, J. V. (1997). *Mind as action.* Oxford University Press.

APPENDIX

Table 1: Categorization of the Distributions of Keyboard Access

Category	Description
Equitable Access	Keyboard access is shared almost equally among all group members. This pattern suggests a collaborative environment where all individuals participate actively.
Dominant Access	Keyboard access is predominantly controlled by one or a select few members. This scenario often indicates unequal participation or suggests a hierarchical or leader-focused interaction pattern.
Exclusive Access	All members except one or a few have access to the keyboard. This pattern may suggest either marginalization of certain individuals or specialization roles within the group, highlighting potential issues of exclusion or peripheral participation.

Table 2: Hypothesized Functions of the Keyboard in Group Power Dynamics

Hypothesis	Description
Control of Input	The keyboard serves as the primary tool for entering data and commands in computer-based activities. This positions the person at the keyboard as a gatekeeper of *what* gets input into the system, thereby controlling the flow and direction of digital tasks. This control can shape the direction and outcomes of collaborative work by determining the specific data and commands that are executed.
Symbol of Technological Proficiency	Proficiency with the keyboard is often perceived as a marker of technological literacy. Individuals who are skilled in using the keyboard may be viewed as more competent or knowledgeable, which can, in turn, position them as more influential within the group. This perception can affect group dynamics by potentially establishing them as leaders or authorities in a collaborative setting.
Gatekeeper of Participation	The individual controlling the keyboard can influence *how* and *when* other group members participate. By managing access to the primary input device, they can either facilitate collaboration by encouraging input from others or hinder it by monopolizing control. This role affects the structure of participation and can centralize or distribute authority within the group.
Influence on Group Dynamics	In scenarios where a single keyboard is shared, it becomes a focal point of interaction. The way participants engage with the keyboard, from positioning themselves to negotiating its use, can affect the power dynamics within the group, influencing communication patterns and collaboration effectiveness.

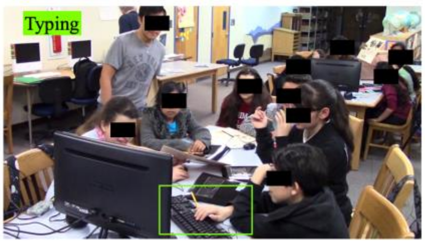


Figure 1: Capture of Sample Video Recording of Typing Activities

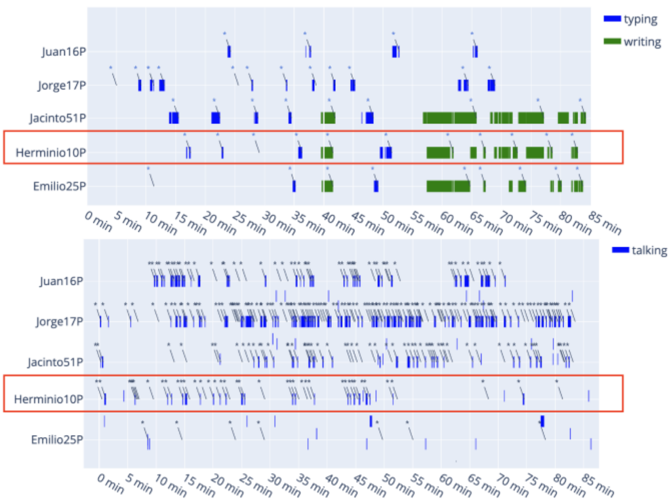


Figure 2: Sample of an Exported Activity Map

Exploring the Effectiveness of AI Generated, On-Demand Expanations within Online Learning Platforms

Author: Eamon Worden
Worcester Polytechnic Institute
elworden@wpi.edu

ABSTRACT: Since GPT-4's release it has shown novel abilities in a variety of domains. This paper explores the use of LLM-generated explanations as on-demand assistance for problems within the ASSISTments platform. In particular, we are studying whether GPT-generated explanations are better than nothing on problems that have no supports and whether GPT-generated explanations are as good as or better than teacher-authored explanations. This study contributes to existing literature since as of yet, there are no studies on the scale of ASSISTments evaluating the effectiveness of GPT support in education. Should GPT explanations prove effective then we plan to continue developing and evaluating explanations, hints, and other supports with GPT within ASSISTments. Our preliminary findings suggest that LLM authored explanations are likely better than nothing and equal to teachers. We also found only 1/415 explanations deployed to contain errors and need to be removed from our online learning platform.

Keywords: AI Generated Assistance, On-Demand Assistance, Online Learning Platform

1 INTRODUCTION

Large language models (LLMs) such as GPT-4, Llama, Claude and Gemini have recently become increasingly mainstream and demonstrate potential for education. LLMs have been used to simulate student behavior, generate supportive content, automated scoring and feedback and personalization. Despite their promise, concerns persist about LLMs' tendency to hallucinate, generate harmful content, and their lack of transparency. While research has proposed effective prompt engineering techniques for harm-reduction such as chain-of-thought prompting, few-shot learning, and self-consistency, the risk of harmful LLM-generated content reaching students remains. Nevertheless, when used responsibly, LLMs offer immense opportunities to scale educational content, saving researchers, teachers, and students time at an unprecedentedly low cost.

The goal of this work is to expand on these prior works by running an experiment with LLM-generated on-demand explanations within ASSISTments at scale. Pardos and Bhandari (2024) have previously run an experiment on GPT-generated hints withing their online learning platform (OLP). They found LLMs generated superior hints to teachers, however, compared to our study they had a significantly smaller sample size, fewer problems and used Mechanical Turk workers rather than real students. Our study gives explanations to problems, rather than hints, and goes to thousands of students using the Illustrative Mathematics curriculum. We also aim to answer two questions, first whether LLMs are better for problems where the alternative is no support, and further whether LLMs are better than teacher-generated content. As such our research questions are as follow:

1) Do LLM-generated explanations improve student learning compared to when no assistance is available?

2) How do LLM-generated explanations compare to teacher-generated explanations already within ASSISTments?

2 BACKGROUND

2.1.1 On-Demand Supports in ASSISTments

ASSISTments in an OLP which provides immediate feedback to students in two forms. First, when a student finishes a problem if it was computer gradeable (such as fill in or multiple-choice) we provide them the correctness of their response, and sometimes a wrong answer feedback message. Second, students may click on a 'Get Help' button to request either a hint or explanation, if available. Explanations are fully worked-out solutions to the given problem, typically containing step-by-step solutions and the final answer.

2.1.2 AI-Generation and Prompt Engineering

Prihar et al. used large language models to generate explanations but found a 50% error rate, however relied on GPT-3, which was then the state-of-the-art. Since then, newer models and improved prompting techniques have achieved accuracy rates of over 90% on GSM8k, a dataset of 8th-grade math word problems. However, these advancements have not yet been deployed on a scale as large as ASSISTments or compared to teacher-authored explanation. We aim to address those gaps in this study. In this paper, we use chain-of-thought prompting to enhance GPT-4's ability to solve and explain math questions. This aims to provide higher-quality explanations for students who request explanations and reduce the hallucination rate. Second, we use self-consistency to attempt to remove all explanations which contain incorrect responses.

3 METHODOLOGY

We worked with teachers, researchers, and ASSISTments employees to develop an effective prompt for authoring explanations to provide to students in ASSISTments. After numerous rounds of prompt development we settled on a separate prompt for fill-in and multiple choice questions due to the different nature of those problems. Below are the prompts for multiple-choice questions (which include "rank" questions, where students rank answers, and select all that apply questions), and fill-in questions.

Prompt 1: Generating Explanations for Multiple-Choice Questions

"Write a step by step explanation for how to solve this {problem_type} problem: {problem}. This is the correct answer: {correct_answer}. Here are all the answers: {all_answers}. Use language and ideas appropriate for {grade}. Use the active voice. Be concise. Write a step by step explanation. Do not repeat the question. Format the output with HTML. Avoid long paragraphs. Write each step as its own paragraph."

Prompt 2: Generating Explanations for Fill-in Questions

"Write a step by step explanation for how to solve this problem: {problem}. This is the correct answer: {correct_answer}. Use language and ideas appropriate for {grade}. Use the active voice. Be concise. Write a step by step explanation. Do not repeat the question. Format the output with HTML. Avoid long paragraphs. Write each step as its own paragraph."

We provide one explanation generated by GPT as it would appear for students using ASSISTments:

Step 1: *First, we need to find the difference between the actual weight of the egg and the weight it's supposed to be. In this case, the egg is supposed to weigh 2.5 oz, but it actually weighs 2.4 oz. So, subtract 2.4 oz from 2.5 oz. The difference is 0.1 oz.*

Step 2: *Next, we need to divide this difference by the weight the egg is supposed to be. This will give us the error as a decimal. So, divide 0.1 oz by 2.5 oz.*

Step 3: *The result from step 2 is a decimal. To convert this decimal into a percentage, we need to multiply it by 100.*

Step 4: *The result from step 3 is the percent error. This is the answer to the problem. In this case, the percent error is 4.*

Below is the prompt used for self-consistency to ensure the quality and correctness of GPT-authored explanations:

"The following is a step by step explanation for how to solve this {problem_type} problem: {problem}. Here is the explanation: {explanation}. Only write an explanation for part {position}. This is the correct answer: {correct_answer}. Here are all the answers: {all_answers}. The explanation should use language and ideas appropriate for {grade}. Score the explanation in ’correctness’ as a 0 if the explanation is incorrect and a 1 if the explanation is correct and makes sense and score the explanation in ’appropriateness’ as a 0 if it uses language or ideas inappropriate for the grade and a 1 if it uses language or ideas appropriate for the grade. Output the answer as a JSON."

3.1.1 Experiment 1: AI Generated Explanations vs No Explanations

We selected 145 problems from the Illustrative Math and EngageNY curriculums which were easily interpretable by ChatGPT (no images or other information), and did not have an existing teacher-authored explanation in the ASSISTments platform. We removed HTML tags from the problem bodies as so that they would not affect GPT4’s performance on math questions. We used prompt 1 to write an explanation for all ’multiple choice’, ’select all that apply’ and ’rank the options’ questions in our 145 problems. For fill-in-the-blank questions, we used prompt 2. We then used prompt 3 to identify the potentially incorrect or inappropriate prompts so we could manually remove them. We removed every explanation that GPT4 determined was either incorrect or not grade-appropriate. We then checked 10 random explanations and found each of them to be correct. We ended up with 130 explanations which were deployed into ASSISTments for experiment 1.

Students are randomized on the problem level into either the treatment group where the student can request a GPT-generated explanation, or the control group where no explanation will be available. We use the below linear model with fixed effects and robust standard errors to analyze the next problem's correctness. For student features, the average correctness of the last five problems is included in the model.

next_problem_correctness ~ control_treatment_assignment * prior_5pr_avg_correctness + Σ(problem_id)

We use next problem correctness as our outcome because within the Illustrative Curriculum, the next problem is almost always of the same skill as the prior problem. Assignments typically focus on one skill thus making the next problem correctness a viable measure for whether an explanation helped a student learn from the explanation. Further, we use prior 5 problem correctness within an assignment as that helps us determine how well a student is doing on a specific assignment, and therefore skill. We use this rather than their prior correctness as students may generally be high performing but struggling on a specific skill, or low performing but excel on a particular skill. In addition, fixed effects are added to capture the current problem. The current problem is added to control for problem-level variances, such as difficulty.

3.1.2 Experiment 2: AI Generated Assistance vs Teacher Generated Assistance

We selected 277 problems from the Illustrative and EngageNY curriculums which were easily interpretable by ChatGPT (no images or other information), and *did* have an existing teacher-authored explanation in ASSISTments. We removed HTML tags from the problem bodies. We utilized the same process and prompts 1-3 to generate explanations for these 277 problems. We determined that 233 were good enough to be deployed into ASSISTments.

Each assignment is randomized into either the treatment group, where the student can request a GPT-generated explanation, or the control group, where the student can request a teacher-generated assistance. Notably students do not know whether they are requesting a teacher or GPT-generated explanation. As the randomization is on the assignment level, the outcome will measure the average correctness of all problems in the current assignment after the problem the student views the first assistance on, as shown in the below equation.

average_correctness_after_first_assistance ~ teacher_ai_assignment * prior_avg_correctness + Σ(sequence_id:previous_problems_in_assignment_count)

For student features, the average correctness of the student prior to the assignment is included. In addition, fixed effects are added to capture the student's current progress within an assignment. For that, we use the problem set identifier (shown as `sequence id`) concatenated with the number of problems the student has already done prior to this problem within the problem set. We perform randomization on this study for two main reasons. The first is that we expect a much larger sample size. A larger sample size allows us to use a more explanatory model without losing all significance. Second we can now use the remainder of the assignment, rather than just next problem correctness which allows us to better determine whether the explanation aided the student's learning for the skill rather than just next problem which can have more variance.

4 RESULTS

4.1.1 Experiment 1: AI Generated Explanations vs No Explanations

Experiment 1 collected 2,806 problem logs across 57 problems from 1,113 students. The dataset was filtered based on whether the student had completed at least five prior problems, whether they had clicked the `Get Help' button to see the condition they were in, and whether the student has a next problem. After filtering, there were 1,195 problem logs across 49 problems from 401 students that could be analyzed.

Predictors	*Estimates*	*CI*	*p*
is treatment [TRUE]	0.04	-0.00 – 0.08	0.063
prior 5 correctness	0.11	0.08 – 0.14	**<0.001**
is treatment [TRUE] × prior 5 correctness	0.03	-0.01 – 0.08	0.173
Observations	1195		
R^2 / R^2 adjusted	0.227 / 0.193		

For **RQ1**, as shown by is_treatment, the treatment effect is barely not significant when alpha = 0.05 (CI: -0.0002175 -- 0.08211). This is further verified after refitting the model using Markov Chain Monte Carlo (MCMC) sampling using 10 chains for 5,000 iterations to further verify the results using the ratio of the posterior distribution, obtaining correctness of 0.95476%. As expected, the prior five correctness of the student is significant, as students who do better on the previous problem are more likely to get the next problem correct. To verify that there is no conflation between the condition and the features, an interaction effect is added, which shows there is no effect between the condition and features.

4.1.2 Experiment 2: AI Generated Explanations vs Teacher-Authored Explanations

Experiment 2 collected 83,631 problem logs across 232 problems from 9,362 students doing 2,950 assignments. The dataset was filtered based on whether the student had completed at least ten prior problems, whether they had requested assistance, whether the student had a next problem within the assignment and removed any instances after the first time the student requested assistance within a single assignment. For any problems the student did not complete, the student was given a 0 for that problem. After filtering, there were 9,176 problem logs across 220 problems from 3,962 students doing 1,837 assignments that could be analyzed.

Predictors	*Estimates*	*CI*	*p*
is ai condition [TRUE]	0.01	-0.02 – 0.03	0.683
prior avg correctness	0.12	0.11 – 0.14	**<0.001**
is ai condition [TRUE] × prior avg correctness	0.01	-0.01 – 0.03	0.393
Observations	9176		
R^2 / R^2 adjusted	0.699 / 0.612		

For **RQ2**, as shown by is_treatment, there is no significant difference between teacher created and LLM-generated assistance when alpha = 0.05 (CI: -0.01961 -- 0.03101). Once again, the prior average correctness of the student is significant. The interaction effect between the condition and features also shows no significant effect.

4.1.3 Student Reports on Incorrect Assistance

Across the time period where the experiments ran, we collected reports from students from both conditions to determine whether the assistance was wrong in some capacity and needed to be replaced. Students reported eleven issues (seven for teacher-authored explanations and four for LLM-generated) across eight problems. Out of the eleven issues, only one LLM-generated assistance was removed due to providing the wrong answer during the first week of the experiment. There were no other issues among the remaining ten reports.

5 DISCUSSION

In Experiment 1, our analysis revealed that the confidence interval for the effect of having GPT-generated hints available ranged from -0.0 to 0.8, suggesting that GPT-generated supports were almost certainly better than no support. We estimate that having GPT-generated supports increases the chance of getting the next problem correct by somewhere between 0% and 8%, likely 4% on average. It is likely that if the experiment ran longer, there would be a significant effect. Notably, the interaction effect between having GPT-generated supports and prior five-problem correctness is not significant. However, the positive estimate indicates that these supports may be more beneficial for students who performed better on the prior five problems.

Experiment 2 shows that GPT-generated supports and teacher-generated supports are equally useful to students. This is a very encouraging result, as it is much faster and cheaper to generate explanations with LLMs compared to asking teachers to take time to write the explanations. Additionally, the absence of an interaction effect between prior average correctness and condition indicates that these supports are equally advantageous for students, irrespective of their performance on the prior five problems.

Of the 269 LLM-generated supports in the experiment which were seen by students, only one LLM generated support was reported by students and removed due to incorrect information. Most reports were because students did not understand how to properly enter an equation within our system. This is promising as it suggests the two steps LLM review was effective in filtering out poorly written and incorrect explanations.

REFERENCES

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., & Anadkat, S. (2023). Gpt-4 technical report. *arXiv Preprint arXiv:2303.08774*.

Baral, S., Worden, E., Lim, W.-C., Luo, Z., Santorelli, C., & Gurung, A. (2024). *Automated Assessment in Math Education: A Comparative Analysis of LLMs for Open-Ended Responses*. 732–737.

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. *Technology Instruction Cognition and Learning*, *3*(1/2), 63.

Gurung, A., Baral, S., Lee, M. P., Sales, A. C., Haim, A., Vanacore, K. P., McReynolds, A. A., Kreisberg, H., Heffernan, C., & Heffernan, N. T. (2023). *How common are common wrong answers? Crowdsourcing remediation at scale*. 70–80.

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*, 470–497.

Pardos, Z. A., & Bhandari, S. (2023). Learning gain differences between ChatGPT and human tutor generated algebra hints. *arXiv Preprint arXiv:2302.06871*.

Patikorn, T., & Heffernan, N. T. (2020). *Effectiveness of crowd-sourcing on-demand assistance from teachers in online learning platforms*. 115–124.

Prihar, E., Lee, M., Hopman, M., Kalai, A. T., Vempala, S., Wang, A., Wickline, G., Murray, A., & Heffernan, N. (2023). *Comparing different approaches to generating mathematics explanations using large language models*. 290–295.

Wang, C., & Blei, D. M. (2018). *A general method for robust Bayesian modeling*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.

Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K. Z., Lasecki, W. S., & Heffernan, N. (2016). *Axis: Generating explanations at scale with learnersourcing and machine learning*. 379–388.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112.

Facilitating Effective Feedback with Human-Centred Generative AI in Higher Education

Zhiping Liang
Centre for Learning Analytics, Monash University, Melbourne, Australia
zhiping.liang@monash.edu

ABSTRACT: Feedback is a crucial aspect of higher education, essential for supporting learners in the learning process and achieving learning outcomes. The rapid development of generative AI(GenAI) has shown potential to support teaching and learning in higher education. With the increasing capabilities of GenAI in supporting teaching and learning, human educators can be augmented to provide high-quality, personalised feedback by following human-centred design paradigms. This PhD project aims to: (i) examine existing human-GenAI collaborative applications in education through a systematic review to inform the design of a human-centred feedback tool; (ii) investigate the effectiveness of GenAI in helping educators provide high-quality feedback to students; and (iii) explore how educators can collaborate with GenAI, leveraging educators' experience to enhance feedback quality. The expected outcome of this project is the development of a novel, human-centred GenAI-assisted feedback tool and empirical evidence demonstrating the effectiveness of the developed feedback tool in real-world educational scenarios.

Keywords: Generative AI, Human- centred design, Feedback Generation, Higher education

1 BACKGROUND

Feedback is a crucial part of the teaching process, and high-quality feedback is essential for enhancing the learning experience and improving student success (Butler & Winne, 1995). However, current research indicates that higher education institutions face significant challenges in meeting the expectations of both students and educators regarding feedback (Yang et al., 2010). While timely and personalised feedback can effectively support student learning, providing such feedback becomes increasingly difficult in large-scale educational settings. The challenges educators face in these environments arise not only from time and resource constraints but also from the diversity of student learning needs. These factors make the task of tailoring feedback to each student's specific needs more complex (Auerbach et al., 2018). As a result, the quality of feedback often becomes compromised, potentially negatively impacting student learning outcomes and leading to disparities in the learning experiences of different students (Fazal et al., 2011).

To meet the need for educators to provide personalised feedback to students, various methods have been developed to support effective feedback generation. Early research primarily employed expert-driven approaches, where rules were established based on expert experience, defining a series of typical student errors and providing corresponding feedback based on the match between student errors and predefined rules (Correia et al., 2017). However, this approach needs to create and maintain a vast number of expert-designed rules, which demands significant human effort, and these rules often lack generalizability (Marwan et al., 2021). With advancements in technology, data-driven automatic feedback approaches have gradually emerged. These methods include training machine learning models on student data to provide feedback based on predictive outcomes (Cavalcanti et al., 2021) or using deep learning algorithms to implicitly learn the rules between student work and expert feedback for feedback generation (Deeva et al., 2021). Although these methods have certainly enhanced the efficiency and personalization of feedback, they generally demand a substantial amount of data to achieve accuracy. Moreover, the model's performance is often restricted by the input features, which limits its generalizability.

The emergence of Generative AI (GenAI) has garnered widespread attention from researchers. With its advanced contextual understanding and real-time natural language generation capabilities, many researchers have begun exploring the application of GenAI in supporting teaching and learning. For instance, GenAI can serve as an assistant to interact with students, answering their questions about the course content or assisting with the initial grading process, allowing educators to focus more on providing detailed feedback (Jeon & Lee, 2023). However, there is currently a lack of clear methods to guide the effective adoption of GenAI in generating high-quality feedback and to validate its effectiveness within higher education. It is crucial to investigate how GenAI can be leveraged to deliver high-quality feedback that enhances student learning. A promising method for exploration is the emerging Prescriptive Learning Analytics (PLA) approach, which combines predictive models with explainable AI techniques to provide transparent insights into student progress and performance, and subsequently prescribe actions for them to take. However, previous automatic feedback generation methods may struggle to transform PLA results into high-quality feedback. Therefore, we posit that combining GenAI with PLA could generate higher-quality feedback and effectively enhance its impact. And we plan to explore the effectiveness and feasibility of integrating GenAI into the PLA framework to generate high-quality feedback.

In addition, building teachers' trust in GenAI remains a significant challenge for its practical application. A possible solution to these challenges is to engage teachers in the decision-making process when incorporating GenAI into teaching (Sun et al., 2024). This approach not only acknowledges teachers' expertise but also enhances feedback quality while minimising potential errors and biases (Shneiderman, 2022). Thus, it is crucial to explore effective collaboration strategies between GenAI and teachers in feedback generation tasks.

In my PhD project, my goal is to leverage GenAI to assist educators in providing high-quality, personalised feedback to students within the context of higher education. To achieve this, I will 1) systematically review and synthesise the latest research that includes GenAI in authentic educational settings for supporting teaching and learning; 2) evaluate the feasibility of using GenAI in combination of PLA for automated feedback generation in authentic course settings; 3) explore how to facilitate collaboration between educators and GenAI for feedback provision in higher education; The outcomes of (1-3) will result in a development of a novel, human-centred, GenAI-assisted feedback tool, and then we will 4) deploy this tool on Moodle, which is the LMS adopted at Monash University, and conduct semester-long longitudinal field studies to evaluate the impact of this feedback tool on real-world teaching practices and student learning performance. These results will be used to answer the following research questions for my PhD project:

Research **Q**uestion:

- **RQ1**: What empirical evidence and insights can be derived from the existing research literature involving human-GenAI collaboration in education?
- **RQ2**: To what extent can GenAI be combined with Prescriptive Learning Analytics to produce high-quality feedback in authentic educational settings?
- **RQ3**: To what extent, and in what ways, can human educators work with GenAI to enhance the feedback quality in higher education?
- **RQ4**: To what extent can the human-centred GenAI-assisted feedback tool improve student engagement and learning performance?

2 METHODOLOGY

2.1 Research Question 1

To answer RQ1, a systematic literature review (SLR) of the GenAI-based studies involving human-GenAI collaboration in education will be conducted. The SLR will follow the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guideline (Page et al., 2021). Although many papers have focused on the application of GenAI in education, we will place emphasis on studies that integrate GenAI to collaborate with educators or students in their real-world teaching and learning scenarios and provide empirical evidence to demonstrate the strengths or weakness of GenAI in these scenarios, as we aim to enable in-depth understanding of the dynamics during such human-GenAI collaboration process. Upon completing the systematic literature review (SLR), we will gain an in-depth understanding of the current state of empirical research on using GenAI to support teachers and students in authentic course settings. We will summarise and analyse the experimental design of these empirical research, how they provide support to participants (whether learners or educators), the empirical evidence they generate, and the potential challenges and opportunities that may arise in the future. Specifically, we use four mainstream bibliographic databases to retrieve relevant peer-reviewed publications, including Web of science, Scopus, ACM digital library, and IEEE Xplore. Besides, we use Google Scholar to retrieve newly published papers or papers from relevant venues that were not indexed in these databases. To retrieve as many relevant papers as possible, we designed the search query to consist of three groups of keywords, which required the retrieved papers to meet the following criteria: (i) using GenAI technology, (ii) focusing on educational research, (iii) GenAI functions as a collaborator or assistant, offering support to educators and learners in their specific educational practices.

2.2 Research Question 2

To answer RQ2, we will evaluate an emerging method of combining GenAI and PLA to generate readily applicable feedback. Specifically, we will perform feature engineering on the student trace data in an introductory course of data science at Monash University. To ensure that actionable insights are obtained through the PLA framework, the feature engineering process will be carried out under the guidance of an experienced lecturer. The results from the PLA framework will be integrated with the guidelines derived from the learner-centred feedback framework (Ryan et al., 2023) and used as prompt input for the GenAI to generate feedback. To further verify the potential of this method in assisting teachers with feedback writing and to evaluate the effectiveness of this automated feedback generation method, we will invite experienced teachers to conduct a comprehensive human evaluation. To assess the quality of generated feedback, we will adapt a feedback rubric based on prior studies in learner-centred feedback (Pinger et al., 2018; van der Lee et al., 2021; Jia et al., 2021), naming “Readily applicable,” “Readability,” “Relational,” and “Specificity”.

And here is the definition of these rubric:

- Readily applicable: In the best judgement of the teacher, whether the feedback could be readily applicable in the authentic course context to help student learning.
- Readability: Rate the readability of feedback concerning grammar, word choice and coherence.
- Relational: To what extent does the feedback utilise specific languages or tones to encourage students and build relationships with students.
- Specificity: To what extent is the feedback specific and pointing out areas of strengths and weakness to be improved upon.

2.3 Research Question 3

In RQ3, to investigate how the expertise of human educators can be used to collaborate with GenAI to further enhance feedback quality, we plan to conduct a controlled study on the Moodle platform. Specifically, we will categorise the feedback into four groups based on the method of writing feedback: (i) the first group will have feedback written by teachers, (ii) the second group will have feedback generated by GenAI, (iii) the third group will have feedback initially generated by GenAI and then revised by teachers based on their teaching experience to enhance the feedback quality, and (iv) the fourth group will have feedback first written by teachers and then revised by GenAI to see if it can improve the feedback quality. Through this study, we can assess the independent and combined effects on feedback generation, while also evaluating the added value of GenAI at different stages of the process and identifying which collaboration method can generate the highest quality feedback, aligning with the criteria outlined in the learner-centred feedback framework (Ryan et al., 2023). Subsequently, to comprehensively evaluate the feedback quality, we will conduct human evaluations and gather educators’ and students’ perceptions through interviews for qualitative analysis. The results of RQ3 will provide important insights for improving feedback quality and developing effective collaboration between GenAI and educators in the future.

2.4 Research Question 4

In RQ4, we plan to integrate the findings from RQ1-3 to develop and deploy a GenAI-assisted feedback tool on Moodle and incorporate it into authentic teaching practices. Therefore, RQ4 will be conducted in the third year of my PhD project, during which we will conduct semester-long longitudinal evaluation studies. The studies will use a quasi-experimental design to assess the impact of feedback on learning outcomes, focusing on the impact on student performance such as assessment scores and final course grades. Considering that the research outcomes may be influenced by the discipline in education, we will aim to conduct experiments in different disciplinary course settings to ensure the robustness of the results. In this quasi-experimental setting, students will be assigned to different study conditions without randomization to accommodate practical constraints often present in educational settings, such as class structures or institutional requirements. To assess the impact of the intervention, we plan to collect data at the beginning, middle, and end of the semester, followed by a comparison of learning outcomes and engagement among different groups within the current student cohort. This approach is designed to ensure the comprehensiveness and completeness of the experiment. By collecting and analysing the student performance and trace data, we aim to evaluate the effectiveness of the proposed GenAI-assisted feedback in improving student learning outcomes. Similarly, we will analyse students' perceptions of the intervention by collecting data through questionnaires or interviews.

3 ETHICAL CONSIDERATION

Currently, this research has obtained ethical approval from the Human Research Ethics Committee at Monash University, and the Project ID is 31325. In the subsequent research, we will update this approval if necessary to ensure compliance with ethical standards.

4 CONTRIBUTION OF SUGGESTED SOLUTION

This project will make four significant contributions to the field of Learning Analytics. First, before applying GenAI in higher education, more rigorous empirical research will be necessary. However, current studies evaluating the effectiveness of GenAI and its collaboration with humans in real teaching and learning environments remain limited. To address this gap and provide support for future research, we will conduct a comprehensive analysis of empirical studies that will test the effectiveness of GenAI in supporting teaching or learning in authentic educational scenarios. Second, we will explore the effectiveness and feasibility of combining GenAI with PLA, aiming to further enhance the quality of the generated feedback. Third, this project will explore the effective collaboration methods between GenAI and educators, advancing the integration of GenAI in teaching. Finally, we will conduct empirical studies to assess the impact of the proposed feedback tool on student course engagement and performance. Through this research, we will provide empirical evidence for this human-GenAI collaborative feedback generation approach, which will support the practical application of the method in authentic teaching environments to better assist the teaching and learning process.

5 ACHIEVED SO FAR

Concerning RQ1, we are conducting a systematic literature review to examine the existing empirical research about the incorporation of GenAI within the educational settings. Regarding RQ2, by integrating GenAI with PLA, we successfully generated readily applicable feedback based on the PLA framework. Our findings indicate that this method can accurately identify 69% of at-risk students in the authentic course setting and provide them with high-quality automated feedback. Given the high dropout and failure rate in introductory programming courses at Monash university, we posit this approach can provide personalised instruction and help pre-empt these dropouts. However, further empirical evidence is required before implementing GenAI in authentic course settings. To enhance the quality of feedback and promote the application of GenAI for feedback generation in higher education, we will explore this topic in greater depth in the study of RQ3. The results of this study have been published in the AIED 2024.

REFERENCES

Auerbach, A. J., Higgins, M., Brickman, P., & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: Expert–novice comparison reveals differences. *CBE—Life Sciences Education*, *17*(1), ar12.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245-281.

Correia, H. P., Leal, J. P., & Paiva, J. C. (2017). Enhancing feedback to students in automated diagram assessment.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, *2*, 100027.

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, *162*, 104094.

Fazal, A., Dillon, T., & Chang, E. (2011). Noise reduction in essay datasets for automated essay grading. In *On the Move to Meaningful Internet Systems: OTM 2011 Workshops: Confederated International Workshops and Posters: EI2N+ NSF ICE, ICSP+ INBAST, ISDE, ORM, OTMA, SWWS+ MONET+ SeDeS, and VADER 2011, Hersonissos, Crete, Greece, October 17-21, 2011. Proceedings* (pp. 484-493). Springer Berlin Heidelberg.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, *28*(12), 15873-15892.

Jia, Q., Cui, J., Xiao, Y., Liu, C., Rashid, P., & Gehringer, E. F. (2021). All-in-one: Multi-task learning bert models for evaluating peer assessments. *arXiv preprint arXiv:2110.03895*.

Marwan, S., Shi, Y., Menezes, I., Chi, M., Barnes, T., & Price, T. W. (2021). Just a Few Expert Constraints Can Help: Humanizing Data-Driven Subgoal Detection for Novice Programming. *International Educational Data Mining Society*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment–effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 160-182.

Sun, C., Major, L., Daltry, R., Moustafa, N., & Friedberg, A. (2024, July). Teacher-AI Collaboration in Content Recommendation for Digital Personalised Learning among Pre-primary Learners in Kenya. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale* (pp. 346-350).

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

van der Lee, C., Gatt, A., van Miltenburg, E., & Krahmer, E. (2021). Human evaluation of automatically generated text: Current trends and best practice guidelines. *Computer Speech & Language*, *67*, 101151.

YANG, M., CARLESS, D., SALTER, D., & LAM, J. (2010, June). Giving and receiving feedback: A Hong Kong perspective. In *The Temasek Polytechnic International Conference on Learning and Teaching*.

Learning to Be Inclusive through Multimodal Large Language Models

Armanda Lewis
New York University
al861@nyu.edu

ABSTRACT: Achieving inclusion has been a recognized goal within educational contexts due to inclusion's connections to improved collaboration and creativity, sustained learning, and equitable instruction. Prior research on detecting and measuring inclusion relies on qualitative observations and self reports, or on psychometric scales and surveys, though inclusion as a skill-to-be-learned is less examined. This dissertation incorporates a quasi-experimental design, where an inclusion scale is first validated based on the concepts of belonging, uniqueness, and exclusion, and mapped to observable behaviors and language. Multimodal large language models (i.e. generative AI) are developed based on these indicators and the final model is tested in an experimental design where groups of learners complete group tasks; task analyses compare inclusion quality between a control group and a model-delivered feedback group. The study's findings contribute to educational research by providing a novel multimodal observational approach for detecting inclusion and facilitating learners' improvement on inclusive skills.

Keywords: multimodal large language models, inclusion, generative AI, feedback loop, human-centered

1 DISSERTATION BACKGROUND AND AIMS

Inclusion, the degree to which individuals experience treatment from the group that satisfies their need for belonging and uniqueness (Shore, et. al., 2011), has increased in attention and stated importance within the U.S. context given changing demographics, increased political polarization, concerns over sustained patterns of historical exclusion, and the critical need to integrate different perspectives to advance some common goal. Inclusion has been linked to improved collaboration (Nembhard & Edmondson, 2006), adaptive thinking within diverse groups (Ely & Thomas, 2001), and improved creativity (Leroy, et. al., 2022). In addition to facilitating high quality group functioning, inclusion is essential for achieving dynamic knowledge building that happens within a social context, i.e. learning. Inclusion contributes to effective learning by integrating the contributions of others, acknowledging disparate perspectives, and increasing social connectedness between oneself and others. Studying inclusion within education, therefore, is a worthwhile pursuit (Barnett, 2020). We define inclusion as a combination of belonging, uniqueness, and exclusion.

1.1 Motivation

Two gaps have been identified within relevant research. (1) Prior research on detecting and measuring inclusion relies either on limited qualitative observations and self reports, or on psychometric scales and surveys. In order to scale our ability to measure inclusion, this dissertation focuses on operationalizing inclusion as a set of observable, multimodal (linguistic and gestural) markers that can be detected by large language models, a form of generative AI which can model a phenomenon, and create feedback based on the phenomenon. (2) Existing scholarship has focused on describing inclusion within groups (Dowell, et. al., 2019) and measuring team dynamics (Jackson, 2024; Reitman, et. al., 2024), though inclusion as a skill-to-be-learned is less examined. This project focuses on facilitating people's improvement on inclusive skills via an automated feedback loop.

1.2 Research Questions

As part of this work, I create multimodal large language models (MLLMs) for inclusion, and gauge the ability of MLLMs to detect inclusion quality and generate near-immediate text-based feedback. This dissertation incorporates a quasi-experimental design, where an inclusion scale is first validated and mapped to observed behaviors and language, and large language models are developed based on these observed indicators. The final model is tested in an experimental design where groups of people complete a series of tasks optimized by inclusion quality. The key research questions include:

- RQ1. Informed by theory and previously validated research, how do we measure inclusion via low level linguistic and behavioral indicators that can be interpreted by humans and large language models?
- RQ2. Which indicators significantly predict inclusion quality when measured by basic linear classifier models, unimodal large language models, and multimodal large language models?
- RQ3. What synergies and gaps do we observe between human-annotated inclusion labels and algorithmic classifications?
- RQ4. How may we leverage the multimodal large language model to provide text-based feedback that supports inclusive skills within small group exchanges?

1.3 Relevant Research and Frameworks

This project on developing and testing MLLMs to support inclusion learning sits at the necessary intersection of three fields: multimodal learning analytics, Natural Language Processing/deep learning, and educational psychology. Multimodal learning analytics, a type of educational research, distills speech, text, audio, and video data within complex learning environments, and seeks to offer rich feedback for improvement on some task or skill (Ochoa et al, 2013; Worsley & Blickstein, 2018). We take as inspiration collaboration research, which details the capture and transformation of multimodal data streams to improve on collaborative learning and acquisition of general collaboration skills (Rummel & Spada, 2005). Our focus is on inclusion as a skill-to-be-learned. NLP, the study and creation of computational systems that can automatically process and generate human language in tasks such as classification, role parsing, question-answering, and translation, has evolved to integrate sound, image and video (Tsimpoukelli, 2021; Wu, et. al., 2023). We harness and further develop such methods to support inclusion learning. Organizational and educational psychology offer insights on the importance of individual and unit-level processes that foster inclusivity within group environments. For example, applied psychology has identified individual-level benefits to inclusion such as regular communication and sense of belongingness as highly correlated with overall group inclusion and positive group functioning; in addition, scholars have identified structural interventions that foster inclusivity (Mor Barak, 2015; Roberson, 2006). We harness these theoretical foundations and measurement findings to operationalize inclusion via scalable and automatic technologies.

2 OVERVIEW OF RESEARCH DESIGN AND PROGRESS THUS FAR

In this three-study project, I focus on U.S., English speaking college students as the population of interest given the existing opportunities for small group learning exchanges and stated priorities of many higher education institutions to promote diversity, inclusion, and equity. The first study, IRB-approved and completed, establishes inclusion quality as a set of behavioral indicators informed by theory, and expert and non-expert human characterizations of inclusive exchanges. Domain identification and item generation were achieved deductively through an analytical literature review that reviewed and reconciled disciplinary interpretations of inclusion and reviewed existing scales. According to best practices for strengthening scales via a robust theoretical foundation and practical sense of usage (Boateng, et. al., 2018), I also utilized inductive methods captured via exploratory steps, where experts provided ground-up information about inclusion based on their professional experiences. Both the deductive (systematic literature review of validated inclusion literature to generate our main constructs of belonging, uniqueness, and exclusion) and inductive (expert content generation) steps were utilized to bolster content validity of inclusion and to generate item. A refined inclusion scale (see Appendix A for example) reported expert interrater reliability (Krippendorff's alpha) as $\boldsymbol{\alpha}$ = 0.925 for belonging and $\boldsymbol{\alpha}$ = 0.855 for uniqueness, and general audience interrater reliability of $\boldsymbol{\alpha}$ > .8 for both dimensions, indicating important synergies in how those with research and/or professional expertise working in inclusion initiatives, and lay audiences characterize the concept.

The next step of Study 1 involved a crowdsourcing experiment, in which a database of group interaction snippets (brief passage-level text, 15-20 second audio excerpts, and 15-20 second video segments) was compiled from open source datasets such as the AMI corpus that contains small group interactions (Carletta, 2006). Following ethical NLP data collection and storage strategies (He, et. al., 2017), human annotations of random samples taken from our corpus were gathered via Amazon Mechanical Turk crowdworkers (i.e. our general audience members), who rated the snippets according to our inclusion scale, as well as a reduced list of behavioral and linguistic markers that could be associated with inclusion. A group of 10 inclusion experts were also recruited to provide a baseline through which we could view the general audience labels. Study 1 findings are being submitted as an article on inclusion quality that links theoretical concepts to observable indicators. The paper details that experts and crowdworkers aligned closely within and across groups on exclusion and inclusion, reporting an interrater reliability of >.8 on aspects of the rubric and indicators, indicating that there are commonalities in how experts and non-experts view inclusion. Belonging showcased more mixed results, with experts consistently coding group referential positive language, open gestures, and contribution frequency as correlating with belonging, where non-experts highlighted language markers primarily.

Study 2, at an advanced stage, creates MLLMs of inclusion quality, and compares machine-generated characterizations with Study 1's human-generated characterizations. We follow for inclusion quality a similar conceptual frame for transitioning between raw data signals and collaboration quality outlined in Praharaj, Scheffel, Drachsler, and Specht (2021). Audio, visual, and physiological data are captured and manipulated to yield basic components (i.e. indicators) of inclusion, such as total speaking time and directional gaze. Indicators can then be grouped to reflect theory-based characteristics of inclusion. Automatic distillation of raw audio and video data into inclusion features include: automatic speech recognition, computational linguistic methods to clean, parse, and analyze transcribed dialogue (eg. word counts, duration, general content analysis, inclusive content analysis), detection of non-linguistic audio (speech prosody), and video signal filtering to detect person placement and basic gestures. The more advanced models can take as inputs raw data that have been batched and processed, and encode them as model features with an optimized transformer architectural design.

Baseline classification models that predict inclusion quality were built first, and then transformer models were finetuned on random holdout samples of the multimodal data snippets created for Study 1 to detect inclusion quality. The selected models (see Appendix B) -- GPT-J (open source version of GPT-3, Brown, et al, 2020), variations of Meta's Llama suite (Touvron, et. al., 2023), and NVLM 1.0 (Dai, et. al., 2024) -- are state-of-the-art autoregressive models, selected based on their unimodal and multimodal capabilities, and due to robust integrations with HuggingFace, a platform that facilitates the transparent development of machine learning models, datasets, and systems (Wolf, et. al., 2020). In all cases, additional bias detection will be implemented.

The main modeling approach in Study 2 is to develop supervised learning pipelines for detecting inclusion quality - one uses a traditional linear classification algorithm, and another leverages MLLM encoding of representation features via sequence classification (see Appendix C). The accuracy will be reflected in the area under the receiver operating characteristic curve (AUROC) score, where .5 indicates performance at the level of chance. These sets of model experiments will allow for comparisons of inclusion quality along three axes: (1) between a basic linear model and large language models; (2) between large language models of varying sizes (number of features) and modality inputs (text only versus multimodal); and (3) between human experts and machines.

Study 3, IRB-approved and awaiting a finalized inclusion quality model, is an experimental educational intervention where the best performing large language model, in terms of its ability to measure inclusion quality, is used in small group exchanges, and then generate feedback based on the exchanges via a simple text-based inclusion quality report. We explore feedback delivered from a general instructor role since: (a) instructor delivered feedback has been cited as time-consuming (Echeverria, Martinez-Maldonado, & Buckingham Shum, 2019); and (b) research indicates that the quality of expert feedback converges when compared to more disparate quality of peer feedback (Ifenthaler, 2009).

The finalized inclusion quality model will be used in a randomized controlled experiment with small groups of recruited college participants, after consent is obtained. Small groups (3 people) will be recorded performing a general activity of brainstorming or deliberation (both correlated with inclusion) at tables - two recorded sessions (separated by a week). The control groups will engage in self-directed feedback after each task and the experimental groups will receive model feedback after each task. Participants will complete a self-report based on our inclusion rubric and we will analyze the inclusion quality and feedback reports generated by the model.

This study design will be a randomized repeated measures ANOVA with four (4) time measurements. We will recruit 30 participants, with 15 randomly placed in the control group and 15 randomly placed in the treatment group. Since we are concerned with small group interactions, we will create random 5 groups of three people for each of the control and treatment groups; we will also switch the order of tasks for each session for various groups. A group sample size of 15 assumes an effect size of .80, .05 alpha, and .80 power, with much higher samples needed for a lower effect size. Since this project aims to establish future studies, and due to limited incentives and time constraints, we elect to limit the sample size and experimental conditions.

3 INTENDED CONTRIBUTIONS

As a result of my study, we will have among the first automated feedback pipelines for supporting inclusion as a skill-to-be-learned, leveraging MLLMs customized to detect inclusion quality and generate feedback. The study's findings will help identify the potentials for scaling AI support for inclusion given its recognized educational impacts, address limitations of existing research that cannot scale due to reliance on qualitative and self-report methods, and contribute to learning analytics research by providing a multimodal observational approach that complements psychometric measurements for inclusion. The study will inform uses of AI in educational downstream tasks, particularly given ethical concerns over model outputs.

REFERENCES

Barnett, R. (2020). Leading with meaning: Why diversity, equity, and inclusion matters in U. S. higher education. *Perspectives in Education*, 38(2), 20–35. https://doi.org/10.38140/pie.v38i2.4521

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6, 149.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few- Shot Learners. ArXiv:2005.14165 [Cs].

Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., Lathoud, G., Lincoln, M., Lisowska, A., McCowan, I., Post, W., Reidsma, D., & Wellner, P. (2006). The AMI Meeting Corpus: A Pre-announcement. In S. Renals & S. Bengio (Eds.), *Machine Learning for Multimodal Interaction* (Vol. 3869, pp. 28–39). Springer Berlin Heidelberg.

Dai, W., Lee, N., Wang, B., Yang, Z., Liu, Z., Barker, J., Rintamaki, T., Shoeybi, M., Catanzaro, B., & Ping, W. (2024). *NVLM: Open Frontier-Class Multimodal LLMs* (No. arXiv:2409.11402). arXiv. https://doi.org/10.48550/arXiv.2409.11402

Dowell, N. M., Lin, Y., Godfrey, A., Cho, H., & Brooks, C. (2019). Promoting inclusivity through time-dynamic discourse analysis in digitally-mediated collaborative learning. In B. McLaren & R. Luckin (Eds.), Proceedings of the 20th International Conference on Artificial Intelligence in Education (pp. 207–219). Chicago, IL: ACM.

Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019). Towards Collaboration Translucence: Giving Meaning to Multimodal Group Data. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16.

Ely, R. J., & Thomas, D. A. (2001). Cultural Diversity at Work: The Effects of Diversity Perspectives on Work Group Processes and Outcomes. *Administrative Science Quarterly*, 46(2), 229–273.

He, H., Chen, D., Balakrishnan, A., & Liang, P. (2018). Decoupling Strategy and Generation in Negotiation Dialogues. ArXiv:1808.09637 [Cs].

Ifenthaler, D. (2009). Model-based feedback for improving expertise and expert performance. Technology, Instruction, *Cognition and Learning*, 7(2), 83–101.

Jackson, M. H. (2024). Modeling of Small Groups in Computational Sciences: A Prospecting Review. *Small Group Research*, 10464964241279164. https://doi.org/10.1177/10464964241279164

Leroy, H., Buengeler, C., Veestraeten, M., Shemla, M., & Hoever, I. J. (2022). Fostering Team Creativity Through Team-Focused Inclusion: The Role of Leader Harvesting the Benefits of Diversity and Cultivating Value-In-Diversity Beliefs. *Group & Organization Management*, 47(4), 798–839.

Mor-Barak, M. E., & Cherin, D. A. (1998). A Tool to Expand Organizational Understanding of Workforce Diversity. *Administration in Social Work*, 22(1), 47–64.

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior*, 27(7), 941–966.

Praharaj, S., Scheffel, M., Schmitz, M., Specht, M., & Drachsler, H. (2021). Towards Automatic Collaboration Analytics for Group Speech Data Using Learning Analytics. *Sensors*, 21(9), 3156.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, 583–590.

Reitman, J. G., Harrison, J. L., Gorman, J. C., Lieber, R., & D'Mello, S. K. (2024). Communicative influence: A novel measure of team dynamics that integrates team cognition theory with collaborative problem solving assessment. Journal of Educational Psychology. Advance online publication. https://doi.org/10.1037/edu0000904

Roberson, Q. M. (2006). Disentangling the Meanings of Diversity and Inclusion in Organizations. *Group & Organization Management*, 31(2), 212–236.

Rummel, N., & Spada, H. (2005). Learning to Collaborate: An Instructional Approach to Promoting Collaborative Problem Solving in Computer-Mediated Settings. *Journal of the Learning Sciences*, 14(2), 201–241.

Shore, L. M., Randel, A. E., Chung, B. G., Dean, M. A., Holcombe Ehrhart, K., & Singh, G. (2011). Inclusion and Diversity in Work Groups: A Review and Model for Future Research. *Journal of Management*, 37(4), 1262–1289.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and Efficient Foundation Language Models* (No. arXiv:2302.13971). arXiv. https://doi.org/10.48550/arXiv.2302.13971

Tsimpoukelli, M., Menick, J., Cabi, S., Eslami, S. M. A., Vinyals, O., & Hill, F. (2021). Multimodal Few-Shot Learning with Frozen Language Models. ArXiv:2106.13884 [Cs].

Wolf, T., Debut, L., Sanh, V., Chaumond, J., et. al. (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing* (No. arXiv:1910.03771). https://doi.org/10.48550/arXiv.1910.03771

Worsley, M., & Blikstein, P. (2018). A Multimodal Analysis of Making. *International Journal of Artificial Intelligence in Education*, 28, 385–419.

Wu, J.; Gan,W.; Chen, Z.;Wan, S.; and Yu, P. S. 2023. Multimodal Large Language Models: A Survey. In *2023 IEEE International Conference on Big Data (BigData),* 2247– 2256.

APPENDIX

Appendix A

Example of inclusion measure: [Inclusion construct] Exclusion ⇒ [Scale item] Presence of toxic language ⇒ [Measured linguistic input] Explicit toxicity count; Explicit toxicity ratio

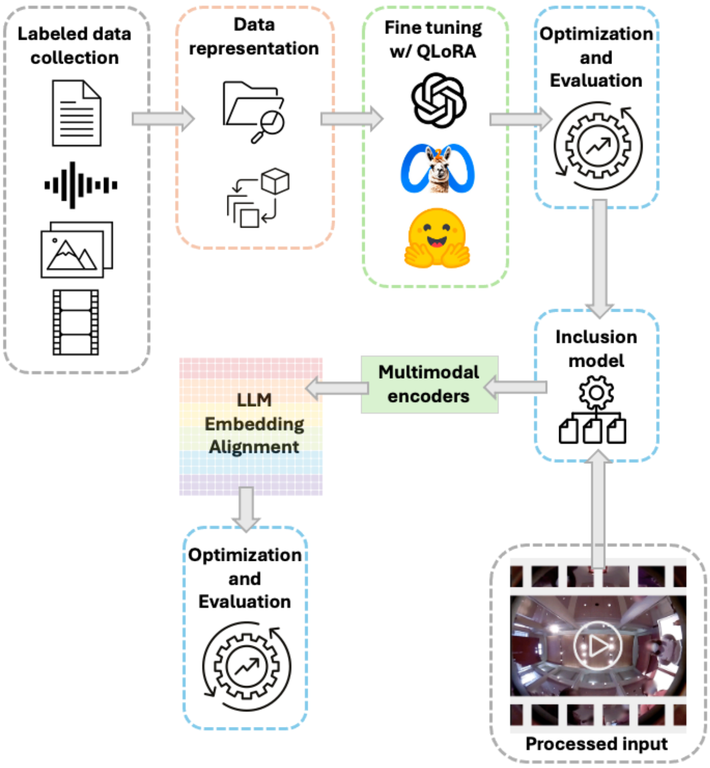
Appendix B

Table 1: Selected Large Language Models

Model	Parameters	Description
GPT-J	6 Billion	Unimodal autoregressive language model with a standard transformer architecture.
Llama 3.1-8B	8 Billion	Unimodal autoregressive language model with an optimized transformer architecture.
Llama 3.2-3B / 11B	3 / 11 Billion	Multimodal autoregressive language model with an optimized transformer architecture.
NVLM 1.0	72 Billion	Multimodal autoregressive language model. My project leverages two of the three available architectural options: decoder only and the hybrid options.

Appendix C

Abstract model pipeline for Phase 2



Beyond Single Subject: Investigating How Students Manage Multiple Subjects Under Full-Time Study Load

Yige Song
The University of Melbourne
yige.song@student.unimelb.edu.au

ABSTRACT: University students typically enrol in multiple subjects during a semester, yet most existing research focuses on analysing students' learning patterns or behaviours within a single subject. Consequently, there is limited understanding of how students strategically manage and approach learning across these subjects to meet their diverse demands. This PhD project addresses this gap through four research questions. The completed investigations on RQ1 and RQ2 examined two concurrently enrolled subjects independently, analysing survey responses and trace data collected from the learning management systems (LMS). These analyses highlight the importance of recognising individual differences and context-specific factors when interpreting students' online learning behaviours. Building on these findings, RQ3 and RQ4 will adopt a student-centred perspective by integrating trace and survey data from multiple subjects at the individual student level. This approach will enable an in-depth investigation of their learning under a full-time study load. By shifting the focus from subjects to a student-centred and holistic perspective, this research aims to provide valuable insights for academic advising and the design of course structures in higher education.

Keywords: Trace Data, Higher Education, Self-regulated Learning, Cross-subject analysis

1 INTRODUCTION

In higher education, typical full-time students enrol in multiple subjects during a semester. To learn effectively, students not only need to self-regulate their learning within each subject (Zimmerman, 2000) but also strategically manage the workload across different subjects (Gerrard et al., 2017). Therefore, to support students' learning, it is important to understand how they approach learning within individual subjects, as well as how they manage multiple subjects holistically.

Existing research has extensively focused on measuring students' self-regulated learning (SRL) and accessing learning behaviours within a single subject. Meanwhile, limited research has been conducted to investigate the nuances of workload management and the resulting cross-subject learning patterns. Learning analytics provides potential for understanding how students approach learning holistically in the university setting. Two types of data will be used in this investigation: survey data and trace data. Trace data, which is the detailed records of students' interaction with learning management systems (LMS), can be analysed to infer learning behaviours and further specific learning metrics, such as time management (Uzir et al., 2020), learning strategy (Jovanović et al., 2017) and SRL (Saint et al., 2020). Survey data, on the other hand, can be used to verify, support, and interpret the behavioural patterns observed from trace data (Li et al., 2020).

In my PhD research, I will utilise both trace data and survey data from students enrolling in overlapping, concurrently delivered subjects. Data science and learning analytics techniques will be applied to these data to reveal how full-time university students approach and manage multiple subjects. The outcome of the PhD project will fill the research gap by improving our understanding of students' holistic online learning behaviours in higher education. The research shifts the focus from a subject-centred to a student-centred, holistic perspective, with implications for personalised academic advising and guiding course structure design in university degree programs.

2 BACKGROUND: MULTI-SUBJECT LEARNING MATTERS

Learning is a complex cognitive process involving the interaction of the learner and the learning environment (Zimmerman, 2000). The process of learning is manifested by learning behaviours, some of which are reflected by the interaction with the learning management system (LMS), as captured by trace data. Existing research has been investigating the trace data to reconstruct the learning process for a better understanding of learning, the learner and the learning environment (Jovanović et al., 2017; Saint et al., 2020; Uzir et al., 2020). However, trace data typically capture the learners' learning trace within a single learning context (i.e. subject or course), rather than combining trace data from different subjects to gain a holistic, student-centred understanding of their learning experience.

This holistic understanding is crucial. For giving personalised support, it is important to understand both students' overall learning experiences as well as their subject-specific approaches. First, their subject approaches differ – our existing findings show that the same student adopts different learning strategies across subjects (Song et al., 2024), and context-specific factors should be considered when interpreting the result with learning theories (Song et al., 2025). Second, students' subject-specific approach can impact their overall learning experience. For example, research has shown that students who have coinciding assessment deadlines from different subjects are overwhelmed (Gerrard et al., 2017), potentially switching to ineffective surface learning approaches (Hattie & Donoghue, 2016), as reflected by trace data (Jovanović et al., 2017). Therefore, to identify the causes of the above "overwhelmed" phenomena for providing effective support, it is essential to gather information from multiple subjects, then integrate and process this data to gain holistic insights.

Achieving this holistic understanding of students' learning requires a shift from subject-centred to student-centred research. A key limitation of subject-centred research is its dependence on the specific learning context in which it takes place. In contrast, the student-centred approach better incorporates the complex interactions between learning contexts and individual students. For example, a study found that although students reported adjusting their learning approach across subjects, within-student consistency and between-student heterogeneity in learning approaches were observed (Vermetten et al., 1999), highlighting the need for modelling the complex interaction between learning contexts and individual differences. Until now, this student-centred approach has mostly remained at the survey level and has not become mainstream. However, with advances in technology and data collection methods, student-centred and idiographic learning analytics have emerged (Saqr et al., 2024), making this type of research more accessible and firmly grounded in theory.

3 CURRENT KNOWLEDGE AND RESEARCH GAPS

As mentioned, most existing research that uses trace data to understand students' learning behaviours focuses on a single subject. One exception is the idiographic type of research conducted by Mohammed Saqr, Sonsoles López-Pernas and colleagues. They examined the transition of the same students' learning behaviours across a sequence of subjects (Saqr, López-Pernas, & Vogelsmeier, 2023) and evaluated the consistency and variation in students' learning patterns over time (Saqr, López-Pernas, Jovanović, et al., 2023). Their findings highlight the heterogeneity of longitudinal behaviour patterns, emphasising the necessity of student-centred approach in trace data analysis.

Note that the subjects in the aforementioned studies were from the same discipline and delivered sequentially, one after the other. What remains unexplored is students' learning across concurrent subjects delivered in the same semester. It is important to emphasise that the latter is not simply an additive extension of the former for two key reasons. First, concurrent delivery requires the integration of trace data from multiple subjects, which may result in designing analytical techniques that are unprecedented in learning analytics research. Second, from students' perspective, managing multiple subjects concurrently involves strategically allocating time and resources. This requires our research to deploy analytical methods capable of capturing this nuanced complexity, which may not be immediately apparent from survey and trace data. As such, the success of this PhD project not only contributes to our understanding of students' holistic learning approaches but also advances research methodology within the learning analytics community.

4 RESEARCH GOALS AND QUESTIONS

This PhD project aims to combine trace data and survey data to investigate how full-time university students approach and manage multiple subjects. It consists of four research questions, as illustrated in Figure 1:

RQ1: How can we extract and compare the online learning patterns of university students across concurrently enrolled subjects? [completed, published as (Song et al., 2024)]

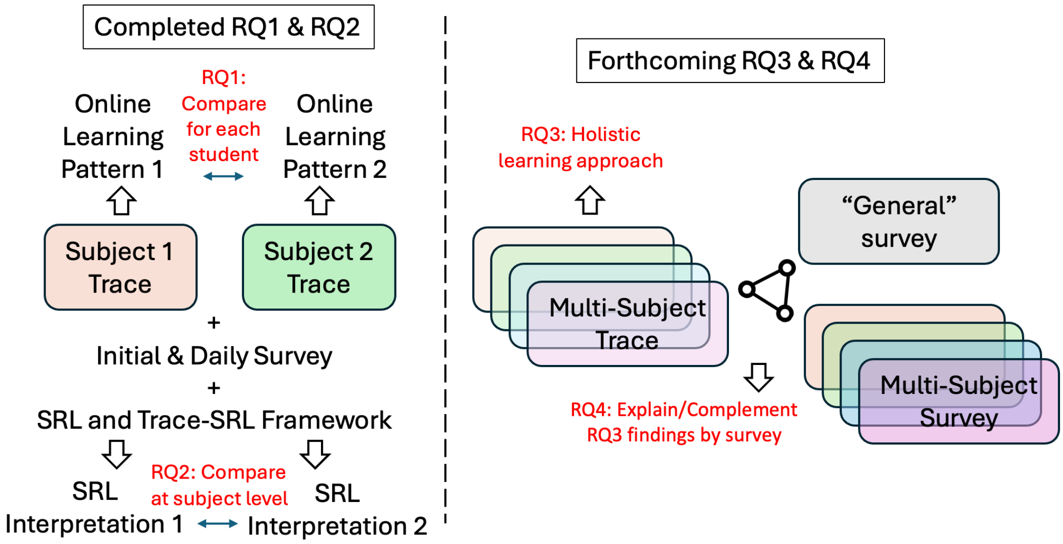
RQ2: How can we measure and validate university students' self-regulated learning (SRL) skills using trace data from concurrently enrolled subjects? [completed, published as (Song et al., 2025)]

RQ3: How do individual university students holistically manage their learning on the LMS under a full-time study load, as evidenced by trace data?

RQ4: How can survey data about students' goals, emotions and metacognition explain or complement the findings of RQ3 from a student-centred perspective?

The investigations on RQ1 and RQ2 have been published, examining the same students' trace data and survey responses in concurrently enrolled subjects. The result highlights the diversity of online learning behaviours across subjects (Song et al., 2024), as well as the challenges in developing a "one-size-fits-all" trace-SRL model to explain the subject-specific behaviours using self-regulated learning theory (Song et al., 2025). However, these investigations were conducted on different subjects independently, identifying patterns but not able to explain them from a holistic, student-centred perspective. To build upon these findings and answer RQ3 and RQ4, we will (1) integrate data from multiple subjects to obtain a holistic view rather than relying on independent, subject-specific analyses, and (2) identify behaviour patterns that are student-specific using trace data, while explaining or complementing these patterns through their survey responses.

Figure 1: Illustration of Research Questions



5 METHODOLOGY (FOR RQ3 AND RQ4)

The research employs quantitative methods. Data for RQ1 and RQ2 has already been collected (with the granted ethics application), while data for RQ3 will be collected separately, and RQ4 will utilise the data from all previous research questions. The RQ1 and RQ2 data combine trace data (from the Canvas LMS) and survey data (daily SRL surveys and an initial survey) collected in 2022 at the University of Melbourne. This data involves 39 participants and their survey and trace data across two concurrently enrolled subjects in Semester 1, 2022.

As previously mentioned, RQ1 revealed that the same student adopts different approaches of learning across different subjects. However, due to the limited sample size, the inclusion of only two subjects, and the lack of integration of trace data across subjects, we were unable to fully understand students' online learning behaviour from a holistic, student-centred perspective. Therefore, RQ3 was proposed as a continuation of this investigation. Since RQ3 relies solely on trace data, which is automatically collected, I plan to access data from at least three subjects (considered a full-time study load) with a significant overlap of students enrolled in all three during the same semester. I am currently in the process of preparing the ethics application to access this data.

The methodology for addressing RQ3 has not yet been finalised. However, a crucial and novel aspect of this investigation involves the integration and modelling of cross-subject trace data. Several potential methods are relevant for this purpose. First, since cross-subject learning behaviours result from complex interactions between students and diverse learning environments, methods used in complex systems, such as recurrent analysis, time series and network analysis, may help uncover and model these interactions (Hilpert & Marchand, 2018; Poquet et al., 2023). Second, existing techniques for trace data analysis, such as process mining and sequence analysis, can provide valuable insights by capturing the temporal and sequential patterns of students' cross-subject learning behaviours.

The investigation of RQ4 involves using surveys on students' goals, emotions, and metacognition to interpret the findings from RQ3. Considering feasibility and data access, we will reuse the survey data collected for RQ1 and RQ2. This means that after deriving a generalisable result on trace data from the RQ3 investigation, we will return to the smaller dataset and use the survey data to explain and complement the RQ3 findings. Combining survey and trace data across multiple subjects presents a significant challenge. However, network analysis techniques may enable us to integrate these data types and model their transitions and interplay throughout the semester. While the specific methodology has not yet been decided, the investigation will be conducted in an idiographic manner (Saqr et al., 2024), emphasising individual differences in students' learning experiences.

6 CURRENT PROGRESS

I started my PhD in February 2023. I have completed the experiments for RQ1 and RQ2. The result has been published in LAK and presented in different venues. I am currently designing the methodology for RQ3 and RQ4, gathering feedback, and preparing the ethics application for RQ3 data. The detailed timeline and plan are shown in Table 1.

Table 1: Research Timeline and Plan

Time	Research Focus	Research Output / Future Plans
2023	RQ1 experiment completed, paper accepted	- Poster presentation at ALASI 2023 - Poster and oral presentation at the University of Melbourne's Doctoral Colloquium - Short paper accepted by LAK24
2024	RQ2 experiment completed, paper submitted	- Oral presentation at PRACtESE SYMPOSIUM - Full paper accepted by LAK25 - Poster presentation at ALASI 2024
2025 (Plan)	RQ3 ethics approval, experiment and paper submission	- RQ3 & RQ4 investigation - Publish the result of RQ3 & RQ4 investigation
2026 (Plan)	Thesis writing	- PhD Thesis

7 CONTRIBUTION

7.1 Theoretical

The research moves beyond typical subject-centred evaluation by adopting a student-centred approach to investigate the same students' learning traces across multiple concurrently enrolled subjects. It enhances our understanding of how students holistically manage their workload and strategically approach different subjects in higher education.

7.2 Practical

Having a holistic view of how students approach and manage different subjects has important implications for academic advising. Additionally, understanding this holistic learning approach across the student body provides institutions with valuable insights to inform both the structure of degree programs and the design of individual subjects.

REFERENCES

Gerrard, D., Newfield, K., Balouchestani Asli, N., & Variawa, C. (2017). Are Students Overworked? Understanding the Workload Expectations and Realities of First-Year Engineering. *2017 ASEE Annual Conference & Exposition Proceedings*, 27612. https://doi.org/10.18260/1-2--27612

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *Npj Science of Learning*, *1*(1), 16013. https://doi.org/10.1038/npjscilearn.2016.13

Hilpert, J. C., & Marchand, G. C. (2018). Complex Systems Research in Educational Psychology: Aligning Theory and Method. *Educational Psychologist*, *53*(3), 185–202. https://doi.org/10.1080/00461520.2018.1469411

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*, 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, *45*, 100727. https://doi.org/10.1016/j.iheduc.2020.100727

Poquet, O., Jovanovic, J., & Pardo, A. (2023). Student Profiles of Change in a University Course: A Complex Dynamical Systems Perspective. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 197–207. https://doi.org/10.1145/3576050.3576077

Saint, J., Whitelock-Wainwright, A., Gasevic, D., & Pardo, A. (2020). Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning From Trace Data. *IEEE Transactions on Learning Technologies*, *13*(4), 861–877. https://doi.org/10.1109/TLT.2020.3027496

Saqr, M., Cheng, R., López-Pernas, S., & Beck, E. D. (2024). Idiographic artificial intelligence to explain students' self-regulation: Toward precision education. *Learning and Individual Differences*, *114*, 102499. https://doi.org/10.1016/j.lindif.2024.102499

Saqr, M., López-Pernas, S., Jovanović, J., & Gašević, D. (2023). Intense, turbulent, or wallowing in the mire: A longitudinal study of cross-course online tactics, strategies, and trajectories. *The Internet and Higher Education*, *57*, 100902. https://doi.org/10.1016/j.iheduc.2022.100902

Saqr, M., López-Pernas, S., & Vogelsmeier, L. V. D. E. (2023). When, how and for whom changes in engagement happen: A transition analysis of instructional variables. *Computers & Education*, *207*, 104934. https://doi.org/10.1016/j.compedu.2023.104934

Song, Y., Oliveira, E., de Barba, P., Kirley, M., & Thompson, P. (2025). Investigating Validity and Generalisability in Trace-Based Measurement of Self-Regulated Learning: A Multidisciplinary Study. *Proceedings of the 15th Learning Analytics and Knowledge Conference*. LAK 25: The 15th Learning Analytics and Knowledge Conference, Dublin Ireland. https://doi.org/10.1145/3706468.3706511

Song, Y., Oliveira, E., Kirley, M., & Thompson, P. (2024). A Case Study on University Student Online Learning Patterns Across Multidisciplinary Subjects. *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 936–942. https://doi.org/10.1145/3636555.3636939

Uzir, N. A., Gašević, D., Jovanović, J., Matcha, W., Lim, L.-A., & Fudge, A. (2020). Analytics of time management and learning strategies for effective online learning in blended environments. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 392–401. https://doi.org/10.1145/3375462.3375493

Vermetten, Y. J., Lodewijks, H. G., & Vermunt, J. D. (1999). Consistency and variability of learning strategies in different university courses. *Higher Education*, *37*(1), 1–21. https://doi.org/10.1023/A:1003573727713

Zimmerman, B. J. (2000). Attaining Self-Regulation. In *Handbook of Self-Regulation* (pp. 13–39). Elsevier. https://doi.org/10.1016/B978-012109890-2/50031-7

Investigating the Impact of Metacognitive Prompts on Strategy and Problem-Solving in Circuit Analysis in a Computer-Based Learning Environment

Jyoti Shaha
Indian Institute of Technology Bombay
Mumbai, India
jyotishaha@iitb.ac.in

ABSTRACT: Metacognitive strategies are important for successful learning in Computer-Based Learning Environments (CBLEs). However, students often struggle with the spontaneous use of these strategies. Metacognitive prompting is one of the techniques used to stimulate students' metacognitive strategies during learning activities. The existing studies on metacognitive prompts have primarily focused on reading and writing tasks in the social sciences, science, and educational psychology domains. There is a notable gap in research examining the effects of metacognitive prompts on engineering problem-solving tasks. Our research aims to address this gap by investigating the impact of metacognitive prompts on engineering students' problem-solving processes. We propose stimulating metacognitive strategies through prompts during engineering problem-solving activities and analyzing their effects on metacognitive strategies and problem-solving. We will identify learners' interaction behaviors using log data and process-mining techniques. Process mining will provide insights to understand the metacognitive strategies Additionally, the research study aims to conduct a quasi-experimental study to compare the learners' metacognitive strategies who receive prompts with those who do not.

Keywords: Metacognition, Metacognitive prompts, Metacognitive strategy, Circuit analysis, Problem-solving, CBLE

1 INTRODUCTION

In current research in Computer-based learning environments (CBLEs), learners often struggle to apply metacognitive skills spontaneously during learning and problem-solving activities. Metacognitive skills refer to the planning, monitoring, control, and evaluation processes involved in learning and problem-solving (P. Güner 2021, Jumari, N. F. 2022). Researchers often employ metacognitive prompting as an effective instructional strategy to address learners' challenges and promote self-regulated learning in CBLEs (K Engelmann, 2021). Several studies have shown that metacognitive prompts enhance learners' awareness and assist in monitoring their learning activities, improving both metacognitive skills and overall learning outcomes (E. Pieger 2018, K Engelmann 2021). Although significant research has explored the effects of metacognitive prompts in areas such as clinical reasoning, biology, educational psychology, and social sciences, most studies have focused on reading and writing tasks. Few have addressed their role in problem-solving, leaving a notable gap in the engineering problem-solving domain. This gap highlights the need for further investigation and empirical studies to assess the impact of metacognitive prompts in engineering problem-solving contexts.

2 RESEARCH PROPOSAL

2.1 Research Goal

The current research gap directed our research focus toward the following research goal: “Analyse the impact of metacognitive prompts on problem-solving and metacognitive strategies in circuit analysis problem-solving within a Computer-Based Learning Environment.” For engineering problem-solving tasks, we focused on circuit analysis problems. Students in basic electrical engineering courses often struggle with circuit analysis problems (Niebler, C. 2023).

The research plan will address the following research questions related to this goal.

RQ1- What is the impact of metacognitive prompts on circuit analysis problem-solving in CBLE?

RQ2- What is the impact of metacognitive prompts on metacognitive strategies in circuit analysis problem-solving in CBLE?

3 PROPOSED METHODOLOGY

3.1 Learning Environment

A computer-based learning environment named MetaGuru is designed to stimulate metacognitive strategies through metacognitive prompts. Figure 1 shows the screenshot of MetaGuru: Learning Environment. The learning environment consists of circuit analysis problem-solving for undergraduate engineering students. MetaGuru covers three core topics in circuit analysis: Basic Concepts of Circuits, DC Circuits, and Circuit Theorems. The content is present through text, videos, pictures, and solved examples.

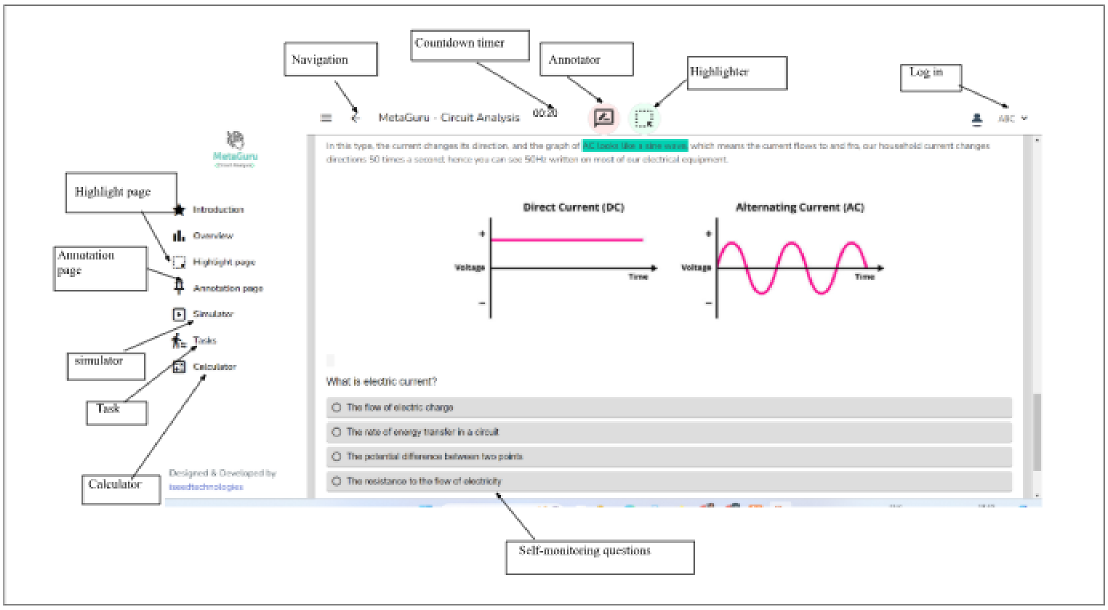


Figure 1: Screenshot of MetaGuru: Learning Environment

The environment features several interactive tools to support learning. These include a highlighter and annotator for active engagement with text, an interactive video player with embedded assessment questions, a countdown timer for time monitoring, and a circuit simulator for practically applying theoretical concepts and helping learners evaluate their answers. MetaGuru also incorporates a dedicated problem-solving task page with circuit analysis problems of varying difficulty levels.

A key feature of MetaGuru is its integration of metacognitive prompts, which are aligned with specific problem-solving steps: orientation, planning, monitoring, evaluation, and reflection. These prompts are embedded to stimulate learners' metacognitive strategies during circuit analysis problem-solving tasks. The system displays orientation and planning prompts on the task page following the problem presentation. learners encounter the evaluation prompt on the solution page prior to submitting their final solution, while the reflection prompt appears after solution submission. Monitor prompt is automatically pop-up on every content page. An example of a metacognitive prompt is, "Think and write about specific concepts, equations, or techniques needed to solve a problem successfully." This prompt aims to guide learners in the orientation phase of problem-solving to stimulate them to consider what concepts are necessary to address a given problem, encouraging them to reflect on their current knowledge and understanding of the problem.

The system is designed to collect comprehensive log data of learners' interactions, enabling researchers to analyse the impact of metacognitive prompts on problem-solving performance. This approach addresses the need for evidence-based strategies in engineering education and contributes to the broader research on metacognition in problem-solving. Table 1 gives the details of actions captured in MetaGuru and description of actions (Shaha J., 2024).

Table 1 Description of Actions captured in MetaGuru

Actions captured	Description of actions
System_access	Log in and log out to MetaGuru
Read	Reading Course material
Video	Information about the video is played, paused, and seek
Highlight	Highlighting feature is used
Annotate	Annotation feature is used
Highlight_View	The page where highlights are saved is viewed
Annotate_view	The page where annotations are saved is viewed
Calculator	Accessing calculator
Simulator	Interaction with simulator
Self_assessment_Question	Self-assessment question in video and text content is attempted
Prompt_Question	Metacognitive Prompt question is attempted
Prompt answer	Response to the prompt

3.2 Study Design

The research study plan employs a mixed-methods approach to investigate the impact of metacognitive prompts on circuit analysis problem-solving. Figure 2 shows the detailed proposed research plan. We will utilize a pre-test/post-test design with experimental and control groups to collect comprehensive quantitative and qualitative data. Initially, all participants will complete a pre-test and the Motivated Strategies for Learning Questionnaire (MSLQ) (García, T., & Pintrich, P. R. 1995). MSLQ questionnaire to evaluate students' self-reported metacognitive strategies in Likert-scale questions to quantify metacognitive strategy usage. After a brief introduction to MetaGuru: a learning environment, participants will engage in a one-hour problem-solving session. The experimental group will receive metacognitive prompts during this phase, while the control group will complete the same tasks without prompts. We will collect multiple data, including log data from MetaGuru, screen recordings of participant interactions, and participant-generated artifacts. All participants will complete a post-test and the MSLQ following the problem-solving tasks. To gain deeper insights, we will conduct 15-minute interviews with a subset of five students from each group. This structured approach will enable us to comprehensively assess metacognitive prompts' impact on problem-solving performance and metacognitive strategy use.

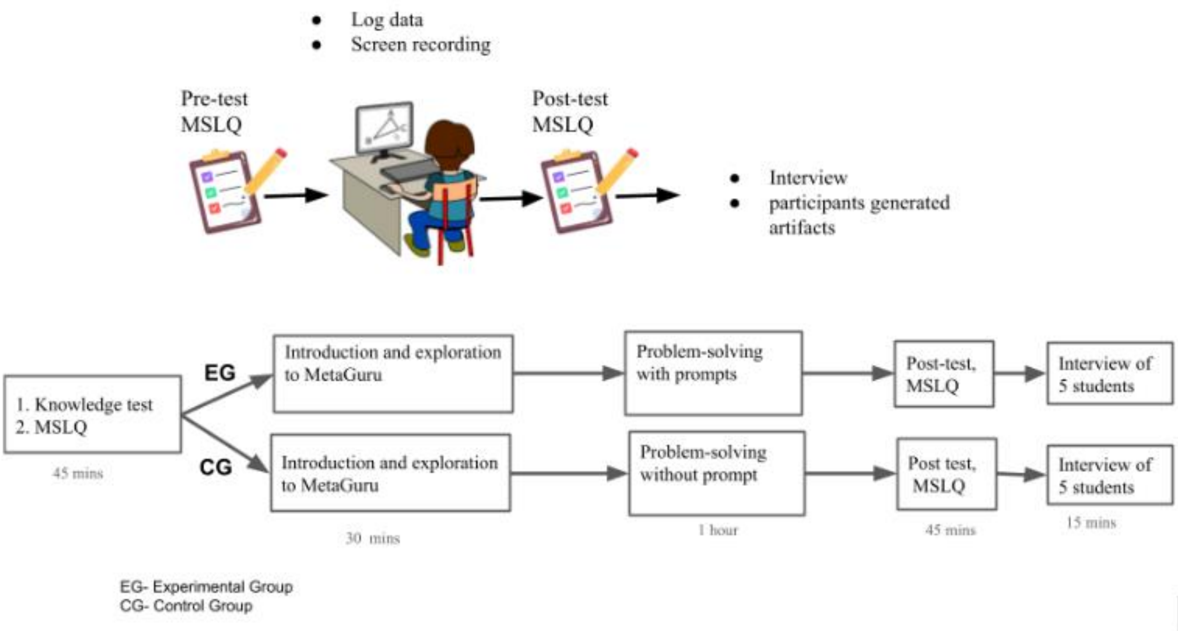


Figure 2: Proposed study design

4 CURRENT STATUS OF WORK

The development of MetaGuru, a learning environment, has progressed through multiple iterations of design, implementation, and evaluation. We conducted pilot studies to assess the validity and effectiveness of the learning environment, focusing on students' responses to metacognitive prompts and their interaction challenges. Based on the initial results of the pilot study and expert feedback, we revised MetaGuru and conducted a second pilot study. We analysed prompt functionality through initial responses, log data, and screen recordings (Shaha J., 2024). Learners' perceptions were assessed through semi-structured interviews. The key findings from the study show that MetaGuru was well-received. However, the effectiveness of metacognitive prompts varied. Participants understood and responded to the orientation, planning, and reflection prompts. While Monitoring and evaluation prompts require improvement.

We have revised the MetaGuru to address technical issues and incorporate refined metacognitive prompts.

5 PROPOSED PLAN

We aim to explore how metacognitive prompts affect learners' metacognitive processes and problem-solving in CBLEs and identify nuances in their actions. To achieve this, we plan to follow the action plan outlined below. -

1. Conducting another study with a larger sample to collect more data to examine the functionality of prompts.

2. Conducting a quasi-experimental study and collecting multimodal data to understand the influence of metacognitive prompts on learners' metacognitive strategies. The study will be conducted with approximately 60 undergraduate engineering students (30 in each group). MetaGuru will incorporate embedded metacognitive prompts for the experimental and control groups with no prompts. Interaction logs from MetaGuru will automatically record timestamps of interactions and responses to prompts. We will collect pre-test and post-test scores to assess problem-solving performance. Additionally, we will administer pre- and post-intervention surveys for knowledge test, MSLQ questionnaire to evaluate students' self-reported metacognitive strategies in Likert-scale questions to quantify metacognitive strategy usage. We will employ process mining techniques to visualize the sequence of interactions in the MetaGuru environment. This approach will help us identify a common sequence of actions taken by students and how these actions differ between the experimental and control groups. We will conduct semi-structural retrospective interviews with a subset of participants to gain deeper insights.

Our research aims to investigate the effect of Metacognitive prompts that can help in personalized scaffolding in the future.

REFERENCES

Greene, J. A., Moos, D. C., & Azevedo, R. (2011). Self-regulation of learning with computer-based learning environments. New Directions for Teaching & Learning, 2011(126).

Zheng, J., Lajoie, S. P., Wang, T., & Li, S. (2023). Supporting self-regulated learning in clinical problem-solving with a computer-based learning environment: the effectiveness of scaffolds. Metacognition and Learning, 18(3), 693-709.

Wang, T., Zheng, J., Tan, C., & Lajoie, S. P. (2023). Computer-based scaffoldings influence students' metacognitive monitoring and problem-solving efficiency in an intelligent tutoring system. Journal of Computer Assisted Learning, 39(5), 1652-1665.

Pieger, E., & Bannert, M. (2018). Differential effects of students' self-directed metacognitive prompts. Computers in Human Behavior, 86, 165-173.

Engelmann, K., Bannert, M., & Melzner, N. (2021). Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments? Research and Practice in Technology Enhanced Learning, 16(1) 3.

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short-and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. Computers in Human Behavior, 52, 293-306.

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.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American psychologist, 34(10), 906.

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and coregulation. European psychologist, 13(4), 277-287.

Güner, P., & Erbay, H. N. (2021). Metacognitive Skills and Problem-Solving. International Journal of Research in Education and Science, 7(3), 715-734.

Zhang, Q., & Lockee, B. B. (2022). Designing a Framework to Facilitate Metacognitive Strategy Development in Computer-Mediated Problem-Solving Instruction. Journal of Formative Design in Learning, 6(2), 127-143.

Marra, R. M., Hacker, D. J., & Plumb, C. (2022). Metacognition and the development of self-directed learning in a problem-based engineering curriculum. Journal of Engineering Education, 111(1), 137-161.

Murata, A., Ohta, Y., & Hayami, T. (2013). Role of metacognition in basic electric circuit problem solving process. In HCI International 2013-Posters' Extended Abstracts: International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 15 (pp. 442-446). Springer Berlin Heidelberg.

Ghazali, N. E., Yusof, K. M., Phang, F. A., Morino, H., Kamioka, E., Arsat, R., ... & Khalid, A. (2022, October). Exploring metacognitive skills among engineering students through global project-based learning. In AIP Conference Proceedings (Vol. 2433, No. 1). AIP Publishing.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. Contemporary educational psychology, 19(4), 460-475.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. Educational psychology review, 7, 351-371.

Aballe, K. S., Tabago, R. F., De Juan, E. J., Suganob, A. L., Cardaño, M. M., Pederiso, A., & Mercado, J. C. (2022). Computer Based Learning and Laboratory Based Learning in Electric Circuits: A Literature Review. International Journal of Multidisciplinary: Applied Business and Education Research, 3(7), 1349-1358.

Kumar, S. N., Lenin Fred, A., Padmanabhan, P., Gulyas, B., Dyson, C., Melba Kani, R., & Ajay Kumar, H. (2021). Multimedia-based learning tools and its scope, applications for virtual learning environment. Computational Intelligence in Digital Pedagogy, 47-63.

García, T., & Pintrich, P. R. (1995). Assessing Students' Motivation and Learning Strategies: The Motivated Strategies for Learning Questionnaire.

Ochilova, V. R. (2021). Metacognition and its history. Frontline Social Sciences and History Journal, 1(03), 18-44

Shaha J., Badhe V., Rajendran R. (2024). Investigating the functionality of metacognitive prompts during the circuit analysis problem-solving. Paper presented to International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2024)

Using BERT to Automate Rubric-based Scoring for Early Elementary Oral Narrative Retell

Author: Molly Leachman
University of California, Irvine
leachmam@uci.edu

ABSTRACT: Recent advancements in natural language processing have significantly enhanced the capacity to analyze language, extending applications from writing to spoken language. This study leverages the Bidirectional Encoder Representations from Transformers (BERT) model to automatically score oral narrative retells by Grade 2 students. The automated scoring system aims to replicate human evaluative processes using a holistic rubric that considers organizational features and narrative structure. Previous research indicates that BERT excels in scoring macrostructural narrative elements, showing high reliability compared to human raters. The present study builds on these findings by fine-tuning BERT to evaluate the oral narrative comprehension of younger children with a more nuanced rubric. Preliminary results suggest that RoBERTa and DistilRoBERTa outperform BERT and offer highly accurate, scalable, and reliable assessment scoring ($QWK_{McDonald's}$ = .98 and .99, respectively). Once complete, this tool could be applied broadly in educational and research contexts to enhance the evaluation of language development and comprehension in early childhood.

Keywords: Automated scoring, natural language processing, language assessment, early childhood

1 BACKGROUND

Recent advancements in natural language processing (NLP) and machine learning techniques have dramatically enhanced our capacity to process and analyze authentic language samples with remarkable accuracy (Kamath et al., 2022; Vaswani et al., 2017). Within education, the application of NLP tools has predominantly centered around writing: revision processes, smart tutoring, instruction, and scoring (e.g., Allen et al., 2017; Crossley et al., 2023; Ludwig et al., 2021). Written text, due to its well-defined structure and ease of conversion into a computation format, has traditionally served as a more consistent and structured dataset for training language-based machine learning models (Kamath et al., 2022). Unlike written text, spoken language, particularly during early childhood, presents unique challenges due to its inherent variability and lack of representation in many large language models derived from internet and/or literary resources.

The analysis of large corpora of essays has already led to sophisticated text analysis tools (i.e., Coh-Metrix; Graesser et al., 2004), intelligent tutoring systems (i.e., iSTART; McNamara et al., 2004), and advanced insights into complex linguistic features such as narrativity, cohesion, and word abstractness (Dowell et al., 2015; Graesser et al., 2004). Although corpora of early spoken language do exist online, these resources are nowhere near as plentiful and diverse as the amount of textual data available made available through books, articles, websites, massive open online courses, and other sources primarily produced by adolescents and adults (She & Ren, 2021). Due to this limiting factor, the field of natural language processing of early childhood language is small, but quickly growing as more data is made available and begins to encompass the complexity of early language development (Yi et al., 2024). This presents a unique opportunity to investigate early childhood language through computational processes and observe whether the scoring of children's oral narrative retell, a common language comprehension assessment type, can be automated when employing a holistic rubric scoring approach. If successful, this approach could help facilitate the analysis of extensive datasets of narrative retell responses, a task that would have previously required a considerable amount of time and human effort.

2 GOALS OF RESEARCH

The primary goal of this research is to develop an automated scoring system using a fine-tuned BERT model to assess oral narrative retell responses from Grade 2 students. This system aims to replicate and, if possible, improve upon the reliability and accuracy of human scoring using a rubric-based evaluation approach. The current study focuses on leveraging BERT's deep learning framework (Devlin et al., 2018) and contextual understanding to evaluate narrative elements such as the inclusion of characters and settings, identification of key plot points, and logical sequencing of ideas, which are essential components in assessing a child's narrative comprehension.

This study aims to answer the following specific questions:

1. Can a commonly used transformer model, BERT, be fine-tuned to score Grade 2 oral narrative retell responses using a rubric approach employing continuous scoring (0 to 25)? Are the automated scoring capabilities of BERT comparable to the scores generated by trained human raters, based on traditional measures of inter-rater reliability and scoring accuracy (i.e., ICC of .90 or higher)?
2. Do other BERT-related transformer models (e.g., RoBERTa, ALBERT, DistilBERT) perform similarly or better, in terms of accuracy and computational efficiency, than the base BERT model when analyzing and scoring Grade 2 oral retell responses?

Answering these questions will not only validate the feasibility of using advanced NLP models for scoring young children's oral narrative retell responses but will also provide a framework for automating similarly nuanced assessments in other educational and clinical settings. If successful, the study could pave the way for broader adoption of NLP tools in early childhood language development research, ultimately contributing to the understanding and support of language acquisition during critical developmental years.

3 OUTLINE OF PROBLEM AND CURRENT SOLUTIONS

Oral narrative retell assessments consist of telling a story to an individual, and asking them to tell the story back to you in as many details as they can remember. Although a relatively straightforward task, the field of educational assessment has long grappled with the challenge of reliably and efficiently scoring oral narrative retell responses. While this approach to measuring language comprehension is one of the earliest in the field of education research (Starch, 1914; Grey, 1919) and has been shown to be an informative method of observing early childhood language development across different developmental profiles (Gillam & Pearson, 2004), it is also resource intensive, especially when aiming to measure overall understanding and familiarity with narrative text structure (Shapiro et al., 2014). Traditional methods to scoring such as total word count as seen in DIBELS (Good & Kaminski, 2002) or exact recall of phrases as seen in the Test of Narrative Language (Gillam & Pearson, 2004) and the Woodcock-Johnson III (Schrank et al., 2018) have been successfully scaled due to their ease of training and implementation, but have also been questioned regarding their ability to capture the true depth of narrative comprehension and linguistic development (McNamara et al., 2015; Shapiro et al., 2014). Research has consistently shown that such methods do not account for the richer aspects of narrative story construction, and therefore miss the nuanced understanding that can be demonstrated by children during retell tasks, especially during formative years of language development (McNamara & Kendeou, 2011; McNamara et al., 2015).

The relevance and effectiveness of rubric scoring can be observed through a recent study from Ralli and colleagues (2021) which illustrated that by using the Story Grammar structure, a similar scoring guideline to the one employed in the current study, subtle patterns in growth were revealed across developmental stages, even when cohorts were merely one year apart. This approach evaluates not only the ability to recall events and specific details from the story but also introduces the importance of documenting text structure elements and overall narrative quality. These frameworks consider how children introduce characters and settings (Shapiro et al., 2014), express motivations and emotions of characters (Urbach, 2012), and follow the chronological order of the story in order to build up to an ultimate resolution (Stein & Glenn, 1975). Rubric approaches help illustrate the ways in which narrative retell offers an opportunity to peer into the thought processing and expressive capabilities of developing children in ways that more traditional methods do not (Urbach, 2012). By examining how children construct narratives and which information they include unprompted, we can acquire in-depth information about the comprehension process at an early stage of language development (Shapiro et al., 2014).

Although the conceptual strengths of rubric scoring are clear and well documented (e.g., Kendeou et al., 2008; Makinen et al., 2014; Roch et al., 2016), implementing these frameworks in practice can prove difficult. Human raters must be meticulously trained to ensure high inter-rater reliability to avoid personal bias and subjectivity when scoring, which can be both time-consuming and resource intensive. Consequently, there has been a growing interest in automating the scoring process using NLP tools to provide more consistent and scalable solutions (e.g., Jones et al., 2019).

Recent advancements in NLP, particularly with transformer models like BERT, have shown promise in handling complex language tasks, such as text summarization, sentiment analysis, and essay scoring (Vaswani et al., 2017; Devlin et al., 2018). However, the application of these models to early childhood language assessment remains limited, to our knowledge. While existing studies have successfully used BERT to score oral narrative tasks in older children, few have explored its effectiveness in younger populations, where linguistic variability and developmental factors present unique challenges (Jones et al., 2021; Karusoo-Musumeci et al., 2022).

4 INNOVATION OF CURRENT SOLUTIONS

The current study is distinct from existing approaches in its use of a fine-tuned BERT model to automate the scoring of oral narrative retell responses when using a holistic rubric-based approach for Grade 2 (mean age = 7.42) children. The ultimate goal of this study is to provide insights into narrative discourse knowledge and overall comprehension development while reducing the need for costly training across human raters and the time needed to score large oral narrative retell datasets. By employing BERT, which has demonstrated strong capabilities in semantic and contextual analysis (Devlin et al., 2018), the project aims to replicate human-level scoring fidelity to rubric guidelines while offering greater efficiency and consistency.

The base-model of BERT (Devlin et al., 2018) has been a popular choice for LLM-based tools (Qiu & Jin, 2024), but recent advancements on the base model have introduced new versions of BERT which have been shown to improve overall task accuracy, speed of processing, and/or computational cost (Cortiz, 2021). RoBERTa (Liu et al., 2019) was trained on a more robust and linguistically diverse language sample, while also making improvements to BERT's original parameters, making it a great option for processing informal language. DistilBERT (Sanh et al., 2019) distilled the BERT base model down to fewer parameters, while maintaining performance accuracy, making processing quicker and computationally lighter–an important consideration for scalability or implementation into future technologies. Lastly, ALBERT (Lan et al., 2019) also aimed to reduce the original parameters of BERT, but additionally employs a different loss function (sentence-order prediction opposed to BERT's next-sentence prediction) which focuses on inter-sentence coherence, thus making it a faster and potentially more accurate option. Considering the goal of automating oral narrative retell scoring is to improve scalability and efficiency of scoring, understanding how these alternative BERT models may improve computational efficiency and timeliness of scoring is an important contribution of the present study.

5 METHODS

The initial BERT model architecture was informed by multiple sources on GitHub and adapted for the current study's specific rubric components and differences in data structure.[1,2] All procedures and analyses were performed with Python using Jupyter Notebook. For the current dataset, there are a total of 529 students who gave oral retell responses to three different stories. Due to major differences in story features, all models were trained to score one story. Therefore, a total of 15 models were run– five models for each BERT version across the three individual stories. Additionally, the current dataset only includes the total score for each story, therefore all models were trained to calculate the total scores only (scores range from 0 to 25).

Prior to fine-tuning, all text was scanned for linguistic conventions entered by human transcribers for the original data analysis (e.g., verb tense separation: run/ing for running, pronunciation errors:

---

[1] https://github.com/ceshine/pytorch-pretrained-BERT/blob/master/notebooks/Sequence%20Regression%20Model.ipynb

[2] https://github.com/sharadkj/BERT_QWK_MISL/blob/master/BERT_QWK.ipynb

thunk:thought if child said ‘thunk’, pluralization: problem/z for problems) that would not be reflected in traditional speech-to-text AI transcription. In order to best reflect realistic speech from a range of early childhood speakers, all lexical and syntactic variations were retained in the text to test BERT’s capabilities for processing naturalistic language (i.e., ‘thunk:thought’ was left as ‘thunk’, ‘aksed:asked’ was left as aksed). Additionally, all numerical values were translated into their corresponding number words, as the text processing and tokenization for BERT requires all special characters (including numerical values) to be removed. Once all retell responses had been cleaned, each BERT variation’s specific tokenization process was employed via the Transformers package from the *HuggingFace* library.

After testing a few approaches to fine-tuning, the most success came from adding rubric-score tags to a randomly selected subsample of responses. A random subsample of 15% of the total responses were tagged using a structured coding system (e.g., [CHAR], [SET], [IE]), which provided contextual cues regarding the rubric features for the model during training. The first step of evaluation was fine-tuning BERT on a subset of the data using regression-based loss functions. During training, early stopping criteria were applied to prevent overfitting. The maximum number of epochs allowed during the first round of fine-tuning was 50, but the majority of models did not exceed 30 epochs due to the early stopping function. These models all employed a learning rate of 5e-5 and weight decay of .01. Each fine-tuned model was saved and employed for the second round of fine-tuning.

The next approach to fine-tuning consisted of a k-fold cross-validation strategy incorporating a larger percentage of the dataset. This allowed for evaluation of the model’s performance across different splits (k = 5) of the data and iterative improvement based on the results from each split. This process helps mitigate the risk of overfitting by ensuring that the model continuously generalized well to unseen data. After the cross-validation fine-tuning had been completed and saved, an unseen, randomly selected subsample of the full dataset (n = 100) was scored using the final fine-tuned model. Performance was evaluated using Intraclass Correlations (ICC), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Quadratic Weighted Kappa (QWK), and the percentage of predictions within one point of the human rater score (% Within 1-point).

6 RESULTS AND FUTURE PLANS

The full results for all BERT variations and stories are reported in Table 1. All BERT variations reached the .90 threshold for reliability (as expressed through ICC and QWK), but most notably, RoBERTa and DistilRoBERTa consistently exhibited the highest performance across all three retell stories. While the different stories did lead to some variation across model scoring ability, RoBERTa and DistilRoBERTa remained highly accurate, as displayed by their percentage of scores within one point of the human rater scores being above 80% for each story, as well as consistently low MAE and RMSE. Furthermore, for the McDonald’s and Dragon stories, these two models achieved ICC and QWK scores of .97 and higher.

One hypothesis for RoBERTa and DistilRoBERTa’s superior performance on this kind of task is that the original pre-training text used for these two models is more robust and diverse than that of the other included BERT variations. The second hypothesis is that RoBERTa employs a different tokenization process which is more flexible in conjunction with the extensive pre-training data. These features in tandem likely allow for the model to appropriately process the complexities often seen in early child language, such as differences in grammatical structures or pronunciation of words.

There are a number of next steps I would like to take based on the reviewer feedback given. Firstly, there is data available from other grades that were collected for the same project. If possible, I would like to try to score these datasets using the fine-tuned models to observe whether the model can be generalized to different age groups and not just Grade 2. Additionally, I am interested in exploring how adaptable the models are to different stories, rather than training one model to score one specific story. While the data for the current study is bound to the stories included in the Test of Narrative Language (Gillam & Pearson, 2004), it is worth exploring if a model can score a story it was not fine-tuned to score, using the rubric-informed logic from the tuning data. I hope to have explored these questions so that I may continue thinking about next steps for this project, such as creating a tool which makes this technology accessible to those who might find automated retell assessment helpful.

Table 1. Results of all tested variations of BERT

	ICC	QWK	MAE	RMSE	% Within 1-point
McDonald's					
BERT	0.95	0.95	1.11	1.48	58.17
DistilBERT	0.91	0.91	1.20	2.02	60.30
RoBERTa	0.98	0.97	0.80	1.17	85.17
DistilRoBERTa	0.99	0.98	0.70	1.09	86.20
ALBERT	0.94	0.94	1.19	1.64	68.25
Shipwreck					
BERT	0.92	0.91	1.04	1.59	74.76
DistilBERT	0.91	0.91	1.14	1.66	71.54
RoBERTa	0.94	0.93	0.87	1.41	84.63
DistilRoBERTa	0.95	0.93	0.90	1.43	83.30
ALBERT	0.90	0.90	1.15	1.69	72.11
Dragon					
BERT	0.94	0.94	1.25	1.77	67.43
DistilBERT	0.95	0.94	1.24	1.79	67.24
RoBERTa	0.98	0.97	0.84	1.24	84.48
DistilRoBERTa	0.98	0.97	0.87	1.23	82.38
ALBERT	0.95	0.92	1.45	1.99	61.11

Note. ICC = Intraclass correlation coefficient; QWK = Quadratic weighted kappa; MAE = Mean absolute error; RMSE = Root mean squared error; % within 1-point = Percentage of predicted scores within one point of the human rater scores.

References

Crossley, S., Wan, Q., Allen, L., & McNamara, D. (2023). Source inclusion in synthesis writing: an NLP approach to understanding argumentation, sourcing, and essay quality. *Reading & Writing*, *36*(4), 1053–1083. https://doi.org/10.1007/s11145-021-10221-x

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://arxiv.org/abs/1810.04805

Dowell, N. M., Graesser, A. C., & Cai, Z. (2016). Language and Discourse Analysis with Coh-Metrix: Applications from Educational Material to Learning Environments at Scale. *Journal of Learning Analytics*, *3*(3), 72–95. https://doi.org/10.18608/jla.2016.33.5

Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language.* Austin, TX: Pro-Ed.

Good, R. H.,& Kaminski, R. A. (2002). DIBELS oral reading fluency passages for first through third grades (Technical Report No. 10). Eugene, OR: University of Oregon.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. https://doi.org/10.3758/bf03195564

Gray, W. S. (1915). *Standardized Oral Reading Paragraphs.* Bloomington, IL: Public School Publishing.

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, *12*(1), 5979. https://doi.org/10.1038/s41598-022-09954-8

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R* (1st ed., Vol. 103, pp. 1–426). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Jones, K. S. (1994). Natural Language Processing: A Historical Review. In *Current Issues in Computational Linguistics: In Honour of Don Walker* (pp. 3–16). Springer Netherlands. https://doi.org/10.1007/978-0-585-35958-8_1

Jones, S., Fox, C., Gillam, S., & Gillam, R. B. (2019). An exploration of automated narrative analysis via machine learning. *PloS One*, *14*(10), e0224634–e0224634. https://doi.org/10.1371/journal.pone.0224634

Jürges, H., Makles, A. M., Naghavi, A., & Schneider, K. (2022). Melting pot kindergarten: The effect of linguistic diversity in early education. *Labour Economics*, *75*, 102119-. https://doi.org/10.1016/j.labeco.2022.102119

Ludwig, S., Mayer, C., Hansen, C., Eilers, K., & Brandt, S. (2021). Automated essay scoring using transformer models. *Psych, 3*, 897-915. https://doi.org/10.3390/psych3040056

Kamath, U., Graham, K. L., & Emara, W. (2022). *Transformers for machine learning: a deep dive* (First edition.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003170082

Karusoo-Musumeci, A., Pearce, W. M., & Donaghy, M. (2022). The effect of workshop training on rater variability in children's oral narrative assessment. *Child Language Teaching and Therapy*, *38*(1), 8–21. https://doi.org/10.1177/02656590211023839

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. Psychological Review, 95(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kudo, T. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 66–71. https://www.aclweb.org/anthology/D18-2029

Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E., & Kunnari, S. (2014). The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight-year-old Finnish children. *First Language*, *34*(1), 24–42. https://doi.org/10.1177/0142723713511000

McNamara, D., Allen, L., Crossley, S., Dascalu, M., & Perret, C. (2017). Natural Language Processing and Learning Analytics. In *The Handbook of Learning Analytics*, 93–104. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition.

McNamara, D., Jacovina, M. E., & Allen, L. (2015). Higher Order Thinking in Comprehension. In *Handbook of Individual Differences in Reading* (pp. 182–194). Routledge. https://doi.org/10.4324/9780203075562-20

McNamara, D. & Kendeou, P., (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education.*, *4*(1).

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 222–233. https://doi.org/10.3758/BF03195567

Python Software Foundation. (2020). Python Language Reference, version 3.8. Available at https://www.python.org/doc/versions/

Ralli, A. M., Kazali, E., Kanellou, M., Mouzaki, A., Antoniou, F., Diamanti, V., & Papaioannou, S. (2021). Oral Language and Story Retelling During Preschool and Primary School Years: Developmental Patterns and Interrelationships. *Journal of Psycholinguistic Research*, *50*(5), 949–965. https://doi.org/10.1007/s10936-021-09758-3

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics, 37*(1), 49–67. https://doi.org/10.1017/S0142716415000417

Rothman, D., & Gulli, A. (2022). *Transformers for natural language processing : build, train, and fine-tuning deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, and GPT-3 / Denis Rothman and Antonio Gulli.* (2nd ed.). Packt Publishing, Limited.

Shapiro, E. S., Fritschmann, N. S., Thomas, L. B., Hughes, C. L., & McDougal, J. (2014). Concurrent and Predictive Validity of Reading Retell as a Brief Measure of Reading Comprehension for Narrative Text. *Reading Psychology*, *35*(7), 644–665. https://doi.org/10.1080/02702711.2013.790328

She, T., & Ren, F. (2021). Enhance the language ability of humanoid robot NAO through deep learning to interact with autistic children. *Electronics*, *10*(19), 2393-. https://doi.org/10.3390/electronics10192393

Schrank, F. A., & Wendling, B. J. (2018). The Woodcock–Johnson IV: Tests of cognitive abilities, tests of oral language, tests of achievement. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 383–451). The Guilford Press.

Starch, D. (1914). *The Measurement of Efficiency in Reading, Writing, Spelling and English.* The College book store.

Urbach, J. (2012). Beyond story grammar: Looking at stories through cultural lenses. *Education and Urban Society*, **44**, 392–411. doi:10.1177/0013124510392567

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998-6008. https://arxiv.org/abs/1706.03762

Wong, T.-T., & Yeh, P.-Y. (2020). Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, *32*(8), 1586–1594. https://doi.org/10.1109/TKDE.2019.2912815

Yi, H., Liu, T., & Lan, G. (2024). *The key artificial intelligence technologies in early childhood education:a review*. https://doi.org/10.1007/s10462-023-10637-7

Sensemaking with Learning Analytics Feedback: Investigating the Impact of Personalized Learning Analytics Dashboards on Students' Intended Actions and Learning Strategies

Tornike Giorgashvili
Goethe University, Frankfurt am Main, Germany
giorgash@sd.uni-frankfurt.de

ABSTRACT: Effective feedback is crucial for promoting self-regulated learning (SRL), yet existing Learning Analytics Dashboards (LADs) often overlook individual differences in learner behaviour and strategy designing feedback systems and visualising data. This PhD project investigates the impact of personalised feedback on learners' sense-making, understanding, intended actions and learning strategies compared to generic feedback. Using the LIST questionnaire, which categorises students according to their cognitive learning strategies, the study examines how learners with different strategy profiles process and respond to personalised and generic feedback. By integrating personalised feedback into LADs, this research aims to enhance learners' reflective processes and promote tailored learning strategies. The study is based on both quantitative and qualitative approaches and literature review, using Epistemic Network Analysis to uncover how feedback is interpreted by different groups of learners. A systematic literature review will synthesise evidence on the impact of personalised feedback in LADs on student reflection, learning strategies, and academic performance. The findings will inform the design of more effective LADs and feedback systems, tailored to promote understanding, reflection and self-regulation in different educational contexts.

Keywords: Learning Analytics Dashboards; personalised feedback; self-regulated learning; sense-making; Epistemic Network Analysis.

1 INTRODUCTION

The term 'self-regulated learning' (SRL) describes the capacity of learners to engage actively in their own learning processes, encompassing activities such as planning, goal setting, self-monitoring and self-evaluation (Zimmerman, 1990). There is a consistent link between SRL and higher academic achievement (Broadbent & Poon, 1990). As education continues to migrate to online platforms, the need for robust SRL skills becomes even more critical. In digital learning environments, where learners often lack direct, immediate guidance from teachers, the responsibility for managing progress and adapting learning strategies falls heavily on the students. In such contexts, digital tools play an indispensable role by providing personalised feedback that not only supports reflection but actively drives the refinement of learning practices, empowering students to take control of their educational journey (Jansen et al., 2017). Learning analytics dashboards (LADs) are one such tool designed to facilitate SRL by visualising student data and providing personalised feedback on learning behaviour and progress (Schwendimann et al., 2016). These dashboards make use of vast amounts of learner data from online environments to support reflection and adjustment of learning strategies (Siemens, 2011). Despite their potential, existing LADs often fail to promote effective SRL due to a lack of grounding in learning theory and an insufficient focus on metacognition and actionable insights

(Matcha et al., 2020). A critical limitation is an inadequate understanding of how students interpret personalised feedback, what actions they intend to take, and how this feedback influences how they approach learning and their long-term choice of learning strategies. Without well-designed, pedagogically aligned feedback systems, students struggle to integrate this information into sustainable improvements in their learning (Li Chen et al., 2021; Li Chen et al., 2023).

The literature indicates that LADs can facilitate awareness and reflection (Uysal & Horzum, 2021). However, their efficacy is variable across different learner profiles. High-performing students engage more deeply with feedback, refining their self-monitoring strategies, whereas lower-performing students often fail to utilise or misinterpret the information provided (Chen et al., 2023). This emphasises the necessity for personalised feedback that is tailored to learners' cognitive strategies and motivational drivers, thereby ensuring that all students, including those with lower SRL capacities, benefit from the feedback (Salehian Kia et al., 2020). In the absence of such personalisation, LADs are likely to reinforce achievement gaps rather than address them. Furthermore, learners encounter difficulties in sensemaking, defined as the process of interpreting and acting on personalised feedback presented through LADs (Poquet (2024)). Although personalised feedback can assist students in identifying deficiencies in their performance and adapting their strategies, its efficacy is contingent upon their capacity to comprehend the information coded into visualisations and link them to their learning practices (Corrin & de Barba, 2015). For many, visualisations can be confusing, and without clear, actionable recommendations, students are unable to transform feedback into meaningful learning actions (Lim et al., 2019). Furthermore, the design of numerous LADs tends to privilege competition over mastery, prompting students to prioritise outperforming their peers over personal growth (Jivet et al., 2017; Wise, 2014). This emphasis on peer comparison can induce anxiety, particularly among students with lower performance levels, thereby further compromising their learning experience (Jivet et al., 2018). The disconnect between the principles of learning analytics and those of pedagogy in the design of learning analytics dashboards serves to compound these challenges. The current generation of LADs frequently prioritises data collection and visualisation over pedagogical alignment, which limits their capacity to foster deeper sensemaking and long-term learning adaptation. It is crucial for students to have trust in the data, positive relationships with instructors, and clear learning goals in order to make sense of the feedback provided by LADs (Klein et al., 2019). In the absence of personalised, targeted feedback, students may find LADs overwhelming, and lower-performing learners may disengage entirely, thereby missing opportunities to improve their learning strategies (Chen et al., 2023).

This PhD project investigates how students interpret personalised feedback in LADs, focusing on how they make sense of it to inform their learning strategies and intended actions. The project involves developing and evaluating an adaptive LAD that enhances sensemaking by aligning feedback with SRL principles. Using Epistemic Network Analysis (ENA) and qualitative analysis of student reflections, the research explores the cognitive processes behind interpreting personalised feedback and how learners adjust their strategies. The main question is: "How can Learning Analytics

Dashboards be designed to enhance sensemaking through personalised feedback, and what is the impact on students' intended actions and learning strategies?"

2 DESIGN AND METHOD

2.1 Theoretical foundation

Personalised feedback is widely recognised as a critical tool for fostering student engagement and improving learning outcomes, particularly in promoting SRL. Research by Wang et al. (2021) showed that feedback, especially when combined with predictions and learning suggestions, significantly increases learners' awareness, motivates them to adjust their learning strategies, and improves their performance in online learning environments. Similarly, Bulut et al. (2019) demonstrated the effectiveness of personalised feedback in increasing motivation and performance in different educational settings, especially when delivered through digital platforms. These findings are consistent with established models by Hattie and Timperley (2007) and Nicol and Macfarlane-Dick (2006), which emphasise the importance of feedback that is specific, timely, actionable and tailored to learners' individual needs in order to promote SRL.

For feedback to be truly effective, students must actively engage with it. Forsythe and Johnson (2017) emphasise the importance of reflection in processing feedback, highlighting that reflective writing improves students' ability to think critically about content and apply feedback to their learning strategies. Boud (2001) also advocates embedding reflective activities within the feedback process, to ensure that students engage meaningfully with both the feedback and the material being taught.

LADs have emerged as a key technology to support SRL by providing personalised feedback and insights into learners' progress. LADs use large datasets generated in digital learning environments to visualise learning behaviours and provide feedback aimed at promoting metacognitive processes such as self-reflection, planning and regulation (Verbert et al., 2014). Schwendimann et al. (2016) argue that LADs are particularly well suited to promoting SRL, as they allow both learners and educators to monitor learning progress in real time. However, effective dashboard design needs to be grounded in learning science principles to ensure that it not only engages learners, but also motivates them to adjust their learning strategies (Gasevic et al., 2015).

Despite the promising potential of LADs, challenges remain in terms of how learners interpret and respond to the feedback provided. The process of sensemaking, or how students interpret dashboard visualisations and integrate them into their learning practices, has not been sufficiently explored. Corrin and de Barba (2015) found that while students were able to identify gaps in their performance based on dashboard feedback, they often struggled to relate the visualised data to their existing learning strategies. This highlights the need for dashboards to provide clearer, more actionable feedback that supports students to make sense of the data.Personalisation within LADs is critical to addressing these challenges. Lim et al (2019) found that students' responses to dashboard feedback often revolved around managing their time and learning environment, suggesting that personalised dashboards are essential to support learners with diverse cognitive strategies. The need for dashboards to support sensemaking is also highlighted by Kitto et al. (2015), who argue that LADs should not only present performance data, but also actively promote key SRL skills, including metacognition and reflection. This research builds on these findings by investigating how personalised feedback delivered through LADs can improve students' sensemaking, comprehension, intended actions and learning strategies. Specifically, the project examines how personalised feedback affects different groups of learners, particularly those with different cognitive learning strategies as measured by the LIST questionnaire. In addition, the study seeks to identify evidence from the existing literature on how personalised feedback affects cognitive self-regulation and learning outcomes through a systematic meta-analysis. By addressing these gaps, this research aims to inform the development of more effective LADs that provide targeted, actionable feedback to enable students to better engage in SRL processes and ultimately improve their academic outcomes.

2.2 Research questions

To fully investigate the impact of personalised feedback delivered through learning analytics dashboards on students' understanding, intended actions and learning strategies, the following research questions were formulated:

A. Systematic literature review (**PF-SR-RQ**): What evidence from existing studies demonstrates the impact of personalised feedback in Learning Analytics Dashboards on cognitive self-regulation, learning strategies, student reflection, and academic performance?

B. Reflection and Learning Strategies (**PF-LAD-RQ**): How does personalised feedback on learning analytics dashboards influence students' reflections and the learning strategies they intend to adopt compared to generic feedback?

C. Cognitive Learning Strategies (**PF-UG-STRAT-RQ**): How does personalised feedback affect the reflection and learning behaviours of students with different cognitive learning strategies, as measured by the LIST questionnaire?

2.3 Approach and experimental design

The research project, which is scheduled to conclude in three years' time, employs a comprehensive and iterative approach to the development, implementation and evaluation of LAD feedback. The objective is to enhance students' sensemaking processes through the provision of personalised feedback. The primary objective is to examine the impact of personalised feedback delivered through LADs on students' understanding, intended actions and learning strategies, particularly in comparison to generic feedback. A systematic literature review (PF-SR-RQ) is currently underway, with publication scheduled for next year. This review will serve as the theoretical foundation of the project by synthesising existing evidence on how personalised feedback in LADs impacts cognitive self-regulation, learning strategies, student reflection, and academic performance. The findings of this review will inform subsequent stages of the project, ensuring that the development of feedback systems is grounded in proven pedagogical principles. The first key milestone of the PhD project is the publication of the paper "Exploring learners' self-reflection and intended actions after consulting learning analytics dashboards in an authentic learning setting" at the EC-TEL 2024 conference, where it has been nominated for Best Paper (Giorgashvili et al., 2024). This paper directly addresses the PF-LAD-RQ research question, examining the impact of personalised feedback on student reflection and learning strategies compared to generic feedback. The study's findings significantly shape the iterative design process of this project.

In this first phase, a LAD was developed specifically for the project. The LAD integrates personalised statistics, textual interpretations of learning analytics indicators, and prompts for self-reflection. Both rule-based algorithms and data-driven analytics, similar to the OnTask system (Pardo et al., 2018). The dashboard operationalises four key dimensions of SRL - elaboration, planning, control and regulation - based on student interaction data in the learning management system. Specifically, Elaboration - measures how often students move between pages of learning material. Planning - Tracks how much time students spend on assignments and how early they start working on them. Control - Assesses behaviours such as using the notepad during quizzes or checking quiz answers before submitting. Regulation - observes how students return to the learning material after completing self-assessment tasks.

The Dashboard consists of three main sections: A. Detailed description of the four SRL dimension feedbacked, B. Visualisation of personalised SRL statistics through bar graphs, C. Detailed and personalised feedback in text form. The feedback mechanisms, based on Hattie and Timperley's (2007) model, included both feedback (interpreting current learning behaviour) and feedforward (suggesting improvements). The dashboard was used in a Moodle-based online course for first-year teacher education students at two German universities. The study followed a randomised controlled trial design, with 742 participants divided into TG and CG. The TG received personalised feedback on their performance on the four SRL dimensions, while the CG received general cohort-based feedback. Students submitted self-reflection texts after each learning unit, providing qualitative data on their interpretation of the feedback and intended actions. The final analysis included 1,251 texts from 417 students (223 TG, 194 CG) who consistently interacted with the LAD.

To build on the EC-TEL study and address the PF-UG-STRAT-RQ, the next step involves using the LIST questionnaire to examine how personalised feedback impacts students with different cognitive learning strategies. The LIST categorises strategies like 'Goals and Planning', 'Control', 'Regulation', and 'Contexts' (Boerner et al., 2005), offering a detailed profile of each learner's approach. By using this data, learners can be grouped based on their strategies, allowing us to explore how personalised feedback in LADs affects their reflections and behaviour. This will help identify whether specific feedback types are more effective for certain cognitive strategies and refine LADs to better support SRL across diverse learner profiles. The progression from PF-SR-RQ to PF-LAD-RQ and then PF-UG-STRAT-RQ will ultimately contribute to the overarching goal of this PhD project.

3 CONCLUSION

This PhD project aims to address key gaps in the design and implementation of LADs by investigating the impact of personalised feedback on students' sense-making, understanding, intended actions and learning strategies. Existing LADs often fail to account for individual differences in cognitive learning strategies, limiting their effectiveness in promoting SRL. By investigating how learners with different strategy profiles process and respond to personalised feedback, this research aims to develop dashboards that better support reflection and tailored learning strategies. The study uses both experimental approaches and systematic meta-analysis to explore how personalised feedback affects different groups of learners, with the LIST questionnaire playing a central role in categorising cognitive strategies. Through the use of Epistemic Network Analysis, the project aims to uncover how feedback is interpreted across different learner profiles, providing visualisations of cognitive processes. The meta-analysis will synthesise existing research to provide comprehensive evidence of the impact of personalised feedback on SRL. Ultimately, the findings from this research will inform the development of more effective, personalised LADs that support comprehension, reflection and self-regulation, thereby enhancing the learning experience for students with diverse cognitive strategies. By integrating these findings into LAD design, this project aims to promote better educational outcomes through improved learner engagement and adaptation.

REFERENCES

Boud, D. (2001). Using journal writing to enhance reflective practice. New Directions for Adult and Continuing Education, (90), 9–17. https://doi.org/10.1002/ace.16

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.

Bulut, O., Cutumisu, M., Aquilina, A. M., & Singh, D. (2019). Effects of digital score reporting and feedback on students' learning in higher education. Frontiers in Education, 4.

Chen, L., Geng, X., Lu, M., Shimada, A., & Yamada, M. (2023). How students use learning analytics dashboards in higher education: A learning performance perspective. SAGE Open, 13.

Chen, L., Lu, M., Goda, Y., Shimada, A., & Yamada, M. (2021). Learning Analytics Dashboard Supporting Metacognition. IEEE Transactions on Learning Technologies.

Corrin, L., & Barba, P. G. (2015). How do students interpret feedback delivered via dashboards? In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge.

Forsythe, A., & Johnson, S. (2017). Thanks, but no-thanks for the feedback: Insights from students' and educators' reflections on feedback dialogues.

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

Giorgashvili, T. et al. (2024). Exploring Learners' Self-reflection and Intended Actions After Consulting Learning Analytics Dashboards in an Authentic Learning Setting.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1).

Jansen, R. S., Van Leeuwen, A., Janssen, J., Kester, L., & Kalz, M. (2017). Validation of the self-regulated online learning questionnaire. Journal of Computing in Higher Education, 29(1), 6-27.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. M. (2017). Awareness is not enough. Pitfalls of learning analytics dashboards in the educational practice. In Proceedings of the Seventh International Conference on Learning Analytics and Knowledge.

Jivet, I., Scheffel, M., Specht, M. M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice.

Kitto, K., Shum, S. B., & Gibson, A. (2015). Embracing imperfection in learning analytics. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (pp. 451-453).

Klein, C., Lester, J., Nguyen, T. N., Justen, A., Rangwala, H., & Johri, A. (2019). Student sensemaking of learning analytics dashboard interventions in higher education.

Lim, L., Dawson, S., Joksimovic, S., & Gašević, D. (2019). Exploring students' sensemaking of learning analytics dashboards: Does frame of reference make a difference?

Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. IEEE Transactions on Learning Technologies, 13, 226-245.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199-218.

Pardo, A., Bartimote-Aufflick, K., Buckingham Shum, S., Dawson, S., Gao, J., & Gašević, D. (2018). OnTask: Delivering data-informed, personalized learning support actions.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82(1), 33.

Poquet, O. (2024). A shared lens around sensemaking in learning analytics: What activity theory, definition of a situation and affordances can offer.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., ... & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Transactions on Learning Technologies, 10(1), 30-41.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic review of learning dashboard research. IEEE Transactions on Learning Technologies, 10(1), 30-41.

Siemens, G. (2011). Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge

Uysal, M., & Horzum, M. B. (2021). Designing and developing a learning analytics dashboard to support self-regulated learning https://doi.org/10.1007/978-3-030-68452-5_1

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

Wang, H., Huang, T., Zhao, Y., & Hu, S. (2023). The impact of dashboard feedback type on learning effectiveness, focusing on learner differences. Sustainability.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (pp. 203-211). https://doi.org/10.1145/2567574.2567588

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

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of Self-Regulation (pp. 13-40). Academic Press.

Learning Analytics from Virtual Reality (LAVR)

Martin Hlosta, Ivan Moser
Swiss Distance University of Applied Sciences (FFHS), Switzerland
{martin.hlosta, ivan.moser}@ffhs.ch

Amir Winer, Nitza Geri
Open University of Israel, Israel
{amirwi, nitzage}@openu.ac.il

Birte Heinemann, Sergej Görzen
RWTH Aachen University, Germany
{heinemann, goerzen}@informatik.rwth-aachen.de

Umesh Ramnarain, Christo van der Westhuizen, Mafor Penn
University of Johannesburg, South Africa
{uramnarain, christovdw, mpenn}@uj.ac.za

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). Building on the successful first LAVR workshop, held within LAK24 in Kyoto, the workshop aims to continue 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. The workshop strives 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, 360-Degree Videos

1 BACKGROUND

Until recently, virtual reality with head-mounted displays was confined only to research laboratories. 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). The data collected from various sensors from these devices present a rich information source 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 (Jiang et al., 2024).

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, accessibility issues or privacy and security concerns. These concerns also apply to learning analytics. The richness of the data generated from VR sensors poses additional challenges to data engineering, as well as new challenges for privacy. For example, a recent study on 55,000+ users found that the motion data from 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). Nonetheless, ongoing advancements in ethical standards and transparency guidelines are driving improvements in the responsible use of learning analytics in educational technologies (Sakr & Abdullah, 2024).

Following the successful first LAVR workshop, held within LAK24 in Kyoto, this workshop aims to serve as 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 as well as synchronous learning experiences. Although being 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
- Effective visualization of data coming from VR
- Privacy and security concerns of using LA from VR in education
- Challenges of algorithmic biases and unintended consequences of LA in VR
- Scalability, availability, and shareability aspects of LA for VR
- Data preparation and challenges of VR for LA (e.g. multiple sensor merging)
- Student/teacher acceptance and perception of using VR for LA
- LA for the design of VR environments and learning experiences
- LA for performance measurement and evaluation of learning in/with VR
- 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 the integration of VR in hybrid learning environments
- LA for empowering instructors in VR learning environments
- LA for supporting the integration of Generative AI in VR learning environments
- LA for educational 360-degree videos

2 Organizational details of the proposed event

Type of event: Mini-tracks/Symposia
Proposed duration: Half-day (expected 3.5 hours based on previous LAK workshops duration)
The workshop/tutorial activities that participants should expect: discussion groups, presentations

Proposed schedule

- 9:00 Ice-breaker and Introductions - 10 minutes
- 9:10 Keynote - 30 minutes
- 9:40 Presentation of the papers (7 submissions + small break) - including showcasing the demo of the VR application
- 11:30 Group activity - Panel discussion and possible discussions about future, opportunities and challenges of using VR and analytics in education
- 12:30 End of workshop

3 Objectives and Outcomes

We intend to continue building the LAVR research community, which was inaugurated at the first LAVR workshop, and bring 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 extend 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. To support this, we will also accept projects in their initial stages that can demonstrate their innovative potential for further analytics. To strengthen the link between research and practitioners, we will also accept demos, showcasing the combined potential of VR and analytics, without the need for rigorous evaluation required for research papers

The workshop website with information, a program and the accepted papers has been published and is available at: https://hlostam.github.io/lavr-lak25/

REFERENCES

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

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

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

Jiang, Z., Zhang, Y., & Chiang, F. K. (2024). Meta-analysis of the effect of 360-degree videos on students' learning outcomes and non-cognitive outcomes. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13464

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.

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*.

Sakr, A., & Abdullah, T. (2024). Virtual, augmented reality and learning analytics impact on learners, and educators: A systematic review. *Education and Information Technologies*, 1-50. https://doi.org/10.1007/s10639-024-12602-5

2nd Workshop on Challenges and Opportunities of Learning Analytics Adoption in Higher Education: Prerequisites for Smart and Engaging Dashboards

L.W. van Meeuwen[1], R. Conijn[1], E. Ventura-Medina[1], M. Specht[2], B. Rienties[3], O. Viberg[4], R. Kaliisa[5], I. Jivet[6]
[1] Eindhoven University of Technology, The Netherlands
[2] Delft University of Technology, The Netherlands
[3] The Open University, The United Kingdom
[4] Royal Institute of Technology, Sweden
[5] University of Oslo, Norway
[6] University of Hagen, Germany

E-mail: L.W.v.Meeuwen@TUe.nl

ABSTRACT: Despite the potential of learning analytics (LA) to enhance student learning in higher education, the adoption of learning analytics is still lagging. In this workshop, participants will share evidence and experiences with implementations of LA applications in higher education. The workshop aims to lower the threshold for a wider audience to engage with study data, thereby scaling LA. The focus is on increasing stakeholders' engagement and data availability, with an emphasis on the methods and conditions for making data available to LA stakeholders. A key area of interest is the use of Learning Analytics Dashboards (LADs), aimed, for example, at improving student self-directed and self-regulated learning by actively engaging stakeholders to use the data. The workshop will provide insights into applications of dashboards, the commonalities and differences between them, as well as the potential and the use of AI to empower LADs. It will also focus on determining the theoretical principles that are not yet implemented in practice, and what is needed to accomplish more evidence-based LAD design. Accordingly, the workshop will provide a platform for a thorough discussion on the mapping between theory and practice in LA adoption.

Keywords: LA Adoption at Scale, Learning Analytics Dashboards, LAD Design, LA and AI, Self-Regulated Learning, Self-Directed Learning.

1 INTRODUCTION

Learning Analytics (LA) can enhance learning and has the potential to increase educational quality in higher education (e.g., Viberg & Gronlund, 2023). Several higher education (HE) institutions have ambitions to exploit opportunities of LA for improving the quality of education (e.g., The Open University, 2015; Lopez-Arteaga et al., 2023). Yet adoption is still lagging (Hernandez-de-Menendez et al., 2022). In the first workshop on Challenges and Opportunities of Learning Analytics Adoption in HE (Van Meeuwen, et al., 2024) this topic was successfully approached from the perspective of 'stakeholder engagement' and 'obtaining data'. The (preliminary) analysis of last year's workshop suggests that a major opportunity for scaling LA lies in lowering the threshold for a wider audience to engage with study data in teaching and learning.

Therefore, our second workshop will focus specifically on the methods and conditions for making data available and actionable to LA stakeholders in HE. The objective is to find ways to actively engage stakeholders in HE to use these data, with the ultimate goal of improving student learning, including self-directed and self-regulated learning. In addition, we aim to find examples where AI has been applied. To this end, in this workshop we focus on displaying data within learning support systems, such as learning analytics dashboards with the particular aim to make these data engaging and actionable.

Workshop aim: Sharing knowledge and practical insights about evidence-informed interactive learning analytics dashboard design to enable student learning.

1.1 Engaging with Data

A common approach to providing students and educators with insights into learners, their learning processes, and their contexts is through the use of LA Dashboards (LADs). LADs could be implemented as stand-alone dashboards or integrated within an Intelligent Tutoring Systems or Learning Management Systems. Since 2020, research interest in LADs has significantly increased, with a notable trend of leveraging LADs to unlock the potential of LA in fostering student autonomy in learning (cf. Masiello et al., 2024). A series of systematic reviews reveal that LADs' long-term use and impact is still limited, potentially due to the shortcomings in both the design and evaluation of LADs (Jivet et al., 2018; Matcha et al., 2019; Kaliisa et al., 2024). This underlines the idea that many bumps still should be overcome towards scalable LA applications beneficial for learning (cf., Alzahrani, 2023).

In particular, the effects found of a LAD on academic performance are limited, with small to negligible effect sizes (Kaliisa et al., 2024). With some studies even showing negative effects of dashboard use (Lonn et al., 2015). These results are partly due to under-powered studies, showing the need to scale up the evaluation of LADs. In addition, it has been argued that the limited impact might be caused by the limited grounding of the LAD design in learning theories (e.g., Matcha et al., 2019). As stated by Masiello et al. (2024): "there is a need for a clear connection between the design of LADs and what educational science asserts works for learning" (p 9.).

Recently, more studies ground their LAD design into learning theories, of which self-regulated learning theory is most common (Heikkinen et al., 2022). Heikkinen et al. (2022) reported on 27 studies which used LADs to support self-regulated learning. Out of these, most studies focus on the performance phase as compared to monitoring and reflection. However, the main focus on performance might improve extrinsic motivation rather than intrinsic motivation, and might not always be actionable.

Therefore, it is necessary to focus on additional strategies, such as adding specific prompts, smart filtering of the data to be presented, letting users interact with the data, or providing targeted strategy instructions − potentially with the use of AI − to make the data actionable and useful over time (Jivet et al., 2021). In addition, (generative) AI can play an increasing role in enhancing the explainability, functionality and data aggregation in LADs (Ouyang & Zhang, 2024), resulting in potential further personalization of the dashboard and creating adaptive interventions. This workshop therefore focuses on: (1) what is needed for learning analytics dashboards to be impactful and engaging over time; (2) grounding of learning analytics dashboard design in theory.

2 WORKSHOP FORMAT AND SUBMISSIONS

To the participants, the workshop first yields insights in applications of dashboards in HE, and what they have in common. The combination of participants with a LA researchers and practitioners background will yield a thorough discussion on the overlap between theory and practice and where practice violates with theory. Second, the activities will focus on determining the theoretical principles that are not yet implemented in practice, finding out why, and gaining insights in what is needed to accomplish more evidence-based LAD design. This comes with the exchange of knowledge on specific and generic functionalities for tooling and (technical) bottlenecks for scaling up.

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 (see Table 1) includes discussions, group discussions, presentations, and voluntary contributions. We asked participants to consider a contribution detailing (case-)studies where the LAD influences students learning, specifying the LAD design process and stakeholder engagement (such as participatory design and institutional collaboration). Contributions of 10-minute demos were submitted in the form of an abstract of up to 300 words. Based on the abstract, the workshop organization carefully reviewed the submissions and selected a compelling array of diverse contributions that fit the workshop structure, workshop schedule, and enrich the workshop. Further interaction before and after the workshop is supported by a website.

2.1 Attract Participants and Community Building

The organizing committee anticipates two main groups of participants. First, scholars who contribute to the LAD domain. Second, participants who want to share more practical experiences about embedding dashboard functionalities in education and curriculum design (e.g., education designers, teaching staff, (program) managers). The organizers will use their professional networks as well as the university alliance networks of their institutions to approach both target groups. The organizers receive examples for LinkedIn posts and information to share on relevant social media and mailing lists to recruit broadly. The participants of the workshop of last year will receive a personal invitation to participate in this year's workshop. An extended website will become available as preparation, getting to know people with LAD-design interests and sharing the findings and relevant documents prior and after the workshop meeting. We are aiming for 25 – 35 participants.

Table 1 Program Design Outline

Duration	Description
20 min	Introduction: Program and recapture last year's results
20 min	Engaging LAD-design: theoretical approach
50 min	Participants show successes and failures, including own reflection on the presented cases
15 min	BREAK
45 min	Theory and Practice: structured group discussion do's and don'ts in smart LAD design
30 min	Conclusion: Characteristics of ideal dashboards; generic features, specific needs

The findings and conclusions from the workshop will be available for the LA community via the workshop website: https://sites.google.com/view/lak25adoptionworkshop/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.

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2023). Supporting self-regulated learning with learning analytics interventions–a systematic literature review. *Education and Information Technologies*, *28*(3), 3059-3088.https://doi.org/10.1007/S10639-022-11281-4

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

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018, March). 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). https://doi.org/10.1145/3170358.3170421

Jivet, I., Wong, J., Scheffel, M., Valle Torre, M., Specht, M., & Drachsler, H. (2021, April). Quantum of choice: How learners' feedback monitoring decisions, goals and self-regulated learning skills are related. In *LAK21: 11th international learning analytics and knowledge conference* (pp. 416-427). https://doi.org/10.1145/3448139.3448179

Kaliisa, R., Misiejuk, K., López-Pernas, S., Khalil, M., & Saqr, M. (2024, March). Have Learning Analytics Dashboards Lived Up to the Hype? A Systematic Review of Impact on Students' Achievement, Motivation, Participation and Attitude. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 295-304). https://doi.org/10.1145/3636555.3636884

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, *47*, 90-97. https://doi.org/10.1016/J.CHB.2014.07.013

Lopez-Arteaga, I., Koenraad, P., Rijk, K. C.H.A.M., (2023). *TU/e Vision on Education.* Eindhoven: University of Technology.

Masiello, I., Mohseni, Z., Palma, F., Nordmark, S., Augustsson, H., & Rundquist, R. (2024). A Current Overview of the Use of Learning Analytics Dashboards. *Education Sciences*, *14*(1), 82. https://doi.org/10.3390/educsci14010082

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. https://doi.org/10.1109/TLT.2019.2916802

Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educational Research Review*, *44*, 100616.

The Open University (2015). *Policy on ethical use of student data for learning analytics*. The Open University. September 2014 (2015), 1–11.

Van Meeuwen, L. W., Rienties, B., Drachsler, H., Viberg, O., Conijn, R., Vonk, C., Brijan, J. W., & Specht, M. (2024). Challenges and Opportunities of Learning Analytics Adoption in Higher Education Institutes: A European Perspective. In *Companion Proceedings of The Fourteenth International Conference on Learning Analytics & Knowledge*. Kyoto, Japan.

Viberg, O., & Grönlund, Å. (Eds.). (2023). Practicable Learning Analytics. Springer Nature.

Hybrid Intelligence: Human-AI Collaboration and Learning

Andy Nguyen[1*], Sanna Järvelä[1], Mutlu Cukurova[2], Carolyn Rose[3], Alyssa Wise[4], Michail Giannakos[5], Kshitij Sharma[5], and Luna Huynh[1]
[1] University of Oulu, Finland
[2] University College London, United Kingdom
[3] Carnegie Mellon University, United States
[4] Vanderbilt University, United States
[5] Norwegian University of Science and Technology, Norway
Andy.nguyen@oulu.fi; sanna.jarvela@oulu.fi; m.cukurova@ucl.ac.uk; cp3a@andrew.cmu.edu; alyssa.wise@vanderbilt.edu; michailg@ntnu.no; kshitij.sharma@ntnu.no; luna.huynh@oulu.fi

ABSTRACT: Hybrid Intelligence aims to enhance the collaboration between humans and machines by fostering mutual understanding and learning from each other. By leveraging the complementary strengths of both humans and AI, Hybrid Intelligence has the potential to achieve superior outcomes that neither human nor Artificial Intelligence (AI) can attain independently. This 1st workshop on Hybrid Intelligence is designed as a platform for dialogue and collaboration, highlighting the transformative potential of Hybrid Intelligence within the context of learning analytics. The workshop acts as a catalyst for enhancing the conceptualization, operationalization, and design of hybrid intelligence. By bringing together a transdisciplinary group of learning scientists, learning analytics practitioners, software engineers, and AI specialists, the workshop aims to facilitate a comprehensive exploration and envisioning of hybrid intelligence in learning analytics research and practice.

Keywords: Hybrid intelligence; artificial intelligence; educational technology; learning sciences

1 BACKGROUND AND PURPOSE

The increasing integration of Artificial Intelligence (AI) across various aspects of life is significantly transforming the educational landscape (Gašević et al., 2023; Nguyen et al., 2024; Viberg et al., 2024). Hybrid Intelligence, which focuses on fostering mutual understanding and collaborative learning between humans and machines, aims to harness the unique strengths of both in their learning processes (Järvelä et al., 2023). Unlike traditional AI, designed to operate independently in performing tasks that typically require human intelligence, such as perception and learning, Hybrid Intelligence involves active collaboration between humans and machines. This approach emphasizes continuous learning and reinforcement from both parties, setting it apart from human-centered AI. By combining the complementary strengths of humans and AI, Hybrid Intelligence holds the potential to achieve outcomes that neither could achieve alone as well as mitigating the negative repercussions of cognitive atrophy of humans (Cukurova, 2024). In this framework, learning analytics plays a pivotal role in facilitating the co-learning processes between humans and AI. Learning analytics not only empowers humans to leverage their distinct abilities—such as creative and flexible thinking, aligning actions with long-term goals and values, and making ethical decisions—but also enhances their understanding of AI. On the other hand, learning analytics enables AI to gain a deeper understanding of human learners and their educational processes (e.g., Sharma et al., 2019). Although Hybrid

Intelligence holds significant potential, it remains an emerging concept, with the development of its conceptualization and operationalization in the context of learning analytics still in its early stages. It is also crucial to identify the indicators of both effective and ineffective human-AI collaboration as they manifest and to understand how these dynamics may evolve over time. This presents challenges in designing effective Hybrid Intelligence systems for educational contexts, highlighting the need for further exploration and innovation in this area. Accordingly, the main purpose of this workshop is to ignite discussions and establish a research agenda for Hybrid Intelligence in learning analytics by bringing together a multidisciplinary group of learning scientists, learning analytics practitioners, software engineers, and AI specialists.

2 ORGANISATIONAL DETAILS

2.1 Workshop format and participants

The workshop is planned as a half-day, in-person event, with a capacity for 15 to 30 participants. It aims to attract a diverse group of attendees, including learning scientists, learning analytics practitioners, software engineers, and AI specialists, all focused on the applications of hybrid intelligence in learning analytics research and practices. The workshop welcomes participants of all skill levels, encouraging them to bring prototype concepts or early-stage projects related to Hybrid Intelligence for discussions and collaborative activities during the session.

2.2 Pre-workshop calls for papers and activities

Once this workshop proposal is approved, a call for papers will be issued to encourage more detailed contributions in this area. Submissions, ranging from 2-4 pages, will be reviewed by the organizing committee and the authors. Workshop participants will have access to both the submitted and accepted papers ahead of the event, enabling them to engage in informed discussions. Additionally, participants will be asked to complete a pre-workshop survey designed to collect information on their previous experiences with Hybrid Intelligence and their potential ideas for its application in research. This data will help lay the groundwork for discussions and collaborations during the workshop.

2.3 Workshop schedule and activities

The workshop is scheduled to take place as part of the pre-conference activities of the main conference and will follow a half-day format lasting up to 4 hours. The schedule is summarized in Table 1 and detailed below.

Table 1: Workshop Agenda

Duration	Activity	Contributor(s)
10 minutes	Welcome & Introduction	Organizers
50 minutes	An Overview and Conceptualisation of Hybrid Intelligence (HI)	Invited Speakers
60 minutes	Research Showcases	Accepted Authors
40 minutes	Roundtable Discussion on the Conceptualisation of HI	Participants
60 minutes	Collaborative Design for HI Systems	Participants

20 minutes	Discussion on future research directions and collaborations	Participants

Welcome & Introduction: Participants will be welcomed, and the workshop will be introduced, along with any housekeeping announcements.

An Overview and Conceptualisation of Hybrid Intelligence In the initial segment of the workshop, we will provide an overview of Hybrid Intelligence and its relationship to learning analytics. The results from the pre-workshop survey will be analyzed to tailor the introductory content on Hybrid Intelligence accordingly. This will include conceptualization work, technology demonstrations, system architectures, and off-the-shelf applications, with a focus on exploring both the opportunities and challenges of implementing Hybrid Intelligence in learning analytics research and practices.

Research Showcases: In the second part of the workshop, authors of accepted papers will give brief flash presentations to summarize their work. Each author will prepare 6 slides for a concise 5-minute presentation. Following these presentations, a set of invited discussants with expertise in both Learning Analytics and AI will offer feedback on each presentation and initiate discussions with the audience, setting the stage for the next portion of the workshop.

Roundtable Discussion on Conceptualisation of Hybrid Intelligence: The goal of the roundtable is to collaboratively shape a shared and comprehensive understanding of Hybrid Intelligence, identify critical research questions, and explore conceptual frameworks that can guide future research and practice. This activity is designed as an interactive and collaborative session, bringing together experts from learning sciences, artificial intelligence, educational technology, and related fields. Participants will be seated in small groups at round tables, fostering an intimate and dynamic environment for in-depth discussions. By the end of the session, each table will share their insights with the larger group, contributing to a collective vision for advancing research in Hybrid Intelligence and its connection to learning analytics.

Collaborative Design for Hybrid Intelligence Systems: This is a hands-on, interactive activity aimed at bringing together participants from diverse backgrounds to co-create innovative solutions. During this activity, participants- ranging from learning scientists and AI specialists to software engineers and educators - will work in small, cross-disciplinary teams to conceptualize and design Hybrid Intelligence systems that effectively integrate learning analytics for provided scenarios, one in an educational setting and one in a workplace setting. Based on the pre-workshop survey results, we will ensure that each team has a diverse mix of expertise and experience with Generative AI, while also aligning participants with shared common interests. The focus will be on leveraging the unique strengths of both humans and AI to enhance educational outcomes. Teams will engage in brainstorming, prototyping, and iterative design processes, with the goal of developing practical, user-centered systems that address real-world challenges in education. By the end of the sessions, each team will present a preliminary design concept that can be further refined and potentially implemented, contributing to the advancement of Hybrid Intelligence in learning analytics.

Discussion on future research directions: All participants will be invited to contribute with ideas to set a potential agenda for hybrid intelligence in learning analytics research.

3 DISSEMINATION STRATEGY

Once the workshop is approved, a dedicated website will be created to serve as the central hub for announcing the call for participation. The website will provide key details, including the workshop's objectives, information about the organizers, contact details, and updates on reports and outputs generated from the event. The outreach efforts will also extend to the following strategies to attract people to our tutorial:

- Social Media platform: Use platforms like LinkedIn and Twitter (X) to share engaging previews, such as teasers and key insights attendees will gain.
- Targeted Emails: Send concise, compelling emails to relevant mailing lists, emphasizing the unique value and content of the tutorial.
- Practical Benefits: Highlight how this workshop will provide actionable insights into AI's application in fields so that audiences will gain practical benefits from the tutorial and can directly apply them to their own research.
- Networking Opportunities: Emphasize the chance to connect with peers and experts across AI, psychology, and cognitive sciences.
- Leverage Conference Platforms: We will promote our tutorial together with the conference via the website and other platforms.

Accepted submissions will be made available either as workshop proceedings, within the LAK companion proceedings, or as part of a CEUR proceedings set.

4 REFERENCES

Cukurova, M. (2024). The Interplay of Learning, Analytics, and Artificial Intelligence in Education. *British Journal of Educational Technology, 1–20.* https://doi.org/10.1111/bjet.13514.

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. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, *54*(5), 1057-1076.

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 1-18.

Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*, 50(*6*), 3004-3031.

Viberg, O., Kizilcec, R. F., Wise, A. F., Jivet, I., & Nixon, N. (2024). Advancing equity and inclusion in educational practices with AI-powered educational decision support systems (AI-EDSS). *British Journal of Educational Technology*.

Second International Workshop on Generative AI and Learning Analytics (GenAI-LA): Evidence of Impact on Human Learning

Lixiang Yan, Andy Nguyen, Ryan Baker, Mutlu Cukurova, Dragan Gasevic, Kaixun Yang, Yueqiao Jin, Linxuan Zhao, and Yuheng Li

ABSTRACT: Building on the resounding success of the First International Workshop on GenAI-LA at LAK24, which ignited conversations and collaborations around practical tools and methodologies, the Second International Workshop on GenAI-LA aims to make even greater strides. The inaugural workshop attracted over 60 participants, published nine workshop papers, and received an impressive overall rating of 4.8/5. Over the past year, significant progress has been made, and this second workshop will bring together learning scientists, learning analytics practitioners, software engineers, and AI specialists. The focus will be on delving deeper into the actual impacts of GenAI technologies on human learning. We will explore the pivotal role of learning analytics in understanding and nurturing essential cognitive, metacognitive, and creative skills. In an era where human-GenAI collaboration is increasingly valued in education and the workplace, this workshop aims to envision and inspire future research in learning analytics and GenAI.

Keywords: Generative AI, Learning Analytics, Educational Technologies

1 INTRODUCTION

The advancement of generative AI (GenAI) technologies represents a critical juncture in the progression of learning analytics and educational technology, presenting a broad spectrum of opportunities alongside a unique set of challenges (Khosravi et al., 2023; Yan et al., 2024). These technologies hold the potential to transform personalised learning, educational content creation, and assessment methods, thereby significantly enhancing learning experiences (Kasneci et al., 2023; Yan et al., 2024). However, the swift advancement and deployment of GenAI in educational settings necessitate a comprehensive examination of its effects on learning processes and outcomes (Cukurova, 2024). A notable deficiency in the current literature is the predominance of opinion pieces and speculative analyses concerning GenAI's capabilities and future applications in education. Much of the existing discourse centres on its potential rather than offering a grounded understanding of GenAI's actual impacts on learning (Hennessy et al., 2024). This highlights the urgent need for empirical research that transcends speculative discussions, aiming to elucidate the tangible impacts and implications of GenAI integration within educational contexts. Decades of research in learning analytics (LA) provide promising methodologies to capture and measure the effects of GenAI on human learning, delivering detailed insights to guide future practice and policy concerning the adoption of GenAI technologies across various educational settings (Yan et al., 2024).

The aim of the workshop is to spark discussions and foster collaboration around the potential applications of LA in capturing GenAI's impacts on human learning. By bringing together a subcommunity of LA researchers and practitioners with expertise in learning sciences, software engineering, and artificial intelligence, we plan to tackle several pivotal questions: What are the critical elements in GenAI analytics that significantly impact user interactions in diverse learning contexts? Which key metrics should be prioritised to evaluate the effectiveness and user engagement in GenAI-supported learning environments? How can we accurately measure and analyse user perceptions and experiences with GenAI tools across various educational settings? The anticipated outcomes of this workshop include 1) Establishing a consolidated network of LA researchers and practitioners with interest in GenAI-LA; 2) Producing a workshop proceeding that features pioneering works utilising LA to understand the impacts of GenAI on human learning; and 3) Developing a set of best practices for capturing GenAI's effects on human learning through LA.

2 BACKGROUNDS

As technological advancements have progressed, human learning has continuously adapted, with each new innovation significantly altering educational methods. The printing press made knowledge accessible to the masses, the Internet revolutionised how information is shared and how people learn together, and now GenAI technologies, such as large language models (LLMs) and diffusion models, are offering new opportunities to rethink education (Gašević et al., 2023). GenAI holds significant promise in automating learning tasks (Yan et al., 2024), delivering timely feedback (Dai et al., 2023), and creating dynamic learning resources (Mazzoli et al., 2023). These capabilities suggest transformative potential in personalised learning, educational content creation, and assessment methods, thereby enhancing learning experiences (Kasneci et al., 2023). However, this potential is accompanied by challenges, such as exacerbating the digital divide (Pontual et al., 2020), potentially diminishing learner agency, and introducing ethical concerns (Yan et al., 2024). Despite the emergence of numerous positioning works, there remains little empirical evidence of GenAI's impact on human learning, necessitating a rigorous and evidence-driven approach (Hennessy et al., 2024).

Learning analytics holds the promise to address this gap in high-quality research, particularly concerning the impacts of GenAI on the learning process (Khosravi et al., 2023; Yan et al., 2024). For example, by leveraging data from various educational technologies, learning analytics can provide detailed insights into how learners interact with GenAI tools. This includes tracking engagement metrics, identifying patterns in learner behaviour, and analysing the effectiveness of personalised learning interventions (Liu et al., 2024). Furthermore, learning analytics can help in understanding the cognitive and metacognitive processes that GenAI tools influence, such as problem-solving strategies, critical thinking, and creativity (Nguyen et al., 2024). Additionally, learning analytics can facilitate the measurement of user perceptions and experiences, providing a comprehensive picture of how GenAI tools are perceived and utilised across different educational contexts (Jin et al., 2023). This empirical evidence is crucial for developing best practices and guidelines for the effective integration of GenAI in educational settings, ensuring that these technologies are used to enhance, rather than hinder, the learning experience. The Second International Workshop on GenAI-LA will examine the real effects GenAI technologies have on human learning, expanding on the initial discussions and partnerships formed during the first workshop.

2.1 Evidence of interest

The First International Workshop on GenAI-LA at LAK24 has been a resounding success, igniting conversations and collaborations around practical tools and methodologies. More than 50 participants attended, nine workshop papers were published, and the overall rating was 4.8/5. This workshop organising committee also guest edited two special sections in the British Journal of Educational Technology, with 91 abstract submissions, and the Journal of Learning Analytics, with 20 full paper submissions, focusing on GenAI, learning, and learning analytics. This evidence underscores the significant interest and engagement within the academic community regarding GenAI and its applications in learning analytics.

3 ORGANISATIONAL DETAILS

3.1 Workshop format, participation, and pre-workshop task

The workshop is designed as a half-day, face-to-face event, with an expected attendance of 30 to 50 participants. It aims to bring together a diverse group of individuals, including learning scientists, learning analytics practitioners, software engineers, and AI experts, who are all interested in the intersection of GenAI and learning analytics. The event is open to anyone with an interest in this field, regardless of their level of expertise. Participants will be encouraged to present their recent research on leveraging learning analytics to understand the impacts of GenAI on human learning. Following the approval of this workshop proposal, a call for papers will be issued to invite detailed contributions on this topic. Submissions, ranging from 4 to 8 pages, will be reviewed by the organising committee and the paper authors. Attendees will have the opportunity to access both submitted and accepted papers in advance, ensuring that discussions during the workshop are well-informed and productive.

3.2 Workshop activities

The workshop is scheduled to occur during the pre-conference activities of the main conference, formatted as a half-day session lasting up to 4 hours (on either March 3 or 4, 2025). The workshop will be divided into three parts:

1. **Opening Keynote (30 mins)**. The workshop will commence with an opening keynote address delivered by an invited speaker, Professor Ryan Baker or Professor Mutlu Cukurova (TBC). This keynote will set the stage by discussing the current state of GenAI and learning analytics, highlighting recent advancements, ongoing challenges, and future directions. The presentation aims to inspire and provide a comprehensive overview that will frame the subsequent sessions.

2. **Workshop Papers (90 mins)**. In this segment, authors of accepted workshop papers will present their findings. Each presentation will be allocated 10 minutes, split into 5 minutes for the presentation and 5 minutes for a Q&A session. This format ensures a dynamic and engaging exchange of ideas, allowing for immediate feedback and discussion. The papers will cover various topics related to the integration of GenAI and learning analytics, showcasing empirical research and case studies.

3. **Collaborative Design Sessions (90 mins)**. Participants will engage in a group-based activity designed to foster collaboration and innovation. Attendees will be divided into small groups based on their experiences and interests. Each group will be assigned a specific learning scenario involving GenAI and tasked with designing a study to use learning analytics to capture the impacts of GenAI on particular learning outcomes or processes. The activity will be structured using the LA cycle framework, guiding participants to identify the learner, data, analytics, and intervention components (60 minutes). After the design phase, each group will present their study design to the entire workshop (30 minutes). Experienced organisers will be available throughout the session to provide guidance and support, helping teams navigate any challenges they encounter.

3.3 Dissemination strategy

Upon the workshop's approval, a dedicated website will be created to serve as the main platform for announcing the call for participation. Outreach efforts will also include posts on Twitter accounts and mailing lists available to the workshop organisers. The website will provide key information such as the workshop's objectives, organiser details, contact information, and subsequent reports and outputs from the event. Accepted submissions will be published as part of a CEUR proceeding.

3.4 Logistics

The workshop will be conducted as an in-person event. The chosen venue will offer flexible seating arrangements with movable desks and chairs to facilitate the collaborative design session. For pre-workshop interactions, a Google form will be used to distribute a pre-workshop survey. This form will also include an invitation to a dedicated Slack channel to ensure smooth communication before and after the workshop.

REFERENCES

Cukurova, M. (2024). The Interplay of Learning, Analytics, and Artificial Intelligence in Education. *British Journal of Educational Technology,* 1-20, https://doi.org/10.1111/bjet.13514

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D., & Chen, G. (2023, July). Can large language models provide feedback to students? A case study on ChatGPT. In 2023 IEEE International Conference on Advanced Learning Technologies (ICALT) (pp. 323-325). IEEE.

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. Computers and Education: Artificial Intelligence, 4, 100130.

Hennessy, S., Cukurova, M., Lewin, C., Mavrikis, M., & Major, L. (2024). BJET Editorial 2024: A call for research rigour. British Journal of Educational Technology, 55(1), 5–9.

Jin, Y., Li, P., Wang, W., Zhang, S., Lin, D., & Yin, C. (2023). GAN-based pencil drawing learning system for art education on large-scale image datasets with learning analytics. Interactive Learning Environments, 31(5), 2544–2561.

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.

Khosravi, H., Viberg, O., Kovanovic, V., & Ferguson, R. (2023). Generative AI and Learning Analytics. Journal of Learning Analytics, 10(3), 1–6.

Liu, J., Li, S., & Dong, Q. (2024). Collaboration with Generative Artificial Intelligence: An Exploratory Study Based on Learning Analytics. Journal of Educational Computing Research, 07356331241242441.

Mazzoli, C. A., Semeraro, F., & Gamberini, L. (2023). Enhancing Cardiac Arrest Education: Exploring the potential use of MidJourney. Resuscitation, 189.

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. Studies in Higher Education, 1-18.

Pontual Falcão, T., Ferreira Mello, R., & Lins Rodrigues, R. (2020). Applications of learning analytics in Latin America. British Journal of Educational Technology, 51(4).

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez‐Maldonado, R., Chen, G., ... & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1), 90-112.

Yan, L., Martinez-Maldonado, R., & Gašević, D. (2024). Generative Artificial Intelligence in Learning Analytics: Contextualising Opportunities and Challenges through the Learning Analytics Cycle. In Proceedings of the 14th International Conference on Learning Analytics & Knowledge (pp. 101-111).

The 8th Workshop on Predicting Performance Based on the Analysis of Reading and Learning Behavior

Brendan Flanagan[1*], Owen H.T. Lu[2], Atsushi Shimada[3], Namrata Srivastava[4], Albert C.M. Yang[5], Hsiao-Ting Tseng[6], Fumiya Okubo[3], Eduardo Davalos Anaya[4], Hiroaki Ogata[1]

[1] Kyoto University, Japan
[2] National Chengchi University, Taiwan
[3] Kyushu University, Japan
[4] Vanderbilt University, USA
[5] National Chung-Hsing University, Taiwan
[6] National Central 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, LAK23 and LAK24 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 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.

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, a joint dataset with students' coding behaviors and predict the final grade score of learners (Authors, 2024), with 26 participants in the last instance of the workshop.

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 we welcome research questions from all participants, and we expect to emphasize the following topic which the organizers feel attention should be paid. Low retention and high failure rates are important problems in education (Villagrá-Arnedo et al., 2017). However, studies have shown that timely interventions for at-risk students can be effective in helping change their behaviors (Arnold et al., 2012; Tanes et al., 2011). Therefore, focusing on the early detection of at-risk students is an essential step to changing student's behavior for greater success. This broader task may be approached from the following perspectives:

- Student reading behavior self-regulation profiles spanning the entire course
- Preview, in-class, and review reading patterns
- Student engagement analysis; and behavior change detection
- Visualization methods to inform and provide meaningful feedback to stakeholders

In addition, this workshop will accept 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.

3 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 main theme of last year's LAK conference, Learning Analytics in the Age of AI, the topic of generative AI continued to feature in the workshop submissions including research that proposed a framework (Ma & Chen) for constructing concept maps from digital learning materials used in e-book reading systems such as BookRoll. To date, this workshop has focused on analyzing the reading behavior data collected as a product of system use and has been used to understand behavior-based learning processes. The automated extraction of concept maps from e-books using large language models could support analyzing the learning process from both knowledge and behavioral contexts (Flanagan et al., 2019), which has previously been constrained by the burden that was placed on domain experts who created concept maps. There was also a submission that continued to build on research into a learning analytics framework for the collection and analysis of affect states and feedback through an emotion-focused dashboard. Previous incarnations of this research focused on the extraction of affect states from visual contexts, such as video feeds from Zoom lectures. The current work proposes the analysis and monitoring of affect states from the oral context through audio voice analysis, and using automated speech-to-text to support sentiment analysis from learner discourse. The analysis and visualization are presented to provide feedback for teachers and students. The proceedings of the workshop can be found on the following website: https://sites.google.com/view/lak25datachallenge.

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., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge map creation for modeling learning behaviors in digital learning environments. Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge, pp. 428-436.

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., Shimada, A., Okubo, F., Tseng, H-T., Yang, A.C.M., Lu, O.H.T., Ogata, H. (2024) The 6th Workshop on Predicting Performance Based on the Analysis of Reading and Learning Behavior. Companion Proceedings of the 14th 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.

LAK25 Trace-SRL: The Workshop on Measuring and Facilitating Self-regulated Learning and Human-AI Co-regulated Learning based on Trace data

Xinyu Li, Linxuan Zhao, Mladen Raković, Dragan Gašević, Yizhou Fan
Monash University, Peking University
Xinyu.li1@monash.edu, fyz@pku.edu.cn

ABSTRACT: This is the proposal for the third interactive workshop on Measuring and Facilitating self-regulated learning (SRL). Measuring SRL using unobtrusive trace data and facilitating SRL through real-time analysis of such data have been identified as highly valuable research directions. However, significant challenges remain in this area, including: (i) the detection, measurement, and validation of SRL processes using trace data is still a debated issue; (ii) the design principles for effective interventions and the complex conditions under which these interventions facilitate learning are not yet well understood; and (iii) the potential benefits of advanced AI techniques, such as ChatGPT, for learners, as well as the mechanisms through which learners can effectively co-regulate with AI, remain unclear. Therefore, we aim to enhance SRL measurement and facilitation through a full-day workshop, providing participants hands-on experience with our AI-powered Trace-SRL tools. We aim to share our platform, tasks, data and project experiences, then discuss an annual international joint study to initiate international collaboration and deepen SRL research. An open call for contributions will be distributed, and the participants will join roundtable-style discussions and hands-on co-design activities. Expected outcomes are forming a community of practice, potential collaborative projects, and possible follow-up joint publications.

Keywords: Self-regulated learning, Trace data, Measurement protocols, Scaffoldings and Dashboards, Human-AI Co-regulated Learning

1. BACKGROUND

1.1 Challenges

Self-regulation improves learning outcomes as revealed by the positive relation between SRL processes and learning measures (Harley et al., 2017). **Measuring SRL**, however, has posed a major challenge to researchers for decades. Various measurement tools and methods have been proposed to help improve the capture of SRL processes, ranging from self-report surveys (Pintrich & et al. 1991), think-aloud protocols (Bannert, 2007), and **trace-based measurement** (Fan et al., 2022). Trace-based methods are gaining popularity because they unobtrusively capture cognitive and metacognitive activities in authentic learning environments (Winne, 2010) and have been employed in multiple studies (Saint et al., 2022, Fan et al., 2022). However, the detection, measurement, and validation of SRL processes with trace data is debatable (Winne, 2020). Hence, we propose this **interactive workshop (aim 1)** for interested researchers to examine the current SRL-related work-to-date, explore how they can build upon existing methods of measurement of SRL, and exchange their lessons learnt from different projects.

While the importance of SRL to learning is widely recognized, effective ways to support learners in regulating their learning remain unclear (Guven & Babayıgıt, 2023). Different types of interventions, such as **scaffolding, dashboards or personalised feedback**, have been designed in learning analytics to effectively support learners' self-regulated learning and ultimately improve their SRL skills. However, there is limited research into the development of these interventions and how design decisions are associated with the execution of SRL and learning outcomes (Lyn et al., 2023). Importantly, the complex conditions and contexts when these interventions facilitate and enhance learning are not known (Guo, 2022). Therefore, this **interactive workshop (aim 2)** aims to address these challenges by sharing how different interventions can be designed, the potential of the interventions, and/or how effective interventions are in supporting SRL. This will lead to new insights concerning the effectiveness of intervention approaches to facilitate self-regulation.

The advancement of AI technologies is revolutionising contemporary education (Chan et al., 2023). Various AI technologies have been integrated into different educational systems to support student learning (Molenaar, 2022), which is inevitable for learners to possess co-regulated learning skills with AI. However, the interaction processes between learners and AI remain insufficiently underexplored. This **interactive workshop (aim 3)** aims to investigate the design of new AI-powered instrumentation tools to detect and measure learning processes during AI interaction, providing insights into the effectiveness of AI in enhancing self-regulation and human-AI co-regulation.

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 SRL measurement and analysis; iii) understand how student data and AI can generate actionable learning insights; iv) design new forms of SRL scaffolding, dashboards or feedback to facilitate teaching and learning. From the participant's perspective, we expect to: i) improve the knowledge and skills in 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 collaborative projects; iv) explore opportunities for joint publications (e.g., a journal special issue) and future workshops. In last year's workshop, we attracted **28 scholars from more than 20 institutions in more than a dozen countries** to participate in our workshop. They provided highly positive feedback for the workshop and expressed their willingness to continue the dialogue, such as the research teams from The University of Hong Kong (HKU) and National Taichung University of Education (NTCU). In their feedback and suggestions, many scholars mentioned the **openness of learning platforms and tools, data sharing, and the importance of international collaborative research**. Therefore, in this year's proposal, we emphasise two objectives different from other workshops or research tracks in LAK25:

- **Provide more hands-on opportunity to experience the measuring and facilitating of SRL using our platform**. Participants will explore a learning analytic project and platform (developed and led by organisers, project name hidden for review) integrated with various instrumentation tools and personalised rule-based/GPT-based 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 bringing together like-minded researchers and teachers, we aim to share our

platform, tasks, data and project experiences, then discuss an annual international joint study plan. For example, asking different teachers to use the same platform and assign similar tasks in their courses. In this way, the field can collect data that can be compared, triangulated and investigated in multiple contexts, which will greatly facilitate international collaborative research and dialogue, and further deepen our understanding of self-regulated learning.

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	
10 minutes	Welcome & Introduction (Morning Section: Measuring SRL)	Organiser 1
40 minutes	2 Presentations about measuring SRL using trace data ● Presentation 1 (12-15 minutes talk + 3-5 minutes Q&A) ● Presentation 2 (12-15 minutes talk + 3-5 minutes Q&A)	Participants
20 minutes	Roundtable Discussion (Previous presenters + Audience) ● Guided by structured questions	Organiser 2 (Host)
30 minutes	Coffee Break and Socialization	All
40 minutes	Presentation of Analytics Platform and Hands-on Task	Organiser 3
40 minutes	Brainstorming about new direction of measuring SRL using trace data Discussing SRL measuring approaches which will be used in the join study	Organiser 2
10 minutes	Summarising the morning section & Next Steps	Organiser 1
	Part 2: Afternoon Section	
10 minutes	Welcome & Introduction (Afternoon Section: Facilitating SRL)	Organiser 1
40 minutes	2 Presentations about measuring SRL using trace data ● Presentation 3 (12-15 minutes talk + 3-5 minutes Q&A) ● Presentation 4 (12-15 minutes talk + 3-5 minutes Q&A)	Participants
20 minutes	Roundtable Discussion (Previous presenters + Audience) ● Guided by structured questions	Organiser 2 (Host)
30 minutes	Coffee Break and Socialization	All
40 minutes	Presentation of Scaffolding system and Hands-on Task	Organiser 3
40 minutes	Brainstorming about new direction of facilitating SRL using trace data Discussing SRL facilitating approaches which will be used in the join study	Organiser 4
10 minutes	Summarising the afternoon section & Next Steps	Organiser 4

The event will be an open and hands-on workshop. The organisation of the workshop will revolve around 3-4 cutting-edge research projects related to trace-based SRL study, 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. The submission timeline will follow the timeline suggested by the conference organisers. All attendees will have the opportunity to discuss with the presenters in the roundtable and brainstorming sessions, and will also have hands-on experiences with SRL measurement and scaffolding design activities guided by **organisers**. We anticipate a registration of about 20-30 participants. We will use #LAKTRACESRL when referencing this event on social media.

After the workshop, **we will organise 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 the call-for-papers and send relevant news to potential participants (including participants from our previous events, e.g., we hosted several workshops in relevant conferences). At the same time, we will send invitations to specific research teams who are working on measuring and facilitating SRL (already have 3-4 teams in mind, and three of them expressed an initial intention to participate). 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

Bannert, M. (2007). Metakognition beim lernen mit hypermedien. Waxmann Verlag.

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.

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.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior*, 112, 106457.

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.

Chan, C. K. Y., & Tsi, L. H. (2023). The AI revolution in education: Will AI replace or assist teachers in higher education?. *arXiv* preprint arXiv:2305.01185.

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., ... & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, 139, 107547.

Guven, M., & Babayıgıt, B. B. (2020). Self-regulated learning skills of undergraduate students and the role of higher education in promoting self-regulation. *Eurasian Journal of Educational Research*, 20(89), 47-70.

Exploring Best Practices for Integrating GenAI into Learning Analytics Dashboards

Alan Berg
University of Amsterdam
a.m.berg@uva.nl

Manuel Valle Torre
Delft University of Technology
m.valletorre@tudelft.nl

Priyanka Pereira
University of Twente
p.d.pereira@utwente.nl

ABSTRACT: Interactive and content-generating dashboards are incrementally becoming the norm in education. However, many dashboard pilots do not survive the vigors of full deployment. This half-day hands-on workshop discusses best practices for strengthening the Learning Analytics (LA) feedback cycle within interactive Large Language Model (LLM) dashboards, such that the LA provides a safety net for the intervention triggered by the generative AI (genAI). We frame the conversation in terms of a Dutch experimental infrastructure. We review the already-established work of the participants in groups, through a mockup session paper or based on live infrastructure. The audience provides feedback on currently-available dashboards, through an interactive review process. We define the requirements, the design practices and the conceptual processes that aid in strengthening the relationship between genAI as an intervention and the entire LA feedback cycle. Documentation of essential requirements, design practices, and conceptual processes will later be prepared for further dissemination and improvement. The workshop thus stimulates and documents the evolution of best practices around the dance between genAI and LA within an educational dashboard setting. The workshop is based on the knowledge gathered from Npuls, the Dutch national digital transformation effort aimed at all adult education levels, specifically the evolving practices collected by the team "Learning Analytics: Best and Worst Practices".

Keywords: Generative AI, Learning Analytics, Best practices, Expert feedback, AIED

1 CONTEXT

Npuls is a National Growth Fund Program of and for all public secondary vocational schools, colleges and universities in the Netherlands. Npuls aims to work together to improve the quality of education, increase the agility of education, and improve the digital skills of teachers and learners (Learning Analytics Magazine: Best and Worst Practices. 2024). Within the program, resides a team focusing on the best and worst practices for deploying Learning Analytics (LA). The team has generated content in the form of magazines and has organized national events (Why Npuls. 2023). From the team's experience building, we realize that there is a need for strong coupling between the LA lifecycle and the interventions created by the interactions of students and teachers with generative AI (genAI), especially within the context of dashboards (Yan et al. 2024). The implications of genAI affect the complete LA lifecycle, for example, in interactions between students and tutoring systems, in data-cleaning procedures, and in creating interactive analytics (Yan, Martinez-Maldonado, and Gasevic. 2024). This half-day hands-on workshop uses participant feedback to explore how the themes can be related practically.

Are you part of a project to deploy educational dashboards with LA for genAI? Are you interested in defining design constraints? Are you considering the policy implications? Is the AI Act on your radar? Are you improving your understanding of the field? This workshop is a place to interact with and learn from your peers as you journey toward incorporating LA in genAI dashboards.

2 DESIGN

The hands-on workshop will run for a half day, with a maximum of 40 participants.

Table 1 outlines the agenda for the workshop. We will begin the workshop by briefly presenting current best practices and theory. We will then invite the participants to provide a three-slide presentation of their or their organizations' genAI dashboards that they are conceptualizing, designing, developing, piloting, or deploying. After a short coffee break, the participants will be divided into groups to examine each presented dashboard and provide feedback. If participants have a specific challenge or focus point in their genAI-enabled LA system, they are invited to mention it during their presentation, so others can join that group discussion. We contextualise the different aspects of genAI and LA proposed by Yan, Martinez-Maldonado, and Gasevic 2024, to help participants structure the discussion or collaboration during this step. Finally, we will discuss and condense the day's experience into key points, which we will later present to an academic audience via a paper submitted to a conference or a journal, as well as via practitioner reports for a broader audience through Npuls.

Table 1: Agenda.

Event	Activity
Introduction (theory and best practices)	Organizer presentation
Participants' dashboards	Participant presentations
Coffee break	-
Review of dashboards	Group discussion
Paper or live design	Mock-up ideas on paper or in the cloud
Summary of lessons learned	Agreement with the group on key points

The supporting infrastructure for the workshop includes a repository to record experimentation and a shared online folder where we can collaboratively edit and review the generated presentations and documents. Additionally, participants receive templates, documented activities supporting the agenda, and an interface for ad-hoc communication. Where possible, we use Npuls AI-related infrastructure.

The workshop outcomes include consensus building and documentation around best practices associated with coupling genAI with LA dashboards; gathering of essential requirements, design practices, and conceptual processes in a collaborative document or live design; feedback on specific dashboards as presented by the workshop participants; awareness building within communities (such as Npuls, SOLAR, participant and organizer networks); and networking with the participants. Concrete outputs include blog posts and/or magazine articles in the Npuls network and a potential paper submission.

We deliver external publicity for the workshop within the Npuls community via the magazine and the community page of the Best and Worst Practices team. The workshop is also publicized through the Special Interest Group for LA at SURF, a post on the website of the European project AI4VET4AI, and the agenda on the website of the LAK25 conference. The workshop organizers also provide further dissemination via their networks.

REFERENCES

Learning Analytics Magazine: Best and Worst Practices. May, 2024. Npuls. Retrieved September 4, 2024, from https://npuls.nl/en/knowledge-base/learning-analytics-magazine-best-and-worst-practices/

Why Npuls. 2023. Npuls. Retrieved September 4, 2024, from https://npuls.nl/en/why-npuls

Yan, L., Zhao, L., Echeverria, V., Jin, Y., Alfredo, R., Li, X., Gasevic, D., & Martinez-Maldonado, R. (2024). VizChat: Enhancing Learning Analytics Dashboards with Contextualised Explanations Using Multimodal Generative AI Chatbots. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), Artificial Intelligence in Education (pp. 180–193). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64299-9_13

Yan, L., Martinez-Maldonado, R., & Gasevic, D. (2024). Generative Artificial Intelligence in Learning Analytics: Contextualising Opportunities and Challenges through the Learning Analytics Cycle. Proceedings of the 14th Learning Analytics and Knowledge Conference, 101–111. https://doi.org/10.1145/3636555.3636856

Transmodal Analysis: A New Conceptual and Methodological Framework for Multimodal Learning Analytics

Author(s): Yeyu Wang, David Williamson Shaffer, Andrew Ruis, Brendan Eagan
University of Wisconsin–Madison
ywang2466@wisc.edu, dws@education.wisc.edu, arruis@wisc.edu, beagan@wisc.edu

Sanna Järvelä, Andy Nguyen
University of Oulu
Sanna.Jarvela@oulu.fi,
andy.nguyen@oulu.fi

Crina Damsa, Rogers Kaliisa
University of Oslo
crina.damsa@iped.uio.no,
rogers.kaliisa@iped.uio.no

Liv Nøhr, Daniel Spikol
University of Copenhagen
lsnl@ind.ku.dk, ds@di.ku.dk

Zack Carpenter, David DeLiema
University of Minnesota
carpe787@umn.edu, ddeliema@umn.edu

Zachari Swiecki
Monash University
zach.swiecki@monash.edu

Mamta Shah
Elsevier
m.shah@elsevier.com

ABSTRACT: Multimodal data integration is one of the major analytical challenges for Multimodal Learning Analytics (MMLA). Continuing conversations begun in previous LAK CROSSMMLA workshops, this workshop will focus on a relatively new conceptual and methodological framework, Transmodal Analysis (TMA), that addresses the data integration challenge using a *functions-not-fusion* approach. In this workshop, TMA adopters, MMLA methodologists, and learning scientists will discuss the affordances and challenges of this new approach. Participants will gain hands-on experience and learn how to use TMA on their own multimodal data. This workshop, thus, provides a venue for the MMLA SIG and others interested in MMLA to exchange expertise and develop future collaborations.

Keywords: Multimodal Learning Analytics, Data Integration, Analytical Framework, Transmodal Analysis (TMA).

1 BACKGROUND AND PURPOSE

Learning analytics aims for a holistic and comprehensive understanding of learning processes (Oviatt et al., 2018). Multimodal learning analytics (MMLA) serves this goal by collecting, integrating, and analyzing learning traces from different tools, environments, and other sources (Ochoa, 2022). CROSSMMLA workshops have been held over the last 7 years at LAK and other conferences (e.g., ISLS, EC-TEL, LASI) to investigate and discuss the affordances and constraints of different approaches to MMLA, due to a growing interest in applications and advancing methodologies of multimodal analytics.

While models of learning are theoretically better when they account for more information about learning processes, integrating multiple data streams with different properties and time scales remains a significant analytical challenge. This workshop will explore the analytical challenge by focusing on Transmodal Analysis (TMA). This emerging conceptual and methodological framework allows researchers to specify *functions* to account for a) varied temporal structures for transmodal interactions, b) different characteristics of learner groups and subgroups, and c) the complex structure of a learning environment. That is, TMA is a *functions-not-fusion* approach, providing a more flexible framework for integrating different data streams while preserving the content and structure of the original data. Despite being a relatively new method, TMA has been well received in the Collaboration & Multimodality Workshop at ISLS Meeting 2023 and has been used by educational researchers in various contexts, such as AI-supported classrooms (Borchers et al., 2024), nursing education (Wang et al., 2023), game-based learning (Carpenter et al., 2023), socially regulated learning, collaborative problem solving, and others. New discoveries based on TMA models have provided insights for empirical educational practices and contributed to the development of educational theories.

With the adoption of any new method, TMA users have encountered challenges with the parameterization and interpretation of models, and they have also identified areas for potential expansion of the method. Thus, this workshop will focus on one particular approach and seed further methodological development in the MMLA community. Participants in this workshop will (1) be introduced to TMA as a conceptual and methodological framework for multimodal analytics with examples from empirical studies; (2) construct a TMA model using their own multimodal data; and (3) participate in a structured discussion about the affordances of TMA in its existing form and opportunities for methodological development to address the needs of MMLA researchers.

As a result of the workshop, participants will (1) be able to apply TMA to their multimodal data, (2) gain a deeper understanding of the affordances and limitations of a functions-not-fusion approach to modeling multimodal data, and (3) participate in and contribute to ongoing efforts in the MMLA community to improve methods for multimodal data integration and modeling.

2 WORKSHOP DETAILS

2.1 Event Type and Structure

During this full-day workshop, up to 30 participants will (a) learn about the conceptual and methodological underpinnings of TMA, (b) hear from 7 MMLA scholars who have used TMA about their findings and experiences, (c) develop TMA models using their own multimodal data, and (d) participate in a facilitated discussion of the affordances of TMA and opportunities for further methodological development. Regardless of analytical skill level and epistemological stance toward multimodal learning analytics, participants will be able to engage with all aspects of the workshop.

2.2 Detailed Schedule

The morning workshop will introduce TMA's conceptual and methodological entailments. This will be followed by a set of short presentations from MMLA scholars working across different learning contexts and domains. The presentations will focus on how and why TMA, as a versatile tool, has been used for different empirical studies and reflect on the affordances and opportunities for methodological advancement. The workshop leaders will then facilitate a discussion between the panelists and participants about the advantages and challenges of using TMA.

The afternoon workshop will focus on helping participants to develop TMA models of their own multimodal data. Workshop leaders will provide a detailed demonstration of how to implement TMA using a simple web-based interface. Then, participants will work in small groups with support from the workshop leaders to construct their TMA models. Each participant will work with their own multimodal data (or the dataset used to demo the approach). One participant from each group will share their model parameters, interpretations, questions and challenges for their first use of TMA. To wrap up the workshop, participants will (1) share the insights and challenges based on their TMA models, (2) discuss affordances and pitfall of TMA compared to other MMLA analytical tools and (3) explore collaboration opportunities and potential contributions in MMLA and relative areas (i.e. multimodal machine learning, Quantitative Ethnography, etc) in a bigger group.

The detailed schedule is shown in the following table:

Time	**Activity**	**Responsible**
9:00 am - 9:15 am	Introduction and purpose for the day	Daniel Spikol
9:15 am – 9:50 am	Overview for Trans-Modal Analysis and Q&A	**Presenter:** David Shaffer
10:00 am –noon	Mini-Conference: Empirical studies and discussion • *Where to Gaze? A Transmodal Analysis Investigation into Socially Shared Regulation of Learning* by Andy Nguyen and Sanna Järvelä. • *Using Multi-Modal Network Models to Visualize and Understand How Players Learn a Mechanic in a Problem-Solving Game* by David DeLiema and Zack Carpenter. • *Developing Nursing Students' Practice Readiness with Shadow Health® Digital Clinical ExperiencesTM: A Transmodal Analysis* by Mamta Shah.	**Panelists:** Sanna Järvelä Andy Nguyen David DeLiema Zack Carpenter Mamta Shah
noon - 1:30 pm	Lunch Break	
1:30 pm – 2:00 pm	TMA webtool demo	**Presenters**: Yeyu Wang and Liv Nøhr
2:00 pm – 3:00 pm	TMA model (Hands-on time)	**Mentors**: Zach Swiecki, Yeyu Wang, David Williamson Shaffer, Rogers Kaliisa, Liv Nøhr
3:15 pm - 4:15 pm	Model sharing	**Facilitators**: Rogers Kaliisa, Andrew Ruis and Brendan Eagan

4:15 pm – 5:00 pm	Debrief and closing	Crina Damsa

2.3 Recruitment and Dissemination

This workshop will be promoted through learning analytics mailing lists and those of adjacent fields (e.g., learning sciences, quantitative ethnography, AIED), social media, and the TMAlak25.org workshop website. To maximize the benefits of the full-day workshop, we will ask participants to complete a survey about their MMLA research interests, data, current approaches to modeling, and analytical skills. Before the conference, workshop leaders will review the survey responses and, if needed, assist participants with data preparation or other tasks that will help them fully participate in the workshop.

3 INTENDED OUTCOMES

As a result of this workshop, participants will (1) have a working knowledge of the conceptual foundations of TMA as a methodology and understand how to apply it to the kinds of multimodal data they work with and (2) contribute to the ongoing improvement of MMLA techniques and methods by engaging in informed discussions of the affordances and challenges of using function-based approaches to integrating multimodal data. We plan to consolidate participants' empirical work and conceptual insights from this workshop into a collaborative paper or symposium, offering new perspectives on modeling multimodal interactions for the broader learning analytics community.

REFERENCES

Borchers, C., Wang, Y., Karumbaiah, S., Ashiq, M., Shaffer, D. W., & Aleven, V. (2024). Revealing Networks: Understanding Effective Teacher Practices in AI-Supported Classrooms using Transmodal Ordered Network Analysis. Proceedings of the 14th Learning Analytics and Knowledge Conference, 371–381. https://doi.org/10.1145/3636555.3636892

Carpenter, Z., Wang, Y., DeLiema, D., Kendeou, P., & Shaffer, D. W. (2023). Using Multi-Modal Network Models to Visualize and Understand How Players Learn a Mechanic in a Problem-Solving Game. The 13th International Learning Analytics and Knowledge Conference, 99–101. https://experts.umn.edu/files/584510749/LAK23_CompanionProceedings.pdf

Ochoa, X. (2022). Chapter 6: Multimodal Learning Analytics—Rationale, Process, Examples, and Direction.

Oviatt, S., Grafsgaard, J., Chen, L., & Ochoa, X. (2018). Multimodal learning analytics: Assessing learners' mental state during the process of learning. The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations—Volume 2 (pp. 331–374). Association for Computing Machinery. https://doi.org/10.1145/3107990.3108003

Wang, Y., Shah, M., Jimenez, F. A., Wilson, C., Ashiq, M., Eagan, B., & Williamson Shaffer, D. (2023). Developing Nursing Students' Practice Readiness with Shadow Health® Digital Clinical Experiences: A Transmodal Analysis. In G. Arastoopour Irgens & S. Knight (Eds.), Advances in Quantitative Ethnography (Vol. 1895, pp. 365–380). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47014-1_25

CROSSMMLA: Multimodal Learning Analytics in the Age of GenAI

Namrata Srivastava[1,2], Roberto Martinez-Maldonado[2], Daniele Di Mitri[3], Daniel Spikol[4], Vanessa Echeverria[2,5], Kateryna Zabolotna[6], Simon Knight[7], Mohammad Khalil[8], Gloria Fernandez Nieto[2], Ruth Cobos[9]

Vanderbilt University[1], Monash University[2], DIPF | Leibniz Institute for Research and Information in Education[3], University of Copenhagen[4], Escuela Superior Politécnica del Litoral[5], University of Oulu[6], University of Technology Sydney[7], University of Bergen[8], Universidad Autónoma de Madrid[9]

ABSTRACT: The CROSSMMLA workshop series has focused on collecting and analyzing educational datasets from multiple modalities of interaction across physical and digital learning spaces. In this year's workshop, we aim to explore how the rise of Generative AI (GenAI) models is transforming the landscape of multimodal learning analytics (MMLA) research, driving new possibilities for understanding and enhancing the learning process. In recent years, GenAI models have made impressive strides, particularly in their ability to process various types of data beyond text, including images, audio and videos. This development has led to the use of Multimodal Large Language and Vision Models in MMLA research. However, a recent analysis highlighted several challenges that need to be addressed when applying GenAI in education, including concerns about data privacy, algorithmic bias, ethical use of AI-generated content, and the scalability of such models in diverse educational settings. Additionally, issues surrounding the transparency of AI decision-making, the environmental costs of training large models, and the societal implications of over-reliance on AI must be considered. To address these challenges, we propose a half-day workshop to explore the opportunities and complexities of Generative AI for advancing MMLA research.

Keywords:

1 WORKSHOP BACKGROUND AND MOTIVATION

As the field of Multimodal Learning Analytics (MMLA) continues to evolve, it is essential to critically examine its methodological underpinnings, particularly as new technologies such as Generative AI (GenAI) transform how we collect, analyze, and interpret educational data. Integrating GenAI models into MMLA presents both exciting opportunities and complex challenges. For instance, GenAI offers significant potential to process multimodal data, ranging from text to images and audio, enabling richer insights into how learning interactions unfold across diverse educational settings (Giannakos et al., 2024; Schneider et al., 2024; Yan et al., 2024). For instance, the use of GenAI could automate content generation, allowing educators to develop personalized, multimodal learning materials quickly and enhance student engagement by providing real-time feedback based on their behavioral and performance data (Giannakos et al., 2024; Yan et al., 2024). However, as we embrace these advancements, engaging in a deeper reflection is crucial to ensure that we address the ethical, technical, and societal issues arising from the use of Generative AI models.

This workshop aims to foster such reflection by exploring how the rise of GenAI could reshape the landscape of MMLA research. GenAI models, particularly Multimodal Large Language and Vision Models, offer novel ways to process complex data, but their use also raises significant concerns about privacy, algorithmic bias, transparency, and the environmental costs of large-scale AI systems (Alwahaby et al., 2022; Giannakos et al., 2024; Yan et al., 2024). As Alwahaby (2022) mentioned in their review, it is important to note that MMLA is different from LA as it involves capturing high-frequency data for all human activity, enabling deeper insights into the learning process (Blikstein, 2013), potentially capturing data in a highly inclusive manner (i.e., face-recognition). With the rise of GenAI techniques in MMLA, multiple dimensions of ethical considerations open up. For instance, data privacy and security become significant concerns when sensitive data, such as video, voice, and biometric information, are involved. Moreover, these models' cost, scalability and adaptability across different educational contexts remain open questions that demand further inquiry (Schneider et al., 2024). Additionally, bias in AI models is a key challenge, particularly when decisions are made using data from diverse populations, as GenAI could unintentionally perpetuate inequities in learning environments if not managed carefully (Alwahaby et al., 2022; Giannakos et al., 2024).

This workshop explores these challenges by bringing together researchers and practitioners to discuss how GenAI can be effectively and responsibly integrated into MMLA. The discussions will focus on practical methods, such as designing ethical, transparent, and scalable GenAI-powered MMLA tools while addressing concerns about privacy and bias (Prinsloo, Slade, & Khalil, 2023). For instance, integrating GenAI with human expertise, ensuring open learner models for transparency, and employing hybrid AI-human systems could mitigate many of these issues.

The CROSSMMLA workshop series has been pivotal in advancing MMLA research, bringing scholars and practitioners together to explore novel methodologies, tools, and applications. This year's workshop will build on this legacy by evaluating how GenAI is shaping the future of MMLA, offering practical insights into how AI-driven tools can enhance learning analytics while maintaining ethical standards and promoting equitable access to AI technologies in education.

2 WORKSHOP DETAILS

2.1 Event Type and Structure

We propose a half-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, but participants should be interested in methodology and research design in 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 GenAI research in MMLA.

2.2 Schedule and Activities

2.2.1 Elevator Pitches and Framing Discussions: 30 minutes

To begin, we will set the tone for active participation, emphasizing that this is not a "sit and listen" event. Participants will take a round of elevator introductions, providing a more engaging, fun (hopefully), and prime the discussions throughout the day. We will ask the participants to introduce themselves and give a brief pitch about their research interests and the type of multimodal data they are working with. As facilitators, we will present key definitions for ideas and curated research examples throughout these activities to frame the discussions, provide context, and promote active participation.

2.2.2 Diving into GenAI Research in MMLA: 1.5 hours

After framing the discussion around GenAI research, the organizers will provide a few examples of how GenAI tools are currently applied in MMLA research. Next, the participants will engage in hands-on activities where they will have the opportunity to apply the tools on either their own dataset or an example dataset provided by the organizers and share their findings and insights.

2.2.3 Future considerations for GenAI Research in MMLA: 1 hour

After the hands-on activity, participants will break into small groups to share their perceptions, motivations, and insights on the challenges and opportunities they foresee in applying the current capabilities of GenAI to MMLA research. As facilitators, we will guide the discussions, focusing on key themes such as data privacy, algorithmic bias, environmental impact, transparency, and the risks of over-reliance on AI systems.

2.2.4 Reflections and Next Steps: 30 mins

To conclude the workshop, we want to summarize the outputs and developments. The participants will take a brief reflection survey that allows us to visually represent our takeaways (text analysis and plots). This will be the starting point for a final discussion.

2.3 Recruitment and Dissemination

We will promote this workshop through the MMLA SIG mailing list and the official CROSSMMLA website (crossmmla.org). This workshop will hold special interest for anyone interested in GenAI tools for MMLA 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 maximize outreach, we will leverage multiple channels, including our website, the MMLA mailing list (https://groups.google.com/g/crossmmla), and our network of colleagues. We will also actively promote the event through our individual social media platforms such as LinkedIn, X (formerly Twitter), and other platforms. The official event hashtag, #crossmmla, will be used across all platforms to centralize discussions and updates.

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.

3 INTENDED OUTCOMES

First, the workshop aims to equip the participants with practical skills and a deeper understanding of integrating Generative AI tools into multimodal learning analytics research. Participants will develop hands-on expertise in applying GenAI to automate data collection, analysis, and feedback generation across different modalities. They will also engage in problem-solving activities that address key challenges such as bias, data privacy, and scalability in GenAI-driven educational research. A significant focus will be placed on the ethical implications of using GenAI, encouraging participants to consider fairness, transparency, and societal impact.

Second, for the community at large, this workshop will generate best practices and methodological insights, offering practical guidance for implementing GenAI in various educational contexts. The discussions will foster multidisciplinary collaboration, helping to build a foundation for future cross-disciplinary initiatives. Additionally, the outcomes from the workshop, including key insights, challenges, and solutions, will be synthesized into a publication or report, contributing to the learning analytics and education research communities. By striking a balance between practical skills, ethical reflection, and collaborative research, the workshop will advance the integration of GenAI in multimodal learning analytics.

REFERENCES

Alwahaby, H., Cukurova, M., Papamitsiou, Z., Giannakos, M. (2022). The Evidence of Impact and Ethical Considerations of Multimodal Learning Analytics: A Systematic Literature Review. In: *The Multimodal Learning Analytics Handbook*. Springer, Cham. https://doi.org/10.1007/978-3-031-08076-0_12

Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 102–106. https://doi.org/10.1145/2460296.2460316

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., … Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 1–27. https://doi.org/10.1080/0144929X.2024.2394886

Schneider, B., Davis, R., Martinez-Maldonado, R., Biswas, G., Worsley, M., & Rummel, N. (2024). Stepping Outside the Ivory Tower: How Can We Implement Multimodal Learning Analytics in Ecological Settings, and Turn Complex Temporal Data Sources into Actionable Insights?. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024*, pp. 323-330. International Society of the Learning Sciences. https://doi.org/10.22318/cscl2024.259119

Prinsloo, P., Slade, S., & Khalil, M. (2023). Multimodal learning analytics—In-between student privacy and encroachment: A systematic review. *British Journal of Educational Technology*, *54*(6), 1566-1586. https://doi.org/10.1111/bjet.13373

Yan, L., Martinez-Maldonado, R., & Gasevic, D. (2024, March). Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 101-111). https://doi.org/10.1145/3636555.3636856

Professional Learning Analytics: Using workplace data for professional learning and development

Anna Janssen(1*), Emma Nicholls(1), Bernard Bucalon(1), Tobias Ley(2), Martin Pusic(3), Judy Kay(1), Tim Shaw(1)
1.The University of Sydney, 2. University for Continuing Education Krems 3. Harvard Medical School
*anna.janssen@sydney.edu.au

ABSTRACT: The Learning analytics (LA) community has acknowledged the educational value of data that is generated through professional interactions with workplace technologies. These datasets are largely underused and could support new just-in-time, adaptive, and formative learning opportunities. Despite this, the area of workplace and professional learning analytics (WPLA) has received less exploration when compared to LA in more traditional instructional contexts. This workshop is designed to bring together researchers and practitioners to showcase what has been done to date, discuss key challenges in developing the area further and set an agenda for future development in this space. We further aim to build a subcommunity of stakeholders (i.e., researchers, industries, educators, and professionals) interested in this area.

Keywords: professional learning environments, workplace learning analytics, informal learning

1 INTRODUCTION

Continuing professional learning is essential for professionals to ensure that they remain effective in their role beyond the completion of formal training or qualification. This is an important facet of lifelong learning, both formal and informal. When professionals engage with workplace technologies, these interactions leave behind a significant digital footprint that remains underutilized for professional development. Such interactions have the potential to unlock data-driven insights related to professional practice that could improve a professional's understanding of their own practice, scaffold reflection, and drive learning and development. Despite the opportunity to leverage data to support such ideas, many challenges exist. The aim of the workshop is to bring together learning analytics (LA) researchers to discuss the exciting, but emerging, space of professional learning, and begin building a cohesive subcommunity of researchers interested in the area. Moreover, emphasis will be placed on understanding how the established broader field of LA can guide both WPLA research and implementation.

2 BACKGROUND

Upon completion of their initial training or qualifications many knowledge workers (e.g., doctors, teachers, architects, lawyers and engineers) engage in professional learning activities to remain up to date with their subject matter knowledge and skillset. In some contexts, these activities may be formally mandated by regulatory bodies in order for knowledge workers to retain professional qualifications (Karas et al., 2020). Even in contexts where learning is not mandated, knowledge workers are often expected to take responsibility for their own personal learning to maintain their professional competence (Kooken et al., 2007). Learning in such contexts can be both formal and informal. Formal learning includes activities such as attending conferences or courses, whereas informal learning includes incidental conversations and learning "on the job". For the latter in particular, individuals may not even be aware that they are learning; it can be both opportunistic and unintended (Eraut, 2004; Littlejohn & Margaryan, 2014). Thus, professionals may miss the opportunity to learn, or it may not be realized.

A significant amount of professional learning activity occurs in the workplace. Workplace learning includes the activities individuals complete at their place of work to improve their competence or based on personal interest (Eraut, 2004; Ruiz-Calleja et al., 2019). Coupled with this, there is an increasing focus on making certain types of professional learning more "data-driven" Tavares et al., 2024). A small but growing body of research on professional learning in LA is focused on workplace learning analytics (WPLA). WPLA focuses on the collection of data in the workplace to support learning (Ruiz-Calleja et al., 2019). Some early work in WPLA include supporting self-regulated learning (SRL) (Siadaty et al., 2012), social learning (Khousa et al., 2015), computer-supported reflective learning (Pammer et al., 2017) and the application and exploration of WPLA into Knowledge Indicating Events (KIEs) (Schoefegger et al., 2010) which are are non-invasive techniques that can identify a user's knowledge based on their interactions with technology (Schoefegger et al., 2010). The Learn-B environment was designed to support knowledge workers in their self-regulated learning processes (Siadaty et al., 2012). The research prototype implemented analytics-based features such as "social waves", progress-o-meters, knowledge sharing profiles, and motivational messages. Pammer et al. (2017) developed a cyclical model of reflective learning at work based on observing reflection in practice, designing ICT that supports workplace reflection, and deploying technology in multiple field trials. Although there is significant potential to enrich workplace learning using learning analytics knowledge and methodology, the sub-field of WPLA has received much less exploration (Ruiz-Calleja et al., 2021) and many challenges exist.

Workplace learning is often informal in nature (Littlejohn et al., 2022) and, unlike platforms like Learning Management Systems (LMSs) or Massive Open Online Courses (MOOCs) that are often leveraged within LA, WPLA often leverages technologies and data that were not initially designed for learning. This makes the "measurement, collection, analysis and reporting of data about learners [professionals] and their contexts [workplaces]" (Siemens & Gašević, 2012), a challenge. There are nuances within professional learning that need to be considered when designing WPLA. For example, professional learning places emphasis on collaboration, active learning and reflection, sustained duration, coaching, and workplace conditions (Cirkony et al., 2024), which may be of less relevance for the traditional LA applications. Moreover, Kump et al. (2012) identified the challenge in creating robust learner models using workplace data as there are few mechanisms to generate explicit feedback on learning in informal settings. Thus, it is not surprising that the main barrier in realizing the potential of WPLA, is how to extract, analyze, and leverage non-traditional data sources to support informal workplace learning. Whilst the workplace offers potentially valuable data to support individual, team, and organizational learning, there is a dearth of knowledge on how to meaningfully leverage it. There is a need for effective scaffolding to support reflection on practice or the understanding of learning triggers. Without this, professionals may lack the capability to interpret data about their practice, learn from prior experiences, or generate new knowledge (Dennerlein et al., 2014).

There are many opportunities to further develop and shape professional learning (Pammer-Schindler et al., 2022), as well as to support collaborative learning knowledge creation in professional learning (Rodriguez-Triana et al., 2020). These include the development of interventions that support adaptive and just-in-time learning aligned with professionals' needs (Littlejohn et al., 2022); more sophisticated contextually relevant interventions e.g., healthcare (Pusic et al., 2023); and accelerated implementation in suitable professional settings (e.g., WPLA for educators in secondary and tertiary contexts). Finally, there is a lack of research into how to measure the success of WPLA such that the best-practice approaches for using the workplace data of varied professionals remain underexplored.

2.1 Evidence of Interest

There is growing interest in further developing the field of professional learning analytics. The first LAK workshop related WPLA was organized in 2016 (Ley et al., 2016) and since then a growing number of oral presentations were held at LAK over the last several years and recent special issues on Professional Learning featured several Learning Analytics submissions. In the context of healthcare, a

Community of Practice (CoP) has been developing looking specifically at data-driven professional learning in the sector since 2018. This workshop would be an invaluable opportunity to connect the CoP with the larger learning analytics community.

3 ORGANISATIONAL DETAILS

3.1 Workshop format, participation, and pre-workshop task

The workshop is scheduled as a half-day, in-person event that will take place during the pre-conference activities of LAK25 in March 2025. It is anticipated that this workshop will appeal to a broad spectrum of LAK attendees including data analysts, practitioners, and learning scientists, with the goal of attracting between 15 and 30 participants. The workshop is open to anyone interested in learning more about WPLA, discussing solutions for some of the key challenges, and setting an agenda for its future development. Workshop objectives include: (i) providing participants an overview of what is happening in the WPLA space; (ii) identifying strategies to overcome key challenges in the field; and (iii) to discuss opportunities and an agenda for growing the field. As a pre-workshop task, interested participants will be invited to complete a brief online survey to capture their expertise and interest areas in WPLA. Data from this survey will be used to inform the final design of the workshop activities to align with attendees' interests. Workshop. Attendees will have the opportunity to submit a short position statement, case study or research contributions describing new work (up to 4 pages).

3.2 Workshop activities

The half day workshop is designed to bring together like minded individuals to connect and grow their networks in WPLA. The workshop has **five** parts:

<u>Overview of PLA and an applied case study (70 mins)</u> | The first workshop will feature a series of short keynotes from invited speakers to provide an overview of professional and workplace learning analytics. This will be followed by presentation of a case study to provide a real-world example of how WPLA has been applied in the healthcare sector.
<u>Research and Innovation Showcase (90 mins)</u> | Accepted workshop papers will be presented by attendees (5 – 10 minutes each). The showcase will also feature interactive elements coordinated by the session chair, including an expert panel discussion from professional learning experts reflecting on key questions aligned with workshop attendee interests.
<u>Think-Tank Discussions (90 mins)</u> | Collaborative group-based activities based on experience/interest. Groups will brainstorm solutions to challenges and identify opportunities for future advancement in the area aligned with key themes. Prompt questions will be used, and each table will have a facilitator to guide discussions.
<u>Agenda setting and next steps (20mins)</u> | Attendees will reflect on the discussions and contribute to setting an agenda for how to grow a community of interest in this space.

3.3 Dissemination strategies

All workshop participants will have the opportunity to provide their contact email to receive a summary of the workshop discussions and post-event actions. Showcase presentations will be published in LAK companion proceedings. All individuals who provide their contact information will be provided information on joining the Practice Analytics CoP, which is a nascent/interest group that has been established by the workshop organizers to start bringing together individuals across sectors and research fields interested in harnessing workplace data to support reflective practice and learning.

REFERENCES

Dennerlein, S. et al., Making sense of bits and pieces: A sensemaking tool for informal workplace learning. In: European Conference on Technology Enhanced Learning: 2014: Springer; 2014: 391-397.

Cirkony, C. et al., (2024). Beyond effective approaches: A rapid review response to designing professional learning. *Professional development in education*, *50*(1), 24-45.

Eraut, M. Informal learning in the workplace. Studies in continuing education 2004, 26(2):247-273.

Khousa, E. et al., (2015). A social learning analytics approach to cognitive apprenticeship. *Smart Learning Environments*, *2*, 1-23.

Karas, M. et al., (2020). Continuing professional development requirements for UK health professionals: a scoping review. *BMJ open*, *10*(3), e032781.

Kooken, J. et al., (2007). How do people learn at the workplace? Investigating four workplace learning assumptions. In *Creating New Learning Experiences on a Global Scale: Second European Conference on Technology Enhanced Learning, EC-TEL 2007, Crete, Greece, September 17-20, 2007. Proceedings 2* (pp. 158-171). Springer Berlin Heidelberg.

Kump, B. et al., (2012). Seeing what the system thinks you know: Visualizing evidence in an open learner model. Paper presented at the Proceedings of the 2nd international conference on learning analytics and knowledge.

Ley, T., et al., (2016, April). Learning analytics for workplace and professional learning. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 484-485).

Littlejohn, A. (2017). Learning and Work: Professional Learning Analytics. In Lang, C., Siemens, G., Wise, A. F., and Gaevic, D., editors, The Handbook of Learning Analytics, pages 269–277. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition.

Lindstaedt, S. N. et al., (2009). Getting to know your user–unobtrusive user model maintenance within work-integrated learning environments. Paper presented at the Learning in the Synergy of Multiple Disciplines: 4th European Conference on Technology Enhanced Learning, EC-TEL 2009 Nice, France, September 29–October 2, 2009 Proceedings 4.

Pammer, V. et al., (2017). Let's talk about reflection at work. *International journal of technology enhanced learning*, *9*(2-3), 151-168.

Pammer-Schindler, V. et al., (2022). Guest Editorial: Designing Technologies to Support Professional and Workplace Learning for Situated Practice. IEEE Transactions on Learning Technologies, 15(5), 523–525. https://doi.org/10.1109/TLT.2022.3207306

Pusic, M. V., et al., (2023). Frameworks for integrating learning analytics with the electronic health record. The Journal of Continuing Education in the Health Professions, 43(1):52.

Prieto, L. P., Ley, T., de Jong, T., & Gillet, D. (2020). Social practices in teacher knowledge creation and innovation adoption: a large-scale study in an online instructional design community for inquiry learning. International Journal of Computer-Supported Collaborative Learning, 15(4), 445–467. https://doi.org/10.1007/s11412-020-09331-5

Ruiz-Calleja A, et al., (2019) An Infrastructure for Workplace Learning Analytics: Tracing Knowledge Creation with the Social Semantic Server. Journal of Learning Analytics, 6(2):120-139.

Ruiz-Calleja, A., et al., (2021). Learning Analytics for Professional and Workplace Learning: A Literature Review. IEEE Transactions on Learning Technologies, 14(3), 353–366. https://doi.org/10.1109/TLT.2021.3092219

Shum, S. et al., (2022). Framing professional learning analytics as reframing oneself. *IEEE Transactions on Learning Technologies*, *15*(5), 634-649.

Siadaty, M. et al., (2012). Learn-B: A social analytics-enabled tool for self-regulated workplace learning. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 115-119).

Siemens, G., & Gasevic, D. (2012). Guest editorial-learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 1-2.

Schoefegger, K. (2010). Towards a user model for personalized recommendations in work-integrated learning: A report on an experimental study with a collaborative tagging system. Procedia Computer Science, 1(2), 2829-2838.

Tavares, W. et al., (2024). Performance data advocacy for continuing professional development in health professions. *Academic Medicine*, *99*(2), 153-158.

Integrating Learning Analytics Systems with Improvement Science: A Collaborative Approach for Enhancing Educational Equity

Author: Ouajdi Manai
Sharjah Private Education Authority
w.manai@spea.shj.ae

Co-Author: Hunada Kanbar, PhD.
Sharjah Education Academy
hunada.kanbar@spea.shj.ae

ABSTRACT: This half-day workshop explores the potential for combining learning analytics systems, such as those embedded in Digital Learning Platforms (DLP), with improvement science methodologies to address educational challenges. Learning analytics systems provide detailed data on student engagement and performance, while improvement science offers a structured framework for continuous, data-informed improvement. The workshop aims to equip participants with strategies for designing improvement cycles that leverage data, foster collaboration between practitioners and researchers, and produce scalable, equitable interventions. Through hands-on activities and case studies, participants will explore practical applications of these methods in various educational contexts.

Keywords: Learning Analytics, Improvement Science, Educational Equity, DLP, SIS, LMS, Digital Learning Platforms, Knowledge Management, PDSA Cycles, Collaboration, Data-Informed

1 WORKSHOP GOALS

The goals of this workshop are:

- **Introduce Core Concepts and Synergy**: Participants will understand how digital learning platforms can support improvement science methodologies to tackle educational challenges, specifically in diverse student populations.

- **Highlight Case Studies**: Provide real-world examples where improvement science and learning analytics have been integrated to address performance variation, identifying root causes, and developing interventions that support equitable learning outcomes.

- **Hands-On Experience with PDSA Cycles**: Participants will design Plan-Do-Study-Act (PDSA) cycles, applying data from digital learning platforms to address problems of practice.

- **Discuss Challenges and Recommendations**: Explore the challenges of integrating learning analytics with improvement science and offer solutions, particularly around data privacy, professional development, and scaling interventions.

2 TARGET AUDIENCE

This workshop is designed for educational researchers, practitioners, data analysts, school administrators, and policymakers with an interest in applying learning analytics and improvement science to enhance equity in education. Participants are not required to have prior knowledge of improvement science but should have a basic understanding of educational data systems.

3 BACKGROUND AND RATIONALE

One of the major issues in education today is the gap between rising expectations for schools and what they can realistically achieve. This challenge is particularly evident when trying to scale research-based interventions across varied contexts. Improvement science, championed by leaders like Anthony Bryk (2015), emphasizes a problem-centered, data-informed approach to addressing systemic issues in education. By focusing on variation in performance, improvement science seeks to uncover what works, for whom, and under what conditions.

Learning analytics systems, such as those embedded in DLPs, provide a wealth of data that can inform this process. However, schools often struggle to utilize this data effectively. Improvement science offers a framework that complements learning analytics, enabling educators to analyze data, identify patterns of variation, and continuously refine interventions through collaborative inquiry.

Drawing on the work of the Carnegie Foundation for the Advancement of Teaching, which has pioneered improvement science and Networked Improvement Communities (NICs), this workshop explores how to accelerate educational improvement through disciplined inquiry and the strategic use of data from learning analytics systems. As Bryk et al. (2015) describe, the six core principles of improvement science provide a roadmap for making systematic, scalable improvements that are sensitive to local context and variability in performance.

4 WORKSHOP STRUCTURE AND TIMELINE

Half-Day Session (3.5 hours)

Session	Time	Content
Introduction to Key Concepts	45 min	• Overview of learning analytics systems and improvement science. • The six core principles of improvement science: problem-specific focus, addressing variation, systems analysis, practical measurement, disciplined inquiry (PDSA cycles), and networked communities.
Case Studies	60 min	• Case examples highlighting how learning analytics systems can identify issues such as uneven performance and guide the design of improvement interventions.

		• Example: Using LMS data to improve feedback timing in classrooms or address reading comprehension issues in STEM subjects. • Discussion of successful applications of PDSA cycles in school districts.
Interactive Group Work: Designing PDSA Cycles	85 min	• Participants form small groups to collaboratively design a PDSA cycle using sample datasets from SIS/LMS platforms. • Groups identify a specific educational challenge (e.g., disparities in student engagement) and develop an intervention to test. • Participants share their designs with the broader group for feedback and refinement.
Challenges and Best Practices	45 min	• Discussion on the practical challenges of integrating learning analytics with improvement science, including data privacy concerns and teacher training. • Facilitators present strategies for overcoming these challenges and ensuring successful collaboration between researchers and educators.
Conclusion and Key Takeaways	15 min	• Summarize workshop insights and provide participants with resources to continue exploring these methodologies in their contexts. • Outline actionable next steps for implementing learning analytics-informed improvement cycles.

5 KNOWLEDGE MANAGEMENT AND NETWORKED IMPROVEMENT COMMUNITIES (NICS)

Improvement science emphasizes the importance of social learning and knowledge dissemination across networks. The success of Networked Improvement Communities (NICs), as discussed by Bryk et al. (2015), lies in their ability to connect educators, researchers, and practitioners in a shared effort to address specific problems of practice. NICs facilitate the flow of knowledge through collaboration, ensuring that the insights gained from one context can be applied in others. This networked approach accelerates learning and fosters a culture of continuous improvement.

The SECI (Socialization, Externalization, Combination, Internalization) model of knowledge creation, proposed by Nonaka and Takeuchi (1995), can also inform how digital learning platforms support improvement work. By converting tacit knowledge (teachers' intuitive understanding of what works) into explicit, shareable insights, learning analytics platforms can serve as tools for knowledge management within NICs, enabling educators to test, refine, and spread effective practices across their networks.

6 ANTICIPATED OUTCOMES

By the end of this workshop, participants will:

- Understand the synergy between learning analytics and improvement science for data-informed educational improvement
- Understand the synergy between learning analytics and improvement science for data-informed educational improvement.
- Be familiar with the six principles of improvement science and their application in real-world educational settings
- Gain strategies for overcoming challenges related to data use, collaboration, and scaling interventions across different contexts

7 RELEVANCE TO THE LAK COMMUNITY

This workshop is directly relevant to the Learning Analytics and Knowledge (LAK) community as it bridges the gap between data generation and its practical application in educational settings. By combining improvement science with learning analytics, this workshop addresses the LAK community's interest in using data to inform continuous improvement and foster more equitable educational outcomes. Participants will leave with a deeper understanding of how to integrate these methodologies into their own work, driving impactful changes in their educational environments.

REFERENCES

Bryk, A. S. (2015). Accelerating How We Learn to Improve. Educational Researcher, 44(9), 467-477. https://doi.org/10.3102/0013189X15621543

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). Learning to improve: How America's schools can get better at getting better. Cambridge, MA: Harvard Education Press.

Lewis, C. (2015). What is improvement science? Do we need it in education? Educational Researcher, 44(1), 54-61. https://doi.org/10.3102/0013189X15570388

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. New York, NY: Oxford University Press.

.

New Horizons in Human-Centered Learning Analytics and AI in Education

[1]Riordan Alfredo, [2]Simon Buckingham Shum, [3]Mutlu Cukurova, [4]Yannis Dimitriadis, [5]Charles Lang, [1]Roberto Martinez-Maldonado, [6]Patricia Santos, [7]Paraskevi Topali, [8]Olga Viberg

[1]Monash University, Australia; [2]University of Technology Sydney, Australia; [3]University College London, UK; [4]Universidad de Valladolid, Spain; [5]Columbia University, USA; [6]Universitat Pompeu Fabra: Barcelona, Spain; [7]Radboud University, Netherlands; [8]KTH Royal Institute of Technology, Sweden
Corresponding author: riordan.alfredo@monash.edu

ABSTRACT: This workshop will explore new horizons in Human-Centered Learning Analytics and Artificial Intelligence (AI) in education, focusing on research, design, and development practices that enhance educational systems. By aligning closely with pedagogical intentions, preferences, needs, and values, these systems aim to amplify and augment the abilities of all educational stakeholders. By examining alternative frameworks and addressing the broader implications of technology for humanity, this workshop aims to foster responsible, inclusive, value-sensitive and sustainable data-powered solutions. This way, we strive for enhanced educational experiences while respecting the agency and well-being of educators and learners, as well as our social bonds and the environment.

Keywords: design, values, well-being, learning/educational environment, human-centered, HCI, Artificial Intelligence (AI), learning analytics (LA)

1 INTRODUCTION

Over the past few years, the field of Learning Analytics (LA) has evolved to place a stronger emphasis on design approaches that consider human and cultural values, learning and educational stakeholders' authentic needs at the centre, with the integration of participatory design practices marking a key shift since 2018 LAK's conference's theme. Human-Centered Learning Analytics (HCLA) has since emerged as a key area of focus, promoting the incorporation of human values, ethical practices, and thoughtful design in the creation of educational analytics (Buckingham Shum et al., 2024) and AI systems (Cukurova, 2024). As Human-Centered AI (HCAI) gains traction across various domains, there is a growing need to ensure that LA and AI in education technologies are effective, ethically sound, learner-focused, and aligned with human values, the pedagogical intentions and broader needs of society. While the concept of "human-centeredness" continues to spark debate and inspire alternative frameworks, the underlying goal remains consistent: **to develop analytics and AI systems that genuinely enhance learning experiences while safeguarding the autonomy and well-being of educational stakeholders**. This workshop will explore the latest developments in Human-Centered LA and AI, examining alternative approaches and addressing the wider impact of these technologies on education. We aim to deepen the discourse around creating ethical, inclusive, and sustainable solutions that meaningfully contribute to education.

2 BACKGROUND

Since the initial efforts to integrate participatory practices into LA back in 2018, the field of HCLA has been steadily gaining momentum (Buckingham Shum et al., 2024). This approach not only emphasises the importance of participatory and co-design practices (Sarmiento & Wise, 2022) but also aims to ensure that human and cultural values (Viberg et al., 2023), responsible practices (Pargman et al., 2023), and design principles are at the forefront when developing and applying analytics and AI systems in educational contexts. These considerations are crucial for creating socio-technical LA systems that support and enhance teaching and learning, aligning technological advancements with the needs and values of educators and learners.

Beyond the scope of LA, Human-Centered AI is emerging as a rapidly expanding research field (Shneiderman, 2022; Capel & Brereton, 2023). This area of study emphasises the integration of human and cultural values, ethical considerations, and user-centric design in the development and application of AI systems (e.g., Viberg et al., 2023). By prioritizing the needs, experiences, and well-being of individuals, HCAI aims to create more responsible, trustworthy, and impactful technologies that align with societal goals and enhance human agency.

Although the term "human-centeredness" is often used ambiguously by various authors (Lang & Davis, 2023) and has been the subject of critique within broader design communities (Norman, 2023), two recent systematic literature reviews have provided a comprehensive overview of how HCLA and AI in education systems have been applied in diverse ways (Alfredo et al., 2024; Topali et al., 2024). These reviews highlight the evolving interpretations and implementations of human-centered principles, shedding light on the various approaches taken to integrate these values into data-intensive educational technologies.

While terms like activity-centered (Gifford & Enyedy, 1999) and even humanity-centered (Norman, 2023) design have been proposed as alternatives for human-centeredness, the core concern remains the same: to develop analytics and AI systems that not only support learning but also respect the agency of educational stakeholders. These systems must address the human-related challenges that arise from potential datafication (Williamson et al., 2020), while also considering their impacts on learners' and educators' wellbeing, the environment and our social structures.

For this reason, this workshop will aim to explore new horizons in human-centered LA and AI in education. By delving into these alternative frameworks and addressing the broader implications of technology in education, we aim to foster a deeper understanding of how to create responsible, ethical, inclusive, value-sensitive and sustainable solutions that truly enhance the educational experience and empower educational stakeholders' autonomy and well-being.

2.1 Evidence of interest

This workshop seeks to build on the momentum from recent years within the LAK and technology-enhanced learning (TEL) communities. There has been a growing interest in this area. The first related workshop was the *LAK Participatory Design workshop* at LAK18 (the theme of LAK18 was *Towards User-Centred Design*), providing an identity to this particular strand of work (Prieto-Alvarez et al., 2018). Then, the first edition of the HCLA workshop happened at LAK21, with subsequent editions at ECTEL21, LASI, LAK22 and LAK23. Some of the co-organisers of this workshop are also involved in the publication of special issues in the Journal for Learning Analytics (Buckingham Shum, et al., 2019) and a Special Section in the British Journal of Educational Technologies (Buckingham Shum, et al., 2024; Viberg et al., 2024).

3 ORGANISATIONAL DETAILS

3.1 Workshop format, participation, and pre-workshop task

- The workshop is envisioned to be an in-person, half-day session. Between 12 and 24 participants, with a shared interest in Human-Centered LA and AI in education, are expected to be part of this workshop. We welcome everyone with an interest in the field, from beginners to experts. We will have a call for papers to welcome more elaborated contributions to this area. The 2-4 pages non-compulsory POSITION papers will be peer-reviewed by members of the organisation team and authors of the papers. All workshop participants will gain access to the submitted and accepted papers before the workshop, which will be discussed during the event.
- Participants not submitting position papers will be asked to complete a survey that will capture previous experiences in this area and current understandings of design aspects that will be relevant to the discussions during the workshop. In particular, participants will be asked to share their views on the new horizons of Human-Centered LA and AI in education.

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. The workshop is divided into four parts:

1. **Overview and Introduction:** The workshop will begin with an overview of key insights drawn from our preliminary survey results and position papers. This segment will introduce several new horizons and perspectives that have been identified as critical for advancing the field of Human-Centered LA and AI in education.
2. **Modified Pecha-Kucha Poster Presentation:** In the second part, authors of position papers will present brief overviews of their ideas. Each presenter will be allotted 2 minutes, during which they will present up to 3 slides, with each slide displayed for 40 seconds. This format will allow for concise and focused presentations, offering a snapshot of diverse perspectives.
3. **Sharing and Guided Critique:** The third part will transition into a collaborative discussion centered on the experience of position papers. To enrich the dialogue, we will invite discussants from related fields—such as human-computer interaction, interaction design, participatory design, and information visualization—alongside critics of human-centered design methods. These discussants will offer critical perspectives on the ideas and design plans presented, fostering a dynamic discussion on the strengths and limitations of techniques in the context of LA and AI in education and providing constructive feedback to the presenters.
4. **Discussion on Next Steps:** The workshop will conclude with an open discussion, inviting all participants to contribute ideas for setting a potential research agenda for HCLA and AI in education. This collaborative effort aims to outline future directions and priorities for research and development in this evolving field. The discussion may result in a manifesto or a document identifying key challenges in HCLA and AI, which could inspire more concrete and actionable outcomes.

3.3 Dissemination strategy

An event website will be made available upon acceptance of this workshop. A call for participation will be generated and published via the website, and through the LinkedIn or Twitter accounts and mailing lists the workshop organisers can access. The website will also include an overview of the workshop's aims, information about the workshop organisers, contact details and reports and other outputs. The accepted papers will be published on the workshop website and open-access publication platform (e.g., CEUR-WS).

3.4 Logistics and tools

The workshop will be conducted in a room that enables collaboration and is equipped with a projector.

REFERENCES

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. Computers and Education: Artificial Intelligence, 6, 100215. https://doi.org/10.1016/j.caeai.2024.100215

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

Buckingham Shum, S., Martínez-Maldonado, R., Dimitriadis, Y., & Santos, P. (2024). Human-Centred Learning Analytics: 2019–24. Br. J. Educ. Technol., 55(3), 755–768. doi: 10.1111/bjet.13442

Capel, T., & Brereton, M. (2023). What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–23. https://doi.org/10.1145/3544548.3580959

Cukurova, M. (2024). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, 1–20. https://doi.org/10.1111/bjet.13514

Gifford, B. R., & Enyedy, N. D. (1999). Activity centered design: Towards a theoretical framework for CSCL. Proceedings of the 1999 Conference on Computer Support for Collaborative Learning - CSCL '99, 22-es. https://doi.org/10.3115/1150240.1150262

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

Norman, D. (2023). Design for a Better World: Meaningful, Sustainable, Humanity Centered. MIT Press.

Pargman, T. C., McGrath, C., Viberg, O., & Knight, S. (2023). New Vistas on Responsible Learning Analytics: A Data Feminist Perspective. Learning Analytics, 10(1), 133–148. doi: 10.18608/jla.2023.7781

Prieto-Alvarez, C., Martinez-Maldonado, R., & Shum, S. B. (2018). *Mapping Learner-Data Journeys: Evolution of a Visual Co-Design Tool*. Paper presented at the ACM Australian Computer-Human Interaction Conference, OzCHI'18.

Sarmiento, J. P., & 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* (pp. 535-541).

Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press.

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M. J., Er, E., Khalil, M., & Açkapınar, G. (2024). Designing human-centered learning analytics and artificial intelligence in education solutions: A systematic literature review. Behaviour & Information Technology, 0(0), 1–28. https://doi.org/10.1080/0144929X.2024.2345295

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

Viberg, O., Kizilcec, R. F., Wise, A. F., Jivet, I., & Nixon, N. (2024). Advancing equity and inclusion in educational practices with AI -powered educational decision support systems (AI - EDSS). British Journal of Educational Technology, 55(5), 1974–1981. https://doi.org/10.1111/bjet.13507

Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in Higher Education: Critical issues and perspectives. Teaching in Higher Education, 25(4), 351–365. https://doi.org/10.1080/13562517.2020.1748811

What are the Grand Challenges of Learning Analytics?

Kirsty Kitto
University of Technology Sydney
kirsty.kitto@uts.edu.au

Catherine A. Manly
Fairleigh Dickinson University
c.manly@fdu.edu

Oleksandra Poquet
Technical University of Munich
sasha.poquet@tum.de

Rebecca Ferguson
The Open University
rebecca.ferguson@open.ac.uk

ABSTRACT: In line with the conference theme, this workshop will "expand the horizons" of Learning Analytics (LA) by bringing together researchers and practitioners from a wide variety of backgrounds to create a community-accepted list of grand challenges. It will work towards finding common elements in various existing research programs and mapping out the new research avenues that are deemed most interesting by the community. This will help the LA community to point to well established "blue skies" requiring more work when applying for funding and large grants. It will also support more junior researchers in seeing the bigger picture when plotting out their research trajectory.

Keywords: Grand Challenges; Theory; Evidence; Synthesis

1 THE PROBLEM

What are the grand challenges of Learning Analytics (LA)? Where is our theoretical contribution and what specifically are we adding to the field of education? Although some attempts have been made to highlight how LA might address large scale challenges (Buckingham Shum, 2023) and indeed, a list of grand challenges for the field has been put forward (Baker, 2019), we are yet to coalesce around a community-defined set of research priorities. Against this backdrop, the aims of large LA research groups are not always aligned, and there have even been recent bandwagon effects where a 'hot topic' emerges and distracts attention from areas with potential for benefitting the field. Most crucially, new entrants to the field and early career researchers sometimes find it difficult to understand the rich background landscape of the field, and why certain problems have been identified as important. Without a clear unifying set of challenges, it is likely that LA will make only incremental contributions, if any. We are at risk of becoming feudalistic, with various teams staying within their safe, identified subfields.

While education itself is often touted as a field that will help us to create a more equal and just society, LA is sometimes accused of supporting agendas that will track people, violate their privacy, and manipulate them towards acts that they might not have undertaken on their own. How can we work towards ensuring that the field is solving big issues that help to ensure the next generation of people are more mindful and accepting of each other and the differences between us, respond less to the abundance of false information, and are able to adjust in ways that are well reasoned rather than simply reactive to societal shifts and new challenges?

2 THE SOLUTION

A number of research disciplines have coalesced around an established list of grand challenges, resulting from periodic workshops bringing together key members of that field. Perhaps the most famous effort in this space was instigated by Hilbert, who in 1900 proposed a set of 23 unsolved problems in mathematics that drove research throughout the 20th century (indeed they still do as only 15 of them have been solved to date). This programmatic approach to defining a field has inspired other research domains. For example, the field of Information Retrieval has held three SWIRL[1] workshops (in 2004, 2012, and 2018) where leading figures in the field were invited to attend and define challenges and opportunities for the field. This type of prioritization can both guide future research for new entrants (Allen, et al., 2012), and support proposals for funding and tenders.

Similar attempts have been made in the educational sciences, but the results have yet to achieve broad impact upon the directions of LA as a field. For example, in 2011 a STELLAR workshop organized as a part of the Alpine Rendezvous series (Mwanza-Simwami, et al. 2011) proposed six grand challenges for Technology-Enhanced Learning (TEL) in Europe but to date the white paper that emerged has received zero citations which leads us to believe that despite deepening the scholarly discussions in this area within the interested community, its broadscale impact upon the field has been limited. So far, none of the proposed lines of work has been pursued to the point of completion or resolution, despite the paper listing likely timeframes and measurable indicators of success. While we believe such workshops that identify grand challenges hold potential for disciplinary scholarship through collaborative development of priorities, the question of how specific the challenges of LA might be, and how they differ from those of TEL remains unanswered. Grand challenges that pertain to artificial intelligence in education (AIED) have also been pulled from other branches of computer science (Kay, 2012) in a process that, while useful for defining relevant challenges and sparking research, has not necessarily been organic to the AIED community itself. Furthermore, their specificity to AIED marks those challenges as potentially too restrictive for our field. Baker's (2019) grand challenges have gained considerable attention (they have been cited 129 times), but are particularly aimed at educational data mining. While they might represent a starting point for discussion, we do not consider them representative of the work occurring in the broader LA community, leaving an opportunity that this workshop will address.

It is time for the learning analytics community to "expand our horizons" by collaboratively defining the problems and opportunities of the field. This workshop will attempt to bring together a range of stakeholders with different voices and backgrounds to define a set of community-accepted and supported grand challenges that can drive the next 10 years of LA research and development.

3 THE WORKSHOP

This full-day interactive workshop devoted to the problem of identifying and then refining a series of grand challenges that the broader LAK community identifies with and considers highly worthy topic for further research. This outcome will help to guide research in the field over the next decade. In

[1] Strategic Workshop in Information Retrieval in Lorne, see https://sites.google.com/view/swirl3/ for more details.

addition to the grand challenges, a set of enabling problems will also be identified, and various LA subfields will be mapped into the different programs of work. This will help established members of the LA community, as well as new entrants and early career researchers, to place their programs of research within the broader field. It will also guide the community as it responds to funding calls and global research initiatives.

3.1 An interactive workshop structure

Participants in the workshop should expect a highly interactive format which encourages cross-field discussion, and collaboration across different teams. A series of group-based activities will synthesise various research programs and challenges across the LA community, exploring their dimensions and considering how they might relate to one another.

We strongly encourage all participants to attend for the full day, as the workshop is designed to scaffold all participants through the process of discovering and refining a draft set of grand challenges that can be taken to the broader community. Note that if you cannot attend this workshop there will follow-up activities throughout LAK25 where space will be provided to inspect, refine, and add to the challenges identified on during the workshop.

3.2 Dissemination and communication

This structure means that the outcomes of the workshop will be disseminated both during LAK, and after, as follows:

- A poster will be created by workshop participants and displayed during the poster session. This poster will include outcomes from the workshop that can be considered, challenged and modified by input from the broader LA community.
- A follow-up interactive panel session will be run on Friday morning of LAK. During this panel session, attendees will be introduced to the grand challenges that emerged from the workshop and invited to further comment upon and refine them.
- Finally, an opinion paper will be created post workshop, for publication in the open peer commentary section of the Journal of Learning Analytics, enabling workshop participants and the broader LA community to engage with the workshop outcomes, extend them, and/or propose amendments.

3.3 Preparation for attendance

All registered participants will be asked to undertake some activities in the leadup to the workshop which both prepare them for the workshop and help to ensure that its outcomes are representative of work in the broader LA community. They will receive a pre-conference survey to get them thinking about the grand challenges faced by LA. They will also be asked to nominate papers that they believe offer promising avenues for future work in the field, along with explanations as to why. This list of papers will be curated and thematically categorized by the organizers with the result used to seed initial discussions at the workshop itself. A small sample of papers will be selected from this nominated list, and circulated prior to the workshop for participants to examine and consider.

3.4 Tentative schedule

Time	What	Details
0930	Introduction	Discuss reason for workshop and objectives. Introduce key papers nominated during pre-conference process and invite any authors in the workshop to speak to why they wrote the paper and what they think has come from it to date.
1000	Phase 1: Icebreaker	Bring your challenge and introduce it to the workshop.
1100	Phase 2: Support and brainstorming	For a challenge to make it to phase 2, it must find a person (other than the original proposer) to support it.
1200	Lunch	Work and discussion will continue over lunch.
1300	Phase 3: Refinement	Groups will give flash presentations of their challenge and participants will have a chance to change to new teams to augment that idea. Groups will map emerging grand challenges to existing work in LA and the learning sciences.
1400	Communication 1: The poster and panel	As their challenge is finalized, groups will gradually shift to creating a poster and planning a panel to share the proposed challenges with all LAK attendees later during the conference.
1500	Communication 2: The mock funding panel	The person leading support for each challenge will present the group's work and make a case for it being a grand challenge. Each workshop attendee will then be given a fixed number of units that they can allocate to challenges as they like. This allocation will be used to rank the challenges for the poster.
1600	Wrap up	Discuss next steps throughout the conference and recruit volunteers for poster session and final panel during LAK25. Plan for post conference activities.

REFERENCES

Allan, J., Croft, B., Moffat, A., & Sanderson, M. (2012). Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. In *ACM SIGIR forum* (Vol. 46, No. 1, pp. 2-32). New York, NY, USA: ACM. http://sigir.org/files/forum/2012J/2012j_sigirforum_A_allanSWIRL2012Report.pdf

Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *Journal of educational data mining*, *11*(1), 1-17.

Buckingham Shum, S. (2023). Trust, Sustainability and Learning@Scale. Keynote Address, *Proceedings of the Tenth ACM Conference on Learning@Scale (L@S '23).* Association for Computing Machinery, New York, NY, USA, pp. 1–2. https://doi.org/10.1145/3573051.3593375.

Kay, J. (2012). AI and education: Grand challenges. *IEEE Intelligent Systems*, *27*(5), 66-69. doi: 10.1109/MIS.2012.92

Mwanza-Simwami, D., Kukulska-Hulme, A., Clough, G., Whitelock, D., Ferguson, R., & Sharples, M. (2011). Methods and models of next generation technology enhanced learning - White Paper. In: Alpine Rendezvous 2011, 28-29 March 2011, La Clusaz, France. https://oro.open.ac.uk/29056/

Foundational ideas and advanced methods in Quantitative Ethnography

Golnaz Arastoopour Irgens
Vanderbilt University, US
golnaz.arastoopour.irgens@vanderbilt.edu

Kamila Misiejuk
FernUniversität in Hagen, Germany
kamila.misiejuk@fernuni-hagen.de

Rogers Kaliisa
University of Oslo, Norway
rogers.kaliisa@iped.uio.no

Shaun Kellogg
NC State University, US
sbkellog@ncsu.edu

Jennifer Scianna, Brendan Eagan, Yeyu Wang, Yuanru Tan
UW Madison, US
jscianna@wisc.edu, beagan@wisc.edu, ywang2466@wisc.edu, yuanru.tan@wisc.edu

Zach Swiecki
Monash University, Australia
zach.swiecki@monash.edu

ABSTRACT: This workshop delves into the foundational concepts and emerging methods in Quantitative Ethnography (QE), focusing on its application within Learning Analytics (LA). As researchers face challenges in analyzing large-scale qualitative data, QE offers a solution by combining qualitative richness with statistical rigour. The workshop, extending from the previous LAK24 session, emphasizes technical proficiency with QE tools, such as the ENA web tool, rENA, and BERT-topic, alongside traditional methods. From novices to intermediates, participants will engage in hands-on activities and discussions centered around data coding, model creation, theoretical saturation, and closing the interpretive loop, enhancing their capacity to integrate QE in their research practices.

Keywords: Quantitative Ethnography, Epistemic Network Analysis, Learning Analytics

1 BACKGROUND AND PURPOSE

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 helpful 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 primary purpose of this workshop is to extend last year's LAK workshop on Integrating QE Methods to Support LA in the Age of AI to focus more on the technical aspects of using QE tools in LA research, including a greater emphasis on tools in R and emergent advanced techniques and theory. *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 modeling patterns of connections in discourse, ENA can help researchers quantify and visualize learning over time for individuals and groups, compare learning across learners or contexts, create learning trajectories, and model individuals' contributions to group discourse (Shaffer et al., 2016). While we will cover using the web tool version of ENA, we will focus more on using the R package RENA and connecting it with other analytic techniques. In addition, this workshop will address the critical steps of qualitative data preprocessing, coding, and closing of the interpretive loop. These steps are significant and tightly connected to the theoretical grounding of learning analytics research (Munk et al., 2017), yet they have yet to be considered and discussed. Leveraging the newer and more advanced QE techniques 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 interpretive 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 full-day hands-on workshop where the participants (a) will be introduced to the foundations of the QE process and (b) learn about and engage with four research topics within the QE framework using a hands-on interactive approach introducing tools, such as ENA webtool, rENA, and BERT-topic. These activities will be grounded in QE theory and inform a discussion of the philosophical and methodological foundations for QE analysis in learning analytics. This workshop is aimed at participants who are new to the QE approach and intermediate-level participants who would like to deepen their knowledge, specifically in the domain of doing QE analysis using R. Participants from any discipline backgrounds and prior knowledge levels interested in integrating qualitative and quantitative methods in their research can benefit from this workshop.

2.1 Recruitment

To recruit participants for our workshop, we will utilize a multifaceted approach, leveraging various channels and networks. Our recruitment strategy includes individual invitations, outreach through social media platforms, announcements on the Learning Analytics Google Group, and promotion via the conference website. The workshop organisers are also actively involved with the International Society of Quantitative Ethnography, which has its own website, newsletter, and mailing list. We will utilise these resources to publicize the workshop and attract potential participants. The organizers will tap into their professional and university alliance networks affiliated with their institutions to engage both target groups effectively. In addition, we extend an invitation to over 170 current and former scholars who have participated in the NSF-Funded Learning Analytics in STEM Education Research (LASER) Institute, many of whom have attended and presented at prior LAK conferences as part of the program. Finally, building on the success of our previous ENA and SNA workshops held at LAK24 in Kyoto, we will extend personal invitations to last year's participants, encouraging them to join this year's event. We are targeting a participant group of 25-30 individuals for this workshop. We anticipate having no difficulty recruiting a diverse group of early and experienced scholars interested in incorporating ENA into their research and teaching practice.

2.2 Structure

During the full-day workshop:

1. All participants will learn about the QE methodology and foundations.

2. All participants can choose to participate in four mini-workshops:

    a. *QE Data preparation and coding*: to learn about data formatting, coding for different data types, coding methods for large datasets, sampling, and coding validation.

    b. *ENA webtool menu*: to learn about specifying parameters, interpreting models, and closing the interpretive loop by diving back to qualitative data.

    c. *rENA:* to learn about reaching theoretical saturation through statistical models and running statistical models outside of the ENA web tool.

    d. *BERTopic:* closing the interpretive loop using computational methods like Neural Topic Modeling.

2.3 Workshop schedule

Time	Activity
9:00 - 9:15	**Introduction**
9:15 - 10:00	**QE data preparation and coding**
10:00 - 10:15	Coffee break
10:15 - 12:00	**Workshop 1:** Participants choose from ENA webtool, rENA or BERTopic

12:00 - 13:00	Lunch break
13:00 - 14:45	**Workshop 2:** Participants choose from ENA webtool, rENA or BERTopic
14:45 - 15:00	Coffee break
15:00 - 16:00	**Group discussion and reflections**
After 16:00	Social activity

2.4 Required equipment

A projector and screen will be required by organisers, as well as tables for collaboration. Attendees will need to bring laptops and adequate internet connectivity to participate in planned activities and access curriculum materials. Specifically, participants will need to access a freely available website that will house all the curriculum materials needed for the workshop. The website will include materials for each activity, including slide decks, coding activities, tutorials, and assessment activities. The source code for all instructional materials will also be housed on GitHub, and coding activities will be made available through Posit Cloud, which allows users to participate fully in the workshop without installing any data analysis software.

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 *Proceedings of the ICQE* (pp. 3-16).

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.

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 LAK* (pp. 569-578).

Writing analytics in the age of large language models: Shaping new possibilities for assessing and promoting writing

Rianne Conijn[1], Antonette Shibani[2], Laura Allen[3], Simon Buckingham Shum[2], Cerstin Mahlow[4]
[1]Eindhoven University of Technology, the Netherlands
[2]University of Technology, Sydney, Australia
[3]University of Minnesota, USA
[4]ZHAW School of Applied Linguistics, Switzerland

Email: m.a.conijn@tue.nl

ABSTRACT: Generative Artificial Intelligence applications powered by large language models (LLMs) have significantly influenced education and in particular, reimagined writing technologies. While LLMs offer huge potential to provide automated writing support to learners it is also important to identify challenges they bring to learning, assessment, and critical interaction with AI. This workshop aims to shape possibilities for writing analytics to promote and assess learning-to-write and writing-to-learn that are appropriate for the generative AI era. In this Seventh workshop of the Writing Analytics series, we propose a symposium-style format to identify how the field can unravel in the age of LLMs. In particular, we focus on (case) studies within two topics: (1) using writing analytics to design and evaluate interactive writing support systems and (2) using writing analytics to evaluate human-AI interactions. In addition, this workshop will serve as a community-building event to invigorate the SOLAR writing analytics community. An overview of the workshop and workshop outcomes will be available at https://wa.utscic.edu.au/lak25-workshop-on-writing-analytics-in-the-age-of-large-language-models/.

Keywords: Writing Analytics, Generative AI, Human-AI Co-Writing, Keystroke Logging, Writing Process, LLM

1 INTRODUCTION

The advent of Large Language Models (LLMs) has had a significant influence on a wide variety of aspects in education, which has resulted in an increasing academic interest within the field of learning analytics as well as education in general (Khosravi et al., 2023). LLMs come with ample opportunities as well as challenges, including threats to academic integrity, overreliance, fairness, privacy concerns, and reduced critical thinking (Kasneci et al., 2023; Memarian & Doleck, 2023). While LLMs are increasingly adopted by learners and educators, it is crucial that learning does not get compromised (Memarian & Doleck, 2023). In this workshop, we specifically focus on the effects of LLMs on writing, with the aim to identify how to promote and assess writing in the age of LLMs, aligning with the goals of writing analytics to support 'learning-to-write' and 'writing-to-learn' effectively in educational contexts (Gibson & Shibani, 2022).

Using LLMs in writing has resulted in various modes of human-AI interaction including co-authoring with AI and multimodal writing assistance powered by LLMs, that envision new forms of writing support (Lee et al., 2024). However, a majority of articles on the use of LLMs still revolves around case studies and opinion pieces (Khosravi et al., 2023; Memarian & Doleck, 2023). There is limited empirical research on the effects of (writer-initiated) use of LLMs and AI-based writing support on learners' writing as well as potential contextual factors (including individual differences, task design, course design, ethical considerations) that might influence the effectiveness.

The small − but increasing − number of evaluation studies on LLMs often still adopt a system-centric view, focusing primarily on the accuracy of the system, or focusing on the perceptions on the user, with limited emphasis on the human-AI interactions evolving during the writing process (Lee et al., 2023). To comprehensively understand learning-to-write and writing-to-learn in the age of LLMs, one needs to focus on written product and user perspectives, but also on the objective user interactions (Shen & Wu, 2023), that is, the (human-AI) writing process. This aligns with the goal of writing analytics, which aims at understanding both the writing product and process, as set out in the first workshop (Buckingham Shum et al., 2016).

In this seventh writing analytics workshop, we aim to invite the current SoLAR Writing Analytics community[1] to envision and shape possibilities for the field in the age of LLMs. In particular, we aim to focus our attention on two key directions:

1. How can we design and evaluate intelligent and interactive writing support systems in effectively aiding learning-to-write and writing-to-learn in the age of LLMs? How do we define 'effectiveness'? How can we ensure learners and educators use the tools effectively? This theme might include empirical studies on the design and evaluation collaborative human-AI writing tools (for example ABScribe, Reza et al., 2024) as well as empirical studies focusing on the ethical considerations in designing and evaluating the tools, including for example trust calibration or use of explainable AI (see e.g., Shen et al., 2023).

2. How can writing analytics support the evaluation of human-AI interactions? How can we evaluate and understand the evolving use of LLMs over time? How can we deal with non-deterministic LLM output? This theme might include studies examining the use of trace analysis, such as keystroke logging and eye-tracking (Lindgren & Sullivan, 2019), authorship visualization (Shibani et al., 2023), and linguistic analysis of sentence histories (Mahlow et al., 2024).

2 WORKSHOP FORMAT AND SUBMISSIONS

We aim to bring together active researchers and practitioners in writing analytics to this half-day workshop. The workshop will run in a symposium format, with calls for participation issued for submission of an extended abstract of 500-750 words, aligning with the two key directions. We welcome contributions providing theoretical, empirical, methodological advances and or critical perspectives in both directions.

[1] https://www.solaresearch.org/community/sigs/writing-analytics-sig/

The workshop will also serve as a community-building event with some time allocated for networking, setting up (new) collaborations, and outlining future activities of the SoLAR Special Interest Group on Writing Analytics (SIGWA) community. The fact that LAK 2025 is organized in Dublin, Ireland, facilitates further possibility to connect with the (European) EARLI Special Interest Group on Writing (https://www.earli.org/sig/sig-12-writing). An overview of the workshop program including the accepted contributions as well as the workshop outcomes following the workshop will be available at https://wa.utscic.edu.au/lak25-workshop-on-writing-analytics-in-the-age-of-large-language-models/

REFERENCES

Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). Critical perspectives on writing analytics. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, 481–483. https://doi.org/10.1145/2883851.2883854

Gibson, A., & Shibani, A. (2022). Natural Language Processing: Writing Analytics. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), Handbook of Learning Analytics (Second Edition, Vol. 2, pp. 96–104). https://doi.org/10.18608/hla22.010

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/J.LINDIF.2023.102274

Khosravi, H., Viberg, O., Kovanovic, V., & Ferguson, R. (2023). Generative AI and Learning Analytics. Journal of Learning Analytics, 10(3), 1–6. https://doi.org/10.18608/jla.2023.8333

Lee, M., Gero, K. I., Chung, J. J. Y., Buckingham Shum, S., Raheja, V., Shen, H., Subhashini, V., Wambsganss, T., Zhou, D., Alghamdi, E., August, T., Bhat, A., Choksi, M. Z., Dutta, S., Guo, J. L. C., Hoque, M. N., Kim, Y., Knight, S., Neshaei, S. P., … Siangluilue, P. (2024). A Design Space for Intelligent and Interactive Writing Assistants. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–35. https://doi.org/10.1145/3613904.3642697

Lee, M., Srivastava, M., Hardy, A., Thickstun, J., Durmus, E., Paranjape, A., Gerard-Ursin, I., Li, X. L., Ladhak, F., Rong, F., Wang, R. E., Kwon, M., Park, J. S., Cao, H., Lee, T., Bommasani, R., Bernstein, M. S., & Liang, P. (2023). Evaluating Human-Language Model Interaction. Transactions on Machine Learning Research. https://openreview.net/forum?id=hjDYJUn9l1

Lindgren, E., & Sullivan, K. (2019). Observing Writing: Insights from Keystroke Logging and Handwriting. In Observing Writing. Brill. https://doi.org/10.1163/9789004392526

Mahlow, C., Ulasik, M. A., & Tuggener, D. (2024). Extraction of transforming sequences and sentence histories from writing process data: a first step towards linguistic modeling of writing. Reading and Writing, 37(2), 443–482. https://doi.org/10.1007/S11145-021-10234-6/FIGURES/14

Memarian, B., & Doleck, T. (2023). ChatGPT in education: Methods, potentials, and limitations. Computers in Human Behavior: Artificial Humans, 1(2), 100022. https://doi.org/10.1016/J.CHBAH.2023.100022

Reza, M., Laundry, N., Musabirov, I., Dushniku, P., Yu, M., Mittal, K., Grossman, T., Liut, M., Kuzminykh, A., & Williams, J. J. (2024). ABScribe: Rapid Exploration & Organization of Multiple Writing Variations in Human-AI Co-Writing Tasks using Large Language Models. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3613904.3641899

Shen, H., Huang, C. Y., Wu, T., & Huang, T. H. K. (2023). ConvXAI : Delivering Heterogeneous AI Explanations via Conversations to Support Human-AI Scientific Writing. Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, 384–387. https://doi.org/10.1145/3584931.3607492

Shen, H., & Wu, T. (2023, March 11). Parachute: Evaluating Interactive Human-LM Co-writing Systems. Proceedings of the CHI'23 In2Writing Workshop. https://doi.org/10.18653/v1/2020.emnlp-main.525

Shibani, A., Rajalakshmi, R., Mattins, F., Selvaraj, S., & Knight, S. (2023). Visual Representation of Co-Authorship with GPT-3: Studying Human-Machine Interaction for Effective Writing. Proceedings of the 16th International Conference on Educational Data Mining, 183–193. https://doi.org/10.5281/zenodo.8115695

The 3rd LAK InnovateDesign Workshop: Building a Triangle between Learning Design, Artificial Intelligence and Learning Management Systems

Organizers 1: Blaženka Divjak, Darko Grabar, Barbi Svetec, Petra Vondra
University of Zagreb, Faculty of Organization and Informatics, Croatia

Organizers 2: Dragan Gašević, Mladen Raković
Monash University, Faculty of Information Technology, Australia

Organizer 3: Bart Rienties
The Open University, Institute of Educational Technology, The United Kingdom

ABSTRACT: The **1st** LAK InnovateDesign workshop was dedicated to introducing an innovative concept of learning design (LD) and a complimentary tool for creating and analyzing LD. The **2nd** workshop focused on the evolving challenges intertwined with AI's role in LD. Building on the previous workshops, the objective for the **3rd** workshop is twofold: 1) 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 AI-supported LD. This encompasses harnessing the possibilities of combining design analytics with learning management system (LMS) data and finding meaningful ways to support that by AI. 2) To introduce participants to the latest developments of an innovative, free, AI-supported LD tool (learning-design.eu) and its capabilities, especially the scaffolding of courses in an LMS based on the LD prepared in the BDP tool. Participants will explore advanced LD analytics using this tool and be invited to collaboratively refine the LD of their courses, programs, or quality assurance endeavours, examining the LA data generated by the tool. They will be invited to exchange ideas and their LDs. This half-day, in-person workshop is a collaborative effort by three esteemed universities from Europe and Australia.

Keywords: learning design, learning analytics, assessment, learning management system, artificial intelligence

1 INTRODUCTION AND BACKGROUND

Learning design (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. It has been presented as a methodology helping teachers and designers in more informed decision-making related to the design of learning activities (Conole, 2013), 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.

To ensure that LD is pedagogically sound, it is essential to ensure constructive alignment (Biggs, 1996) between intended learning outcomes (LOs), teaching and learning activities (TLAs), and assessment (Divjak et al., 2024), and pay attention to the respective student workload. Achieving this can be strongly backed by learning analytics (LA) (Divjak et al., 2022, Divjak et al., 2023), which has been increasingly used to support LD (Rienties et al., 2017). Especially rich insights supporting the development and continuous improvement of LD can be provided by sophisticated AI-based LA using LMS data (e.g. to make predictions, as in Divjak et al., 2024), but there is also a significant potential of AI in providing assistance in the creative LD process and real-time feedback based on LD data. On the other hand, today there is also a great necessity to include AI-related LOs, activities and content into LD in a meaningful and sound way (Dai et al., 2023; Crompton & Burke, 2023). The aim of this workshop is therefore to discuss not only how to use AI as an element of LD, but also how to use it to support the creative process of LD.

Considering the recognised benefits of LD in enhancing teaching and learning in a digital age and supporting 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). However, the BDP also introduces a great amount of innovation, with continuous updates reflecting state-of-the art technological advancements, introduced in the design cycle process.

To start with, the BDP concept and tool enable linking course LOs with the study program LOs, providing an institutional perspective, which is valuable as 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 TLAs, assessment, feedback and student workload, supporting a student-directed approach (Divjak et al., 2024). 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. The analytics are provided in real-time, through a dedicated dashboard, and can be used as a valuable input directing the LD process. Furthermore, the tool enables collaborative work and sharing of LDs, as well as the export of LDs. Here, one of the latest and most advanced export functionalities enables the scaffolding of a course designed in the BDP tool automatically in the Moodle LMS, providing a high practical value in course preparation. Finally, the latest developments are related to exploring the possibilities of generative AI in providing real-time assistance in the LD process.

The BDP 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 1800 courses and MOOCs, by over 1800 users from more than 40 countries.

2 LEARNING OUTCOMES, WORKSHOP STRUCTURE AND WEBSITE

Based on the capacity-building at the workshop, participants will be 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, supported by AI, and (3) create LDs and scaffold courses in an open LMS. The half-day workshop, organized in cooperation of three universities, will be held face-to-face, consisting of the parts presented in the table below. The expected number of participants is between 15 and 30. Participation in previous workshops is not a prerequisite for this year's session.

Table 1. The proposed agenda of the workshop

Duration	Description	Responsible
10 min	INTRODUCTION	Organizer 1
	SHARING and DISCUSSION	Organizer 2
20 min	Principles of sound LD - presentation and discussion	Organizer 1, 2, 3
60 min	How can the BDP tool be used together with AI and courses automatically scaffolded in an open LMS - presentation Showcasing and interaction with LDs - interaction Finding interesting RQs - discussion	Organizer 1, 2 3
30 min	BREAK for tea and coffee	
	HANDS-ON COLLABORATION ON LEARNING DESIGN	
60 min	Designing with AI-supported BDP LD tool - work in groups	Organizer 1, 2 & 3
20 min	Presentation of LDs and discussion	Organizer 2
10 min	FUTURE STEPS AND CONCLUSIONS	Organizer 1, 2 & 3

The workshop will be supported by a dedicated website, where all related information will be shared, and which will support pre-workshop data gathering and planning, including the application of participants. To advertise the workshop to participants, we will use the workshop and SoLAR websites, and social media. After the workshop, the website and the social media will be used to support ongoing dissemination. The website will include the following sections: About, Background, Literature and Material, Workshop Agenda, and Submission Area.

3 SHARING OF EXPERIENCES, RESEARCH AND CHALLENGES

The workshop will start with a few short presentations by the workshop organizers, focusing on the current research, practices and experiences in the use of LD, complemented by interactive showcases of selected course LDs. A special focus will be on the principles of sound LD and how LA and LD analytics can support sound LD and how AI-related LD can be implemented. The workshop organizers will present how the BDP tool can be used hand in hand with AI and courses can be automatically scaffolded in an open LMS. Time will also be ensured for discussion of all participants, leading to open research questions and challenges, as well as presentation of other relevant tools supporting AI-enhanced LD.

4 HANDS-ON COLLABORATION ON LEARNING DESIGN

Ahead of the workshop, participants will be 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 will work collaboratively, grouped based on their own preferences and the similarity of courses/LOs they would like to work on.

The groups will be invited to design their courses and LOs using the BDP tool, with assistance provided by an AI chatbot. If applicable, the participants will scaffold their LDs automatically in the Moodle LMS. We also welcome contributions who want to use a different LD/AI/LMS tool as we are keen to learn from the diverse practices within the LA community.

Participants will work on the detailed planning of TLAs, assessment, feedback, modes of delivery, etc. In the process, they will consult the analyses provided by the tool, as well as suggestions provided by an AI chatbot, 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 will take approximately 1 hour and each group will be supported by one of the organizers. After the collaborative part, in the plenary session, groups will be invited to share their LDs and mutually discuss their outputs**.**

5 FUTURE STEPS AND CONCLUSIONS

Finally, the participants will be 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 will be shared with the participants after the workshop. There will be a possibility to establish further collaboration to work on a project and/or a publication. All participants will be 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

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347–364. https://doi.org/10.1007/BF00138871

Conole, G. (2013). *Designing for Learning in an Open World*. Springer New York.

Crompton, Burke,D. Artificial intelligence in higher education: the state of the field. Int J Educ Technol High Educ 20, 22 (2023). 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.*

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 Technolog*y, 54(1), 313–334. https://doi.org/10.1111/bjet.13290

Divjak, B., Svetec, B., & Horvat, D. (2024). How can valid and reliable automatic formative assessment predict the acquisition of learning outcomes? *Journal of Computer Assisted Learning*, 1–17. https://doi.org/10.1111/jcal.12953

Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y., Gasevic, D., & Chen, G. (2023, April 13). Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. https://doi.org/10.35542/osf.io/hcgzj

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.

The Second Workshop in New Measures & Metrics in Education

Charles Lang
Teachers College, Columbia University
charles.lang@tc.columbia.edu

Geraldine Gray
Technological University, Dublin
geraldine.gray@tudublin.ie

Ruth Cobos
Universidad Autónoma de Madrid
ruth.cobos@uam.es

Detra Price Dennis
Ohio State University
price.151@osu.edu

ABSTRACT: Learning analytics offers tremendous potential to improve educational outcomes, but new measures and metrics often remain isolated within institutions. Building on the success of the First Workshop on New Measures & Metrics LAK24, this half-day workshop aims to collaboratively advance the development and dissemination of innovative analytic measures in education. Similarly to last year, submissions of new metrics will be compiled on a website and presented at the event. A focus this year will be on demonstrating measures rather than only describing them though. 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

- Type of event: Workshop
- Proposed duration: Half day
- Workshop activities: Short presentations, affinity group-based discussion, competition
- Proposed schedule:
  - Introduction: Landscape, objectives, format
  - Open discussion about current issues with metric development
  - Mini-presentations of submitted new metrics
  - Vote on most promising new metric and prize award
  - Summary
- Participant numbers: 15-30
- Recruitment channels: Listservs, personal networks, student groups, social media, event advertisement.

2 WORKSHOP OBJECTIVES/INTENDED OOUTCOMES

2.1 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. This workshop will bring together researchers and practitioners to facilitate collaborative ideation, refinement and dissemination of new measures and metrics in learning analytics.

The adoption of novel learning analytics metrics faces substantial barriers within educational institutions. Many schools lack the financial resources, staff, infrastructure, and technical capabilities needed to implement new measures. 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. Privacy and ethical concerns surrounding data use further complicate adoption. Additionally, new metrics may misalign with established assessments and accreditation standards favored by administrators and faculty who tend to resist altering familiar practices. 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. The learning analytics community has an interest in addressing these concerns in a methodical and impactful way. One strategic approach is to surface, test and promote new measures for adoption – a role this workshop will fulfill.

2.2 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 metrics. By exchanging knowledge and experiences, participants can gain insights into challenges and strategies to operationalize and distribute metrics more effectively.

2.3 Key Format

The key activity will be the presentation and discussion of new measures that participants will submit 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. Submissions will include a description, a sample data set and a visualization. Election and wards for the most promising measures will be a key outcome of the event.

2.4 Dissemination

The workshop outcomes will be disseminated through multiple channels. A public website will compile promising metrics and serve as a reference for the community. The organizers also intend to synthesize insights and produce a review article on the state of the field. 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.

3 PLANNED MECHANISMS FOR COMMUNICATING INFORMATION & RESOURCES

- How will potential attendees find out about the intended content and structure of the event?
    - Participants will be sent website with slide deck of submitted metrics to be discussed
- How will you share information and resources with participants before, during and after the event?
    - Email and website
- Consider what tools will best meet your purpose, for example a website, shared google drive, mailing list, Slack workspace
    - Github and website

4 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.

From Data to Discovery: LLMs for Qualitative Analysis in Education

Author(s): Amanda Barany, Ryan S. Baker
University of Pennsylvania
amanda.barany@gmail.com
ryanshaunbaker@gmail.com

Author: Andrew Katz
Virginia Tech Engineering Education
akatz4@vt.edu

Author: Jionghao Lin
Hong Kong University
Carnegie Mellon University
Monash University
jionghal@andrew.cmu.edu

ABSTRACT: This workshop brings together researchers who have explored the use of large language models (LLMs) for the processing and analysis of qualitative data, in both the learning analytics community and other research communities. The workshop will feature presentations of research examples and demonstrations of these applications , providing insights into the methodologies and tools that have proven effective for automating qualitative analysis across research contexts. Additionally, the session will address challenges such as data privacy, ethical considerations, and ways to build community and shared resources. Attendees will share their experiences and contribute to a collective understanding of best practices in the use of AI for qualitative research. Participants will engage in discussions and hands-on activities to understand the capabilities and limitations of LLMs in handling qualitative data. An output of the workshop will include a plan for developing a systematic review of progress in using LLMs for qualitative data analysis.

Keywords: Large Language Models, Qualitative Data Analysis, AI in Education, Natural Language Processing

1 BACKGROUND

Although researchers have used classical Natural Language Processing (NLP) methods for many years to support qualitative research, a significant shift has occurred with the advent of Large Language Models (LLMs), in terms of quality and feasibility (e.g., Zambrano et al., 2023; Chew et al., 2023; Sinha et al., 2024; Kirsten et al., 2024). This has resulted in an increase in recent publications on the application of LLMs and AI for qualitative analysis, including discussions of the ethical, technical, and logistical issues and best practices (Törnberg, 2023; Xiao et al., 2023; Lopez-Fierro & Nguyen, 2024; Meng et al., 2024). Given this diverse and widespread use, a consolidated view of existing work and practical guidance on the application of AI in qualitative research is needed to more clearly map both the current state of the field and next steps.

We aim to bridge this gap by organizing a workshop that will bring together leaders in the field to discuss cutting-edge research on current applications of LLMs and AI for qualitative data analysis. Through short talks, facilitated breakout sessions, and collaborative discussions, the workshop will discuss key issues, explore best practices, outline future research directions, and create a plan for developing a systematic review of progress in using LLMs for qualitative analysis. By fostering dialogue among scholars and practitioners, the workshop will contribute to advancing the use of AI in qualitative research, ultimately enhancing the field's capacity to leverage these tools effectively and ethically.

2 WORKSHOP AIMS AND STRUCTURE

2.1 Aims and Contributions

The aim of this workshop is to expand the field’s knowledge regarding the current and ongoing applications of Large Language Models (LLMs) and AI techniques for qualitative data analysis by bringing together a diverse group of leaders conducting research in this area. Workshop organizers will solicit and coordinate submissions on the topic which will shared during the workshop in presentations and poster lightning talks. The aims of the workshop are as follows:

1. Learn about emerging work on the use of LLMs for qualitative data analysis within and related to the field of learning analytics.
2. An exploration of the alignment between specific uses of LLMs, and different qualitative research methodologies, framed around applied research in diverse educational contexts.
3. Discussion of key issues and next steps, encompassing practical, methodological, and ethical aspects, for the future advancement of qualitative data analysis using AI.
4. Create a plan for developing a systematic review of progress in using LLMs for qualitative analysis.

We anticipate that the following benefits and contributions will emerge from the workshop discussion and subsequent research outputs:

1. A collaborative, accessible overview of existing research on LLMs in qualitative data analysis for both presenters and attendees.
2. Dialogue between scholars around the potential applications of LLM and AI techniques to qualitative research from different theoretical approaches and applied contexts.
3. Sharing of study designs and analysis procedures such as:
    - Preparing qualitative data for analysis using LLMs.
    - The development of codes and codebooks for qualitative analysis.
    - The application of codes (coding).
4. Techniques for ensuring validity and reliability of analysis procedures.
5. A discussion of best practices for AI-driven qualitative research, such as:
    - Ensuring interpretability, validity, and reliability in research outputs.
    - Addressing ethical considerations and data privacy concerns.
    - Supporting fairness and transparency in research practices.
6. A discussion of affordances and constraints of LLMs and AI for qualitative research, such as:
    - Scalability and efficiency versus potential for error and oversimplification.

- Automation of repetitive tasks versus the need for continuous human oversight.
- Cross-language analysis versus limitations in handling linguistic diversity.

A roadmap for areas of emergent research, methodological innovations, possible future collaborations, and notes generated from workshop discussions will be created and shared with the broader learning analytics community, along with a plan for building forward to a systematic review.

2.2 Proposed Workshop Structure

The workshop will be structured as follows:

1. **Introduction:** A brief overview of workshop goals, setting the stage for the day's discussions.
2. **Short Talks:** Participating scholars will present a 10-minute talk on:
   - The context and populations of their qualitative research using LLMs and AI.
   - Theoretical frameworks and alignment with AI techniques.
   - Specific tools and techniques used, their affordances and constraints.
   - Proposed next steps and future research directions.
3. **Poster Session and Lightning Talks**
   - Five-minute presentations focusing on contributions of their work, proposed next steps, and how their topic aligns with major questions for the field.
   - Poster Review and Discussion: Non-presenting attendees share feedback and ideas for ongoing work.
4. **Discussion and Brainstorming (60 minutes)**
   - Fifteen-minute sessions: Small group discussions focusing on what is known, current gaps, and major questions for the field.
   - Group Sharing: Each group will share their insights with all participants, synthesizing key themes and ideas.
5. **Facilitated Breakout Sessions:** In-depth small group sessions led by organizers and panelists addressing practical, methodological, and theoretical next steps for AI-driven qualitative research. Topics will include best practices, emerging technologies, and contextual and ethical considerations.
6. **Collaboration Planning and Community Building:** Rapid, structured meetings where participants have brief one-on-one or small group interactions to discuss potential collaborations, share ideas, and build connections within the research community.
7. **Wrap-Up and Review Paper Discussion:** An open forum for final questions, ideas, and contributions to a review paper on the area.

This structure is intended to support a comprehensive exploration of current research, collaborative brainstorming, and networking opportunities for participants.

3 CONCLUSIONS

By combining presentations, collaborative sessions, and networking opportunities, we seek to enhance the community’s collective understanding of LLMs' capabilities and limitations, foster innovative research practices, and establish steps forward for advancing AI-driven qualitative analysis in education. The workshop's outputs, including slides and recordings from presentations, notes, and artifacts from discussions and community-building, will be made available to workshop attendees and the broader community at the conclusion of the workshop. In addition, interested attendees and presenters will be invited to participate in the development of a systematic review and position paper on the use of LLMs for qualitative research. Our hope is that this work will contribute valuable insights and guidance for the ongoing development and ethical application of AI technologies in research.

4 PROGRAM COMMITTEE

Anthony Botelho, abotelho.777@gmail.com, The University of Florida College of Education
Ashish Gurung, agurung@andrew.cmu.edu, Carnegie Mellon University
Yoon Jeon (YJ) Kim, yj.kim@wisc.edu, University of Wisconsin-Madison
Ayaz Karimov, ayaz.karimov@jyu.fi, University of Jyväskylä
Yi-Chieh (EJ) Lee, yclee@nus.edu.sg, National University of Singapore
Ha Nguyen, ha.nguyen@unc.edu, University of North Carolina at Chapel Hill
Jeanne McClure, jmmcclu3@ncsu.edu, North Carolina State University
Jaclyn Ocumpaugh, ojaclyn@upenn.edu, University of Pennsylvania
Luc Paquette, luc.paquette@gmail.com, University of Illinois
Saríah Lopez-Fierro, sariah.lopezfierro@usu.edu, Utah State University

REFERENCES

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv preprint*, arXiv:2306.14924. Retrieved from https://arxiv.org/abs/2306.14924

Kirsten, E., Buckmann, A., Mhaidli, A., & Becker, S. (2024). Decoding complexity: Exploring human-AI concordance in qualitative coding. *arXiv preprint*, arXiv:2403.06607. Retrieved from https://arxiv.org/abs/2403.06607

Lopez-Fierro, S., & Nguyen, H. (2024). Making human-AI contributions transparent in qualitative coding. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning* (CSCL 2024, pp. 3-10). International Society of the Learning Sciences.

Meng, H., Yang, Y., Li, Y., Lee, J., & Lee, Y. C. (2024). Exploring the potential of human-LLM synergy in advancing qualitative analysis: A case study on mental-illness stigma. *arXiv preprint*, arXiv:2405.05758. Retrieved from https://arxiv.org/abs/2405.05758

Sinha, R., Solola, I., Nguyen, H., Swanson, H., & Lawrence, L. (2024, June). The role of generative AI in qualitative research: GPT-4's contributions to a grounded theory analysis. In *Proceedings of the Symposium on Learning, Design and Technology* (pp. 17-25).

Törnberg, P. (2023). ChatGPT-4 outperforms experts and crowd workers in annotating political Twitter messages with zero-shot learning. *arXiv preprint*, arXiv:2304.06588. Retrieved from https://arxiv.org/abs/2304.06588

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. Y. (2023, March). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 75-78). https://doi.org/10.1145/3581754.3584136

Zambrano, A. F., Liu, X., Barany, A., Baker, R. S., Kim, J., & Nasiar, N. (2023, October). From nCoder to ChatGPT: From automated coding to refining human coding. In *International Conference on Quantitative Ethnography* (pp. 470-485). Cham: Springer Nature Switzerland.

LAK25 Assess:

The 5th Workshop on Learning Analytics and Assessment

Dragan Gašević and Mladen Raković
Centre for Learning Analytics, Monash University
{dragan.gasevic, mladen.rakovic}@monash.edu
Blaženka Divjak
Faculty of Organization and Informatics, University of Zagreb
bdivjak@foi.hr
Yoon Jeon Kim
Department of Curriculum and Instruction, University of Wisconsin-Madison
yj.kim@wisc.edu
Abhinava Barthakur
Education Futures, University of South Australia
abhinava.barthakur@unisa.edu.au
Lukas Liu
Faculty of Education, Hong Kong University
lukasliu@hku.hk

ABSTRACT: The first four editions of the Workshop on Learning Analytics and Assessment were successfully organized at LAK21-24 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. Many 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., Baker, Lindrum, Lindrum, & Perkowski, 2015; Gardner & Brooks, 2018; Greene et al., 2019), analysis of learning strategies and 21st century skills (e.g., Srivastava et al., 2022), adaptive learner support and personalized feedback at scale (e.g., 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. 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). 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 four editions of the Workshop on Learning Analytics and Assessment were successfully organized at LAK21-LAK24 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 fifth 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 investigate 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 Half-Day Workshop Schedule

Table 1: Proposed schedule.

Timing	Description	Contributors
10 minutes	Welcome, introductions and plan for today	Organizers
20 minutes	Learning analytics for assessment of self-regulated learning 7 minutes per presentation + 3 minutes for Q&A per presentations	Presenters 1 & 2
25 minutes	Assessment of self-regulated learning - roundtable	Participants
20 minutes	Analytics for formative assessment of reading and writing	Presenters 3 & 4
25 minutes	Formative assessment of reading and writing - roundtable	Participants
30 minutes	Morning Tea	
30 minutes	Assessment and generative AI 20 minutes for presentation + 10 minutes for Q&A	Keynote
20 minutes	AI in assessment 7 minutes per presentation + 3 minutes for Q&A per presentations	Presenters 5 & 6
25 minutes	AI in assessment - roundtable	Participants
5 minutes	Next steps and close	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 1 October 2024, deadline for abstract submissions 4 Dec 2024, and notification of acceptance 20 Dec 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

Baker, R. S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education E-Learning Courses. International Educational Data Mining Society (pp. 150-155)

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. User Modeling and User-Adapted Interaction, 28(2), 127-203.

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., ... & Panter, A. T. (2019). Modeling temporal self-regulatory processing in a higher education biology course. Learning and Instruction, 101201.

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.

Srivastava, N., Fan, Y., Rakovic, M., Singh, S., Jovanovic, J., Van Der Graaf, J., ... & Gasevic, D. (2022). Effects of internal and external conditions on strategies of self-regulated learning: A learning analytics study. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 392-403).

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.

Equity-centered R&D: Why we need it and how to do it

Britte Haugan Cheng
Menlo Education Research
bcheng@menloedu.org

Caitlin Mills
University of Minnesota
cmills@umn.edu

ABSTRACT: This workshop focused on equity-centered research and development (R&D); specifically, we brought together researchers to talk about how the field of learning analytics (LA) was well-positioned to make significant contributions to the larger educational research and development (R&D) space by taking equity seriously in their efforts. Equitable R&D had not been a huge focus in the learning analytics community to date, but we discussed ways that LA researchers, regardless of their primary research focus (i.e. one did not have to study equity to deeply engage with equitable practices), could take actionable steps to ensure that their research integrated such R&D methods and processes.

Keywords: equity-centered R&D, inclusive R&D

1 WORKSHOP BACKGROUND

This workshop considered how the learning analytics (LA) community is well-positioned to integrate research questions, methods, and processes that centered equity in order to make significant contributions to the larger educational research and development (R&D) space. Given that equitable R&D has not, historically, been a focus in the learning analytics community, our workshop focused on how LA researchers, regardless of their primary research questions, can take steps to ensure that their work integrated equitable R&D methods and processes.

1.1 Research and Development (R&D) Processes

One of the core aspects of the LA community was designing new tools for teachers and students – often as part of research and development (R&D) cycles, where the ultimate goal is to create new tools (i.e. dashboards), interventions, or improvements. The research and development processes featured at LAK, however, tend to only involve a relatively narrow set of voices in this process overall. This could mean that students and teachers are often not deeply woven into the R&D process, or their various identities are not considered. Although there were some exceptions to this – especially in very recent years [1], [2], [3], [4], this has not historically been the case: for example, Williamson and Kizilcec [5] found that diversity, equity, and inclusion were rarely considered in the designs of an LA dashboard (i.e. considering demographics of the dashboard was designed for, etc.)

– let alone be included as participants in the design. Further, studies do not typically consider, for example, representation of gender identity, racial/ethnic identity, socio-economic background, or other ways that samples are representative of a larger population unless the research question(s) require attention to these dimensions.

Beyond considering representation of participants in data sets, the majority of studies in learning analytics do not raise questions of access (opportunity to engage or learn), other systemic/infrastructure factors, or situational power dynamics that may impact students' experiences and behaviors when considering achievement outcomes. Moreover, while some studies consider the learning setting as a mediator of student learning processes and outcomes (e.g., via the lens of resources; [6]), few have examined the cultural aspects of the learning context as key design considerations. Implications of these frequent omissions as part of field-wide theoretical advancement or for the development of educational products risked a continual systematic exclusion of historically marginalized students and participants in the contributions of the learning analytics community.

1.2 Equity-centered R&D

Although it is important for the learning analytics community to center equity in the formation of research questions across the field, employing methods that are designed with equity in mind is a field-wide shift that can increase the relevance of the community's collective work. As part of her work in mathematics education, Gutierrez's framework defines 4 dimensions of equity that researchers can consider in designing studies [7]; they include *access* (the tangible resources that students have available to them to participate in [learning]), *achievement* (how student outcomes are defined and whether/why there are systemic gaps in those outcomes), *identity* (students' personal, cultural or linguistic capacities), and *power* (who has 'voice', whose knowledge counts, etc.). Haynes and colleagues offer guidance to researchers to consider identity from a more complex perspective, one of intersectionality, as a means to reflect on unexamined biases in our methods [8]. Among other approaches, they call for researchers to use a critical lens to uncover micro- and macro-level power relations and to explicitly address how those power relationships shaped their methods. They ask researchers to critique their own positionality and biases and to explicitly name the strategies used to disrupt the ways that power has shaped the research study methods.

1.3 Why now?

Taking on equity as part of our methods does not require refocusing our research questions or professional trajectories whole cloth, but it does require that we attend to the impacts of our methodological choices as part of the larger R&D enterprise. Our workshop proposed to critically engage in discussions about the ways in which LA researchers can bring these perspectives into their practices. This included creating a space to learn from the researchers within LA who were doing this well, as well as creating time and structure for others to find ways to incorporate these practices to continue advancing the field in equitable ways. Emerging scholars are eager to do this work, and our workshop invited folks in to learn, create community, and find aligning frameworks to guide future endeavors in this space.

2 ORGANIZATION OF PROPOSED WORKSHOP

This half-day workshop included three main components. First, we had a workshop opening presentation, which presented the goals of the day, and discussed background on inclusive R&D. This also included small groups where people shared where they were in their work and how they would like to center equity moving forward. They stayed with these groups throughout the workshop.

The second component took on a mini-conference format, where a limited number of papers were presented along with discussion. For this section, we invited research teams to present on: 1) completed projects or work-in-progress that engaged in inclusive R&D or 2) opinion/commentaries that reflected on these topics and grappled with open questions about learning analytics' relationship to equity-centered R&D. The final section of the workshop involved guided small group brainstorming about how folks would like to see equity show up in their research process and any gaps they viewed in their work that they would like to fill.

These themes were converted to questions, for a wrap-up discussion with the larger group. After the workshop, we invited participants to continue the conversation on this topic by creating a special issue at an interdisciplinary journal, summarizing workshop findings, and discussing how others engaged in equity-centered R&D. The workshop had open participation, with open registration for anyone interested.

1.1 Workshop Format and Activities

The schedule was as follows:

- 09:00am - 10:00am Workshop opening + small groups
- 10:00am - 11:00am Paper presentations + discussion
- 11:00am - 11:15am Coffee Break
- 11:15am - 12:15pm Interactive activity: guided small group brainstorming (incorporating equity focus/methods in your work)
- 12:15am - 12:45pm Open Discussion: Identifying guiding questions + resource sharing
- 12:45pm - 01:00pm Workshop closing

1.2 Workshop Objectives and Intended Outcomes

The key outcomes of the workshop were:

1. To identify brightspots and barriers to equity centered R&D within the LAK community
2. Identify ways to bring an equity-centered lens to any research topic or domain

3. Identify resources and opportunities to support equity-centered R&D (how to find partners, R&D networks, funding opportunities, tools and resources for research (SEERNEt and AIMS RDI)

4. Develop plans to formalize findings from the workshop and ways to share resources that result from the workshop

REFERENCES

[1] M. Khalil, P. Prinsloo, and S. Slade, "Fairness, trust, transparency, equity, and responsibility in learning analytics," J. Learn. Anal., vol. 10, no. 1, pp. 1–7, 2023.

[2] A. Grimm, A. Steegh, M. Kubsch, and K. Neumann, "Learning Analytics in Physics Education: Equity-Focused Decision-Making Lacks Guidance!," J. Learn. Anal., vol. 10, no. 1, pp. 71–84, 2023.

[3] R. Heiser, M. E. D. Stritto, A. Brown, and B. Croft, "Amplifying student and administrator perspectives on equity and bias in learning analytics: Alone together in higher education," J. Learn. Anal., vol. 10, no. 1, pp. 8–23, 2023.

[4] J. Anderson and M. Devlin, "Data analytics in adaptive learning for equitable outcomes," in Data Analytics and Adaptive Learning, Routledge, 2023, pp. 170–188. Accessed: Sep. 07, 2024. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003244271-13/data-analytics-adapti ve-learning-equitable-outcomes-jeremy-anderson-maura-devlin

[5] K. Williamson and R. F. Kizilcec, "Learning Analytics Dashboard Research Has Neglected Diversity, Equity and Inclusion," in Proceedings of the Eighth ACM Conference on Learning@ Scale, 2021, pp. 287–290.

[6] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," Internet High. Educ., vol. 28, pp. 68–84, Jan. 2016, doi: 10.1016/j.iheduc.2015.10.002.

[7] R. Gutiérrez, "Context Matters: How Should We Conceptualize Equity in Mathematics Education?," in Equity in Discourse for Mathematics Education, B. Herbel-Eisenmann, J. Choppin, D. Wagner, and D. Pimm, Eds., Dordrecht: Springer Netherlands, 2012, pp. 17–33. doi: 10.1007/978-94-007-2813-4_2.

[8] C. Haynes, N. M. Joseph, L. D. Patton, S. Stewart, and E. L. Allen, "Toward an Understanding of Intersectionality Methodology: A 30-Year Literature Synthesis of Black Women's Experiences in Higher Education," Rev. Educ. Res., vol. 90, no. 6, pp. 751–787, Dec. 2020, doi: 10.3102/0034654320946822.

Expansive Actionable Learning Analytics for Schools: Challenges and Opportunities

Facilitator: JohnPaul Kennedy
University of South Australia
John.Kennedy@unisa.edu.au

ABSTRACT: Actionable learning analytics have shown promise in improving educational outcomes, particularly in higher education, but their application to the K-12 context has been somewhat inconsistent. While some educators and researchers see benefits, others are concerned about data misuse and privacy issues. The growing volume of student data available in school systems presents both opportunities and challenges, with concerns about ethics, consent, and potential harm balancing against the possibilities of individualized educational support and tailored learning experiences. This half-day collaborative workshop will engage researchers and educators with an interest in expanding the use of actionable learning analytics in the K-12 context in structured discussions. Through the use of roundtable and rapid Delphi protocols the workshop participants will identify three research themes for prioritization and the advancement of the field.

Keywords: Actionable Learning Analytics, K-12 Educational Change, Collaborative research analysis, Research priorities.

1 OVERVIEW

The use of actionable learning analytics (ALA) to improve educational outcomes is now an established field although its use has been more prevalent in higher education contexts than in schools for both pragmatic and practical reasons. Within the K-12 context ALA has had a mixed response with researchers (e.g., Catherine et al., 2024; Laura et al., 2022) reporting some educators being open to the educational benefits of ALA while others remain unconvinced and/or concerned by potentially harmful impacts such as data misinterpretation or privacy infringement. Bond et al. (2023) identified that while ALA has strong potential to increase students' engagement in their learning there is little compelling evidence to suggest that this has occurred in practice. Perhaps this dissonance between potentiality and actuality is due to differences in pedagogical aims or philosophical incongruences between researchers and educators or perhaps there is something more fundamentally different about K-12 education.

In the last decade, the volume and breadth of student data available to educators and researchers in the K-12 school context has grown to such an extent that it has become both a treasure trove and an ethical minefield. While the ethics of ALA have always been an area for active research (e.g., Abelardo et al., 2014; Sharon et al., 2013) concerns around the use of data in schools, where students cannot generally provide legally binding informed consent, are often amplified. Combining this with the potential risk for harm (Andrea & Andrea, 2020; Willis et al., 2016) leads to a situation where schools are not always willing or able to share data with researchers in a meaningful way. Nevertheless, if researchers are to bring the ALA field to its potential, there needs to be a shift towards understanding learning environments, not as a collection of independent processes, but as a complex, integrated and holistic system (see Dawson et al., 2019).

This workshop aimed to explore some of the challenges and opportunities for ALA that the relative uniqueness of individual school contexts, cultures, and values bring to the fore. By bringing together researchers, educators, and other contributors with experience of this sector, this collaborative workshop aims to establish three research themes that the participants believe need to be prioritized by the research community in order to expand the impact of ALA within the global K-12 community.

2 WORKSHOP STRUCTURE

This half-day, collaborative workshop (Table 1) was designed to guide up to 40 participants through a series of structured roundtable discussions to explore some of the challenges that are unique to this learning context and to delve into possible approaches to solutions. All participants with a research or implementation interest in LA in schools, regardless of level of experience, were encouraged to attend and share their observations, insights, and perspectives.

Table 1: Workshop Structure.

	Protocol and Prompting Question
Session 1	Round table 1 Prompting Question: *What are the challenges we face when implementing ALA in K-12 settings?*
	Whole group synthesis and discussion
Session 2	Round table 2 Prompting Question: *What approaches or solutions might we adopt to address these challenges?*
	Whole group synthesis and discussion followed by Rapid Delphi stage 1a
Session 3	Rapid Delphi Process to discern by consensus *What are three highest priority research areas for expanding ALA in the K-12 context?*

Participants were be divided randomly to form roundtable discussion groups of around 8 participants. Working from a simple prompting question (Table 1), each roundtable will follow a protocol as outlined in Table 2. While many formal and informal protocols exist for facilitating an effective roundtable discussion of peers, a process similar to that by Devis et al. (2023) will be used to ensure that each participant gets an equal voice in the conversation and that no one idea or agenda can dominate the discussion. At the end of each roundtable insights and key ideas from each group will be shared more broadly with the workshop group and discussed in a search for aspects of harmony and dissonance within the larger group of experts. The roundtable groups will be shuffled between rounds of activities to ensure that as many different perspectives as possible are captured through the discussions.

Table 2: Roundtable Protocol.

Timing	Activity
8 × 2 minutes	Each participant gives their initial response to the prompting question in turn without interruption or comment by others.
8 × 3 minutes	Each participant gives continues the discussion by building on either their own initial statement, responding to others, highlighting what they perceive to be commonalities or differences, contributing additional personal experience, or similar.
15 minutes	Group prepares summary of key points from their discussion to share with the workshop group as a whole.

In the third and final session, a rapid protocol (Table 3) based on the Delphi method (Dalkey, 1969) will be adopted to discern, rank, and gain consensus on the participants' views around the three most pressing research themes in need of investigation in order to expand the utility of ALA in the K-12 context. This Rapid-Delphi technique is implemented using a custom web application[1] that will be freely available to use after the workshop.

Table 3: Rapid Delphi Protocol.

Stage	Timing	Activity
1a	5 minutes	Each participant writes and shares what they consider to be the most important 1 or 2 research themes that need to be addressed. Research themes are written into a digital tool.
1b	10 minutes	The facilitator shares all research themes on the screen and invites questions from the participants to enhance or clarify any statements. Statements are updated live on the screen as the discussion proceeds
2	5 minutes	Participants select what they perceive to be the most important 5 research themes and rank them in importance using the digital tool.
3	20 minutes	The amalgamated ranked list of themes is shared with participants and an open discussion is facilitated. The order of themes may be adjusted by the group through consensus. It is important to note that consensus does not equate to a simple majority vote; rather, it indicates a position where those who dissent agree that they have been fully heard.
4	5 minutes	Agreement is reached over the final three research priorities to be shared with the LAK community.

[1] https://github.com/DrJPK/rapid-delphi

3 DISSEMINATING FINDINGS

The findings of the workshop will be available at the workshop website (https://efa.unisa.edu.au/K12LA) after the conference and will be published as a journal article during 2025. A non-identifiable summary of data produced during the workshop will also be made available at this site under a CC BY-NC license to assist other researchers in identifying potentially fruitful themes for future research.

The ethical aspects of this workshop have been considered by the Human Research Ethics Committee (206788) of the University of South Australia as required by the Australian government research requirements, specified in the National Statement on Ethical Conduct in Human Research (2023).

REFERENCES

Abelardo, P., Abelardo, P., George, S., & George, S. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12152

Andrea, L. B., & Andrea, L. B. (2020). The use of learning analytics and the potential risk of harm for K-12 students participating in digital learning environments. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-020-09854-6

Bond, M., Viberg, O., & Bergdahl, N. (2023). The current state of using learning analytics to measure and support K-12 student engagement: A scoping review. *International Conference on Learning Analytics and Knowledge*. https://doi.org/10.1145/3576050.3576085

Catherine, P., Sam, V., Richard, T. B., II, Catherine, C., Christine, S., & Zandra De, A. (2024). A review of learning analytics opportunities and challenges for K-12 education. *Heliyon*. https://doi.org/10.1016/j.heliyon.2024.e25767

Dalkey, N. (1969). An experimental study of group opinion: The Delphi method [Article]. *Futures*, *1*(5), 408-426. https://doi.org/10.1016/S0016-3287(69)80025-X

Dawson, S., Laat, M. d., Joksimović, S., Poquet, O., & Siemens, G. (2019). Increasing the Impact of Learning Analytics. *International Conference on Learning Analytics and Knowledge*. https://doi.org/10.1145/3303772.3303784

Devis, D., Fowler, S., Vieira, M., Giannoni, K., Gabriel, F., Kennedy, J., & Leonard, S. (2023). *From insight to action: strategies for cultivating equity and empowering women in industry*. https://www.unisa.edu.au/siteassets/academic-units/unisa-education-futures/docs/report-from-insight-to-action-strategies-for-cultivating-equity-and-empowering-women-in-industry.pdf

Laura, H., Mohammed, S., Sonsoles, L.-P., & Teemu, V. (2022). A systematic narrative review of learning analytics research in K-12 and schools. *FLAIEC*. https://www.semanticscholar.org/paper/48451dd8acb31a537c8b3ec11da894ff464e2386

Sharon, S., Sharon, S., Paul, P., & Paul, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*. https://doi.org/10.1177/0002764213479366

Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective [Article]. *Educational Technology Research and Development*, *64*(5), 881-901. https://doi.org/10.1007/s11423-016-9463-4

Game-Based Learning Analytics

Maurice Boothe Jr., Xavier Ochoa
New York University
mab1488@nyu.edu, xao1@nyu.edu

David Gagnon, Luke Swanson
University of Wisconsin - Madison
djgagnon@wisc.edu, swansonl@cs.wisc.edu

Erik Harpstead
Carnegie Mellon University
harpstead@cmu.edu

ABSTRACT: Game-Based Learning Analytics (GBLA) is an emerging field that combines game design principles with learning analytics to create personalized, engaging, and data-driven educational experiences. Despite significant contributions in this space, there is an absence of structured collaboration within the learning analytics community. This workshop aims to address that gap by bringing together researchers and practitioners interested in the intersection of games and learning analytics. Through this half-day workshop, we aim to formalize a community around GBLA, set foundational principles, and coordinate initial scholarly contributions while laying the groundwork to formalize a community of practice.

Keywords: game-based learning, learning analytics, special interest group

1 WORKSHOP BACKGROUND

Games are a rich medium with potential to facilitate and enhance students' learning (Plass et al., 2013). Researchers in game design have coined the term "serious" or "educational" games for these games that are intentionally built to help students acquire or practice a given knowledge or skill (Loh, 2015). The development of these types of games as learning environments has evolved into an active area of research known as Game-Based Learning or GBL (Plass et al., 2020). GBL has shown that serious games utilize successful game design principles to generate personalized, engaging, and deep learning experiences (Gee, 2006) and may help students achieve better learning outcomes compared to conventional ways of learning (Mayer, 2014; Mayer, 2019).

There have been numerous games-related contributions in the field of learning analytics. In the most recent International Learning Analytics and Knowledge Conference (LAK), there were three accepted research papers directly related to learning analytics in game contexts (Mangaroska et al., 2024; Vanacore et al., 2024; Wang et al., 2024) as well as four more accepted contributions in the corresponding companion proceedings (Gamper & Schroeder, 2024; Hlosta et al., 2024; Li et al., 2024; Poquet et al., 2024). This combined with a special issue in the Journal of Learning Analytics titled "Analytics for Game Based Learning" (Kim et al., 2022) indicates a significant presence of contributions that combine learning analytics with game-based learning. Despite these contributions, there remains a lack of collaboration related to games in the learning analytics community.

There is an opportunity for the founding of a special interest group within SoLAR to facilitate connections and collaboration on the unique challenges that emerge when applying learning analytics to educational game contexts such as parsing out learning mechanics data from game mechanics data, navigating the granularity and density of game log data, and the ethics of collecting protected information from playful experiences. This prospective group will collaborate with other related groups such as the Open Game Data Initiative on projects that facilitate engaging with and analyzing the data from these types of games (eg. Gagnon et al., 2019; Gagnon et al., 2022; Liu et al., 2023; Scianna and Kim, 2024; Swanson et al., 2022) as well as address the design of dashboards, feedback systems, pedagogical agents, narratives and more as it relates to the capture and analysis of learning analytics in games. There is also an opportunity to consider the different stakeholders that participate in the design process of game-based learning analytics (Boothe et al., 2025).

2 WORKSHOP DETAILS

We propose a half-day workshop for up to 20 participants who have a connection or interest in the design, implementation, or analysis of learning analytics in educational games. The purpose of this workshop is to invite these members "into the room" in order to discuss relevant issues, principles, and priorities with the overarching topic for the workshop: Game-Based Learning Analytics or GBLA. An immediate result of this workshop is the establishment of priorities, coordination of potential contributors, and a plan of action for initial and forthcoming contributions.

2.1 Event Type & Structure

This half-day workshop will consist of invited speakers, informal “micro-talks”, and a series of participant-driven discussion activities.

2.2 Schedule & Activities

2.2.1 Keynote Speaker: 45 minutes

The workshop will begin with a keynote address to provide foundational aspects of game-based learning and connections to learning analytics. The keynote will have made contributions to game-based learning literature and ideally have implemented learning analytics into one or more projects. Prospective speakers include Ryan Baker of Carnegie Mellon University, YJ Kim and V. Elizabeth Owen of University of Wisconsin - Madison, and Elizabeth Rowe of Technical Education Resource Center.

2.2.2 Community-Building: 1 hour

Participants will be invited to participate in a “microtalk” in which they briefly share their prior work related to GBLA to establish a social presence and demonstrate expertise across the group. As part of this activity, abstracts on relevant topics will be solicited. Accepted abstracts will have an opportunity to expand into a short paper for submission. Information related to this activity will be communicated to participants in advance of the workshop.

2.2.3 Establishing Principles of Game-Based Learning Analytics: 45 minutes

Participants will be invited to identify factors relevant to GBLA, exploring aspects that emerge from the two originating fields. Through a series of collaborative activities, this brainstorm will be compiled and synthesized into a series of initial principles that will guide future considerations within GBLA. Through these group discussions and guided activities, the workshop will also establish a “working definition” of Game-Based Learning Analytics.

2.2.4 *Identifying and Prioritizing Issues in Game-Based Learning Analytics: 1 hour*

Having established a series of principles for GBLA, participants will be invited to identify issues or challenges that emerge from the space. These will be collected and prioritized for consideration of future scholarly contributions.

2.2.5 *Reflection and Coordination of Scholarly Contribution: 30 minutes*

We will invite feedback on the format, content, and activities at the conclusion of the workshop as well as coordinate a post-conference session to discuss topics for a potential scholarly contribution.

2.3 Recruitment & Dissemination

As part of organizing this community of practice, several platforms will be leveraged for engagement including the creation of a new Google Site as a central location for disseminating details about the event. Recruitment messages will also be sent to an email list maintained by the organizers as well as to existing online communities (including Open Game Data, IDGA) and email groups (such as Learning Engineering, CHI Play, and Educational Data Mining).

3 INTENDED OUTCOMES

This workshop seeks to: 1) formalize a community of practice surrounding Game-Based Learning Analytics and its needed infrastructure, 2) establish an initial set of active projects, principles, issues, and priorities for the community, 3) validate the need for this community as part of an application to SoLAR as a special interest group, and 4) coordinate an initial scholarly contribution.

REFERENCES

Boothe, M., Gopalakrishnan, M., Huynh, M., Wang, Y., Ochoa, X. (2025). Game-Based Learning Analytics: Insights from an Integrated Design Process. In: Plass, J.L., Ochoa, X. (eds) Serious Games. JCSG 2024. Lecture Notes in Computer Science, vol 15259. Springer, Cham. https://doi-org.proxy.library.nyu.edu/10.1007/978-3-031-74138-8_9

Gagnon, D. J., Baker, R. S., Gagnon, S., Swanson, L., Spevacek, N., Andres, J., Harpstead, E., Scianna, J., Slater, S., & San Pedro, M. O. C. Z. (2022, September 18). Exploring players' experience of humor and snark in a grade 3-6 history practices game. https://escholarship.org/uc/item/80r6r2jp

Gagnon, D. J., Harpstead, E., & Slater, S. (2019). Comparison of Off the Shelf Data Mining Methodologies in Educational Game Analytics. EDM (Workshops), 38–43. https://ceur-ws.org/Vol-2592/paper5.pdf

Gamper, P., & Schroeder, U. (2024). Exploring auto generated solution spaces of a serious game for introductory programing courses. Companion Proceedings 14th International Conference on Learning Analytics & Knowledge, 281–283. https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24_CompanionProceedings.pdf

Gee, J. P. (2006). Are video games good for learning? Nordic Journal of Digital Literacy, 1(3), 172–183.

Hlosta, M., Moser, I., Winer, A., Geri, N., Ramnarain, U., & Westhuizen, C. V. der. (2024). Learning Analytics from Virtual Reality (LAVR). Companion Proceedings 14th International Conference on Learning Analytics & Knowledge, 382–385. https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24_CompanionProceedings.pdf

Kim, Y. J., Valiente, J. A. R., Ifenthaler, D., Harpstead, E., & Rowe, E. (2022). Analytics for Game-Based Learning. Journal of Learning Analytics, 9(3), Article 3. https://doi.org/10.18608/jla.2022.7929

Li, L., Feng, M., & Bang, H. J. (2024). A Data-Centric Personalized Learning Technology Solution to Accelerate Early Math Skills of Young Learners. Companion Proceedings 14th International Conference on Learning Analytics & Knowledge, 55–57. https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24_CompanionProceedings.pdf

Liu, X., Slater, S., Andres, J. Ma. A. L., Swanson, L., Scianna, J., Gagnon, D., & Baker, R. S. (2023). Struggling to Detect Struggle in Students Playing a Science Exploration Game. Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play, 83–88. https://doi.org/10.1145/3573382.3616080

Loh, C. S., Sheng, Y., & Ifenthaler, D. (Eds.). (2015). Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement. Springer International Publishing. http://link.springer.com/10.1007/978-3-319-05834-4

Mangaroska, K., Larssen, K., Amundsen, A., Vesin, B., & Giannakos, M. (2024). Understanding engagement through game learning analytics and design elements: Insights from a word game case study. Proceedings of the 14th Learning Analytics and Knowledge Conference, 305–315. https://doi.org/10.1145/3636555.3636885

Mayer, R. E. (Ed.). (2014). The Cambridge Handbook of Multimedia Learning (Second Edition). Cambridge University Press.

Mayer, R. E. (2019). Computer games in education. Annual Review of Psychology, 70(1), 531–549. https://doi.org/10.1146/annurev-psych-010418-102744

Plass, J. L., Mayer, R. E., & Homer, B. D. (2020). Handbook of game-based learning. MIT Press. http://ebookcentral.proquest.com/lib/nyulibrary-ebooks/detail.action?docID=6018189

Plass, J., O'Keefe, P., Homer, B., Case, J., Hayward, E., Stein, M., & Perlin, K. (2013). The Impact of Individual, Competitive, and Collaborative Mathematics Game Play on Learning, Performance, and Motivation. Journal of Educational Psychology, 105, 1050–1066. https://doi.org/10.1037/a0032688

Poquet, O., Dindar, M., Allen, L., Dever, D., & Elizabeth Cloude. (2024). Advancing Learning Analytics with Complex Dynamical Systems: Trends and Challenges in Non-Linear Modeling of Learning Data. Companion Proceedings 14th International Conference on Learning Analytics & Knowledge, 382–385. https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24_CompanionProceedings.pdf

Scianna, J., & Kim, Y. J. (2024). Assessing Experimentation: Understanding Implications of Player Choices. https://repository.isls.org//handle/1/10760

Swanson, L., Gagnon, D., & Scianna, J. (2022, October 17). A Pilot Study on Teacher-Facing Real-Time Classroom Game Dashboards. arXiv.Org. https://arxiv.org/abs/2210.09427v1

Vanacore, K., Gurung, A., Sales, A., & Heffernan, N. T. (2024). The Effect of Assistance on Gamers: Assessing The Impact of On-Demand Hints & Feedback Availability on Learning for Students Who Game the System. Proceedings of the 14th Learning Analytics and Knowledge Conference, 462–472. https://doi.org/10.1145/3636555.3636904

Wang, K. D., Liu, H., DeLiema, D., Haber, N., & Salehi, S. (2024). Discovering Players' Problem-Solving Behavioral Characteristics in a Puzzle Game through Sequence Mining. Proceedings of the 14th Learning Analytics and Knowledge Conference, 498–506. https://doi.org/10.1145/3636555.3636907

AI-assisted Data Storytelling in Education: Tools, Methods and Future Practices

Gloria Milena Fernandez-Nieto[1], Vanessa Echeverria[1,4], Namrata Srivastava[2], Stanislav Pozdniakov[3], Yi-shan Tsai[1], Mikaela Milesi[1], Adriano Pinargote[4], Maurice Boothe Jr.[5], Roberto Martinez-Maldonado[1]
Monash University[1], Vanderbilt University[2], The University of Queensland[3], Escuela Superior Politécnica del Litoral[4], New York University
gloriamilena.fernandeznieto@monas.edu, vanessa.echeverria@monash.edu, namrata.srivastava@vanderbilt.edu, s.pozdniakov@uq.edu.au, yi-shan.tsai@monash.edu, mikaela.milesi@monash.edu, adriano.pinargote@cti.espol.edu.ec, mab1488@nyu.edu, Roberto.MartinezMaldonado@monash.edu

ABSTRACT: Data Storytelling (DS) in education has provided tools and methods to support data experts to make stories more accessible to non-data experts (i.e., learners, educators, and professional staff) while also allowing data savvy stakeholders (i.e., researchers), using human-centred approaches, in the creation process. With the rise of Generative AI (GenAI), interest has grown in exploring its potential to automate the process of creating effective data stories, as this is often a time-consuming task. This workshop seeks to foster critical discussion and hands-on activities around the opportunities and challenges of integrating GenAI tools and methods into DS in educational contexts. Key topics for exploration include: (i) How can GenAI be integrated into DS stages (analysis, planning, implementation, and communication) to automate the generation of actionable data stories that improve teaching and learning outcomes? (ii) How can researchers, designers, and educational stakeholders adapt and adopt GenAI tools and produce meaningful learning and teaching stories? (iii) What challenges and risks arise from including GenAI in DS stages? This workshop aims to bring together experts in storytelling and GenAI within the LA community to discuss and shape the future of DS in LA, addressing both its challenges and opportunities.

Keywords: Data Storytelling, Generative AI, Human-GenAI partnership, AI-assisted Data Storytelling

1 WORKSHOP BACKGROUND

1.1 Motivation

In recent years, the Learning Analytics (LA) community has embraced new methods from Human -Centred Interaction (HCI) and Information Visualisation (InfoVis) communities, such as the adoption of Data Storytelling (DS) in LA interfaces. DS, which incorporates narratives, visual elements, and storytelling techniques in large and complex data, aims to support and guide the interpretation and communication of educational insights to non-data experts stakeholders (i.e., students and teachers) (Ryan, 2016). For instance, Echeverria et al. (2018) demonstrated that DS visual elements (e.g., titles and highlights) reduce the effort needed to interpret visualisations, while Martinez-Maldonado et al. (2020) used a layered storytelling approach to communicate team performance data. Similarly, Fernandez-Nieto et al. (2021) crafted learner data stories to promote reflection and help students identify areas for improvement in clinical simulations. Pozdniakov et al. (2023) found that teachers with low visualisation literacy benefited from DS-based visual guidance. Other explorations of DS in education include its application in communicating feedback to students (Maheshi et al., 2024) and its potential for creating narratives in digital games, focusing on character and choice design (Boothe et al., 2024). While these studies show the potential of DS in improving data interpretation in LA, there remain significant challenges, particularly concerning DS automation and scalability, ethical considerations such as Fairness, Accountability, Transparency, and Ethics (FATE).

The current GenAI's capabilities to process, generate, and combine together diverse data types (e.g., datasets, text, images, audio) have enabled a human-AI partnership to support humans in diverse and complex tasks. While recent studies have begun to explore AI-assisted automation of crafting data stories to enhance accessibility and scalability (Li et al., 2023; Li et al., 2024; Ye et al., 2024), this remains a nascent area of empirical research. Given the limited work in this field, this workshop seeks to explore how GenAI can support various stages of Data Storytelling, including analysis, planning, implementation, and communication (Li et al., 2024) to support teaching and learning. Specifically, this workshop will explore benefits and challenges of using GenAI as a design material during the DS creation with stakeholders, and GenAI-enabled authoring tools for crafting data stories, uncovering topics such as overreliance, trust, human and AI roles, bias mitigation and impact on learning.

The workshop will provide participants with a space to reflect on and critically discuss the following aspects: How can GenAI be integrated into DS stages? How can researchers and practitioners adapt and adopt GenAI tools to produce meaningful learning data stories? What are the challenges and risks of relying on AI-assisted data stories? To begin, the organisers will present a review of seminal work (tools, techniques, empirical results) on using GenAI to automatically craft data stories in LA and other fields, such as InfoVis and HCI. This review will serve as a starting point for discussions on the opportunities, challenges, and risks of using GenAI to automate DS, focusing on how these stories can help educational stakeholders make sense of data traces and take actionable steps to improve their practices.

Topics of interest for this year's workshop include:

- **Human-GenAI Partnership in Crafting Data Stories.** Exploring methodologies and frameworks that support a Human-GenAI partnership during the DS design process.
- **GenAI for Automated DS in LA.** Exploring how GenAI can automate different stages of DS—from data analysis to narrative generation—toward the development of GenAI-enabled authoring tools for DS.
- **Evaluating the Impact of GenAI-generated Data Stories for Education**: Evaluating the effectiveness of GenAI-generated data stories in improving teaching and learning across different contexts.
- **Opportunities and Risks of GenAI in Automating Data Storytelling**: Identifying the potential benefits, challenges, and ethical concerns associated with using GenAI to automate the crafting of data stories. Reflecting on how these stories can remain meaningful and trustworthy for educators and learners.

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 include: 1) enable researchers and practitioners to discuss around the role of GenAI in supporting how data stories for education are crafted, 2) to discuss design implications and potential risks of using GenAI for crafting data stories for education, and 3) 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 and envision the future of DS in LA.

2 WORKSHOP DETAILS

2.1 Proposed Half-Day Workshop Schedule

The workshop is planned as a half-day event, and the following activities will be organised.

2.1.1 *Welcome and contextualisation (30 min)*

In the first activity of this workshop, the organisers will introduce current explorations of GenAI in crafting DS in LA and other research fields such as InfoVis and HCI. The motivation of this activity is to broaden participants' understanding of Human-AI collaborations for crafting DS in education and draw on inspirations from similar endeavours in other contexts.

2.1.2 *Hand-on exploration of GenAI and its current uses for DS (1.5 hours)*

Participants will engage in a hands-on activity, exploring the uses of GenAI across all stages of DS –analysis, planning, implementation, and communication (Li et al., 2024)–, and human-centred approaches to designing data stories. The activity will focus on how GenAI can be integrated into each stage to craft and design data stories for educational purposes. The organisers will provide examples and current practices of GenAI applications for crafting DS and encourage participants to apply these insights to their own domains, envisioning potential Human-GenAI collaborations for designing and creating data stories.

2.1.3 *Working groups (1 hour)*

After the hands-on activity, participants will engage in small group discussions to share their perceptions, motivations, and insights on the challenges and opportunities they foresee in the current capabilities of GenAI and its implications for DS research in LA. During this discussion, participants will reflect on the use of GenAI at various stages of crafting data stories and will discuss considerations and recommendations for future design practices and applications of GenAI in educational data storytelling.

2.1.4 *Wrap-up and conclusions (30 min)*

The workshop will conclude with a summary of key insights from the hands-on activities and group discussions, highlighting critical takeaways and future directions for exploring GenAI in educational data storytelling.

2.2 Intended outcomes

Participants attending the workshop will gain: (i) deep insights into GenAI's role in crafting DS; (ii) practical experience with GenAI tools through hands-on activities to craft DS; (iii) a critical perspective on the opportunities and challenges of using GenAI for the design and automation of DS;

(iv) and the chance to collaborate with DS experts in LA and practitioners to explore the future of DS in the GenAI era.

To support these outcomes beyond the workshop, the program, materials, and interactions will be made available through the workshop website, *DS-LAK25 Website* (to be created). 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.

REFERENCES

Ryan, L. (2016). *The visual imperative: Creating a visual culture of data discovery*. Amsterdam, Netherlands: Elsevier Science.

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 *CHI '20*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376148

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.

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 *LAK2023*. Association for Computing Machinery, New York, NY, USA, 89–99. https://doi.org/10.1145/3576050.3576063

Maheshi B., Milesi M. E., Palihena H., Zheng A., Martinez-Maldonado R., and Tsai Y. (2024). *Data Storytelling for Feedback Analytics*. In LAK Workshops (pp. 150-161)

Boothe Jr, M. A., and Brenneman J. S. (2024). *From Visualizing to Narrativizing: Powerful Data Storytelling through Non-Player Characters*. LAK Workshops. 2024. (pp. 162-170).

Li H., Ying L., Zhang H., Wu Y., Qu H., and Wang Y. 2023. *Notable: On-the-fly Assistant for Data Storytelling in Computational Notebooks*. In *CHI '23*. Association for Computing Machinery, New York, NY, USA, Article 173, 1–16. https://doi.org/10.1145/3544548.3580965

Ye, H., Zhang, J., Wu, Y., & Cao, N. (2024). *Generative AI for visualization: State of the art and future directions*. *IEEE Transactions on Visualization and Computer Graphics, 30*(1), 1211-1221. https://doi.org/10.1016/j.visinf.2024.04.003

Li H., Wang Y., and Qu H. (2024). *Where Are We So Far? Understanding Data Storytelling Tools from the Perspective of Human-AI Collaboration*. In CHI '24. Association for Computing Machinery, New York, NY, USA, Article 845, 1–19. https://doi.org/10.1145/3613904.3642726

The 2nd Workshop on Adaptive Lifelong Learning (ALL25)

Alireza Gharahighehi
KU Leuven, Itec, imec research group, Kortrijk, Belgium
KU Leuven Campus Kulak, Department of Public Health and Primary Care, Kortrijk, Belgium
alireza.gharahighehi@kuleuven.be

Rani Van Schoors
KU Leuven, Itec, imec research group, Kortrijk, Belgium
KU Leuven, Centre for Instructional Psychology and Technology, Leuven, Belgium
rani.vanschoors@kuleuven.be

Paraskevi Topali
Radboud University, Faculty of Social Sciences, National Educational Lab AI (NOLAI), Nijmegen, the Netherlands
evi.topali@ru.nl

Jeroen Ooge
Utrecht University, Department of Information and Computing Sciences, Utrecht, the Netherlands
j.ooge@uu.nl

ABSTRACT: This half-day interactive workshop emphasizes the role of adaptive lifelong learning in the dynamic landscape of learning analytics (LA). When learners, trainees, teachers and trainers are confronted with AI in an educational context, they often face challenges such as information overload or the necessity to exhibit high degrees of flexibility to adapt to the rapid changes and continuous evolution of learning tools. By highlighting the virtues of AI for adaptive learning, the workshop examines the impact of AI as a source of support for bridging learning gaps and differences, to streamline automation and so on. Furthermore, the workshop emphasizes the critical necessity for multidisciplinary expertise in the evolving LA domain, moving beyond its roots in computer science (technical knowledge) to encompass a broader educational perspective (didactical knowledge). With a specific focus on multi-criteria adaptive learning, the workshop advocates for collaboration among learners, educators, policymakers, researchers, and EdTech companies. The goal is to facilitate the development of relevant and evidence-based adaptive learning tools that significantly enhance and support lifelong learning.

Keywords: adaptive learning, lifelong learning, explainability, personalized learning, human-centred design

1 RELEVANCE AND IMPORTANCE

1.1 Why Lifelong Learning?

Driven by digitization and datafication, innovative technologies are increasingly integrated into educational research and policies [4, 5, 8]. Within this context, the use of AI in education is a timeless issue for educational practitioners and researchers worldwide [5, 11] because the field of AI is characterized by rapid progress and technological pushes towards more automation. This results in skill gaps and forces individuals and organizations to continually update their knowledge and capabilities to remain relevant in an increasingly AI-driven world. In education specifically, learners and trainees face significant challenges related to demonstrating high flexibility and keeping up with the rapid changes and continuous development of new tools [15]. As a result, and further accelerated by the COVID-19 pandemic, millions of lifelong learners have turned to online learners.

1.2 Why adaptive lifelong learning?

There is a growing interest in adaptive learning because of its many presumed benefits [3, 9, 10], Such as positive impact on cognitive and non-cognitive learning outcomes [6, 13, 14]. Furthermore, many praise adaptive learning for considering the heterogeneity of learners and trainees and their needs by appropriately personalizing exercises, scaffolds, and assessments [4]. Moving away from the traditional 'one-size-fits-all' learning and training approach, it is believed that adaptive tools can remediate learning gaps, especially in the post-pandemic context [1, 8, 12, 13]. To personalize learning with adaptive learning systems, learner models play a key role. These dynamic models represent learners' evolving knowledge and understanding [2] and can be based on various cognitive, pragmatic, or data-driven approaches. Aligned with the call for more transparent and explainable AI systems [10], researchers have argued for transparent learner models. For example, Bull and Kay [2] proposed Open Learner Models through which learners can better understand and maintain their learner models. Furthermore, Kay and Kummerfeld [7] discussed the importance of scrutable user modeling for addressing challenges such as privacy, the invisibility of personalization, errors in user models, wasted user models, and the controllability of user models. Compared to formal learning contexts, adaptation is even more essential in lifelong learning because learners are afforded more choices and flexibility, requiring learners to adjust to the evolving educational landscape.

1.3 Why Adaptation Based on Multiple-Criteria?

This workshop focuses on multi-criteria adaptive lifelong learning, encompassing descriptive, predictive, and prescriptive learning analytics, as well as language processing, speech, and image recognition/processing. The objective is to tailor learning to various factors, including knowledge, skills, and motivation. Multi-objective [17] and multi-task [18] learner models, or ensemble methods [19] can be applied to achieve such adaptive systems. We also examine the multi-dimensionality of learning contexts, covering both individual and group levels of adaptation. Our approach transcends traditional classroom settings, facilitating learning anywhere and anytime, such as at home. We emphasize research involving multiple stakeholders, highlighting the importance of robust partnerships among learners, teachers, policymakers, researchers, and EdTech companies. By sharing knowledge, co-designing interventions, and engaging in ongoing dialogue, we can develop more evidence-based adaptive tools that are applicable and valued in real-world educational settings.

2 ORGANIZATION

This is the second edition of this workshop. We plan to build on the success of the first edition [16], which was co-located with AIED 2024 and featured invited talks, paper presentations, and interactive group discussions. The organizing team includes expertise in both technical and didactical perspectives. Technically, the team specializes in various aspects of personalization, such as recommendations in

MOOCs [20], addressing the cold-start problem in adaptive systems [21], and explainability [22] and control [23] in AI-based adaptive learning systems. From a didactical standpoint, the team focuses on the potential of artificial intelligence in education [24] and human-centered AI and learning analytics.

3 CONTENT AND THEMES

This workshop aims to unite a community of researchers interested in various aspects of lifelong learning, particularly in adaptation and personalization. We invite contributions covering all topics related to adaptive lifelong learning, from both technical and didactical perspectives, with a specific focus on, but not limited to, the following list:

- Tailoring learning to various factors, including knowledge, skills and motivation
- Extending the range of adaptation criteria beyond mere relevance, incorporating factors such as fairness, diversity, bias, and pedagogical aspects
- Group-aware and context-aware adaptations in lifelong learning
- Lifelong learning adaptation involving multiple stakeholders
- Adaptive learning in Massive Open Online Courses (MOOCs)
- Quantifying learners' engagement and dropout risk
- Adaptive or personalized nudging strategies within lifelong learning
- Multi-modal adaptive lifelong learning
- Adaptive educational games for enhancing lifelong learning
- Adaptive computer-assisted language learning
- Adaptive communication with learners (e.g., feedback, dashboard, chatbots, etc.)
- Adaptive simulations in workstations
- Explainable adaptations in lifelong learning
- Generating adaptive learning trajectories

4 TAREGT AUDIENCE, COMMUNICATION AND DISSEMINATION

The workshop is designed for researchers, PhD researchers, and master's students who are actively engaged in lifelong learning, adaptive learning, learning in MOOCs, language learning, recommendation systems, educational data mining, and learning analytics. It is also relevant for employees at EdTech companies interested in technologies for lifelong learning. Overall, we anticipate 15–30 participants. The outcome of the workshop, including accepted papers, presentations, and group discussions, will be shared through our workshop proceedings, website, and a concluding paper.

ACKNOWLEDGEMENTS

This workshop is supported by Itec, imec research group at KU Leuven. The authors also acknowledge support from the Flemish Government (AI Research Program).

REFERENCES

[1] Breines, M.R. and Gallagher, M. 2023. A return to Teacherbot: rethinking the development of educational technology at the University of Edinburgh. *Teaching in Higher Education*. 28, 3 (2023).

[2] Bull, S. and Kay, J. 2016. SMILI: A Framework for Interfaces to Learning Data in Open Learner Models, Learning Analytics and Related Fields. *IJAIED*. 26, 1 (2016).

[3] Cardona, M.A. et al. 2023. *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*.

[4] Fadel, C. et al. 2019. Artificial intelligence in education: Promises and implications for teaching and learning. *Journal of Computer Assisted Learning*. 14, 4 (2019).

[5] Gill, K.S. 2022. Nowotny, Helga (2021). In AI we trust: power, illusion and control of predictive algorithms, Polity, Cambridge, UK, ISBN-13: 978-1509548811. *AI & SOCIETY*. 37, 1 (2022).

[6] Holmes, W. and Porayska-Pomsta, K. 2022. *The ethics of artificial intelligence in education: Practices, challenges, and debates*.

[7] Kay, J. and Kummerfeld, B. 2012. Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *ACM Transactions on Interactive Intelligent Systems*. 2, 4.

[8] Knox, J. 2023. *AI and Education in China: Imagining the Future, Excavating the Past*.

[9] Maslej, N. et al. 2023. The Artificial Intelligence Index 2023 Annual Report. *Institute for Human-Centered AI, Standford University*.

[10] Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*.

[11] Salomon, G. 2002. Technology and Pedagogy: Why Don't We See the Promised Revolution? *Educational Technology*. 42, 2.

[12] Sims, A. et al. 2022. UNICEF Public Consultation on Draft Policy Guidance on AI for Children. *SSRN Electronic Journal*. (2022). DOI:https://doi.org/10.2139/ssrn.4088822.

[13] Zhai, X. et al. 2021. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*.

[14] Zhang, K. and Aslan, A.B. 2021. AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*.

[15] Zimmerman, M. 2018. *Teaching AI: Exploring new frontiers for learning*.

[16] Gharahighehi, A., Van Schoors, R., Topali, P. and Ooge, J., 2024. Adaptive lifelong learning (ALL). In International Conference on Artificial Intelligence in Education (pp. 452-459). Cham: Springer Nature Switzerland.

[17] H. Li, Z. Zhong, J. Shi, H. Li, Y. Zhang, 2021. Multi-objective optimization-based recommendation for massive online learning resources, IEEE Sensors Journal 21, 25274–25281.

[18] M. Geden, A. Emerson, J. Rowe, R. Azevedo, J. Lester, Predictive student modeling in educational games with multi-task learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 654–661.

[19] A. Gharahighehi, C. Vens, K. Pliakos, An ensemble hypergraph learning framework for recommendation, in: Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings 24, Springer, 2021, pp. 295–304.

[20] Gharahighehi, A., Venturini, M., Ghinis, A., Cornillie, F. and Vens, C., 2023. Extending bayesian personalized ranking with survival analysis for mooc recommendation. In Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (pp. 56-59).

[21] Gharahighehi, A., Pliakos, K. and Vens, C., 2022. Addressing the Cold-start problem in collaborative filtering through positive-unlabeled learning and multi-target prediction. IEEE Access, 10, pp.117189-117198.

[22] Ooge, J., Kato, S. and Verbert, K., 2022. Explaining recommendations in e-learning: Effects on adolescents' trust. In Proceedings of the 27th International Conference on Intelligent User Interfaces (pp. 93-105).

[23] Szymanski, M., Ooge, J., De Croon, R., Vanden Abeele, V. and Verbert, K., 2024. Feedback, Control, or Explanations? Supporting Teachers With Steerable Distractor-Generating AI. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 690-700).

[24] Itec, 2024: Learning, teaching & training in the era of artificial intelligence: Challenges and opportunities for evidence-based educational research. Acco

AI in Education: Towards Developing International Standards

Haiming Liu, Kate Borthwick, Jize Yan
University of Southampton
h.liu@soton.ac.uk, k.borthwick@soton.ac.uk, j.yan@southampton.ac.uk

ABSTRACT: This workshop addresses the urgent need for international standards in the use of artificial intelligence (AI) in education. By gathering key stakeholders, including educators, AI scientists, industry leaders, and regulators, this event aims to identify the requirements for ethical, safe, and effective AI integration in educational contexts. Through discussions on the current needs of higher education, enterprise opportunities, and the development of actionable standards, the workshop will establish a roadmap for global collaboration. A primary outcome will be the creation of a white paper to guide the development of principled international standards for AI in education, ensuring tools are designed responsibly to support learning and professional growth.

Keywords: AI in Education, Standards, Responsible AI, Educational Technology, Human-centric AI

1 INTRODUCTION

AI in education has never been more important topic before. The Generative AI technologies emerged in 2022 is a game-changing and disruptive innovation for education. We want to embrace the change and innovations in education technology development. To achieve this we need to ensure all stakeholders such as educational technology companies, educators, students and regulators are all supported for the change through creating frameworks and standards for designing and developing human-centric AI education technologies. Currently there is a global call for action on principled regulations for AI in Education with the booming of AI in Education technologies developed by research institutions and companies, such as TurnitIn, Keath.ai, Learnado.ai, Habitat Learn, etc. It is a desperate call for practitioners, AI scientists, educators, regulators, teachers and student to get together to discuss and create the international standards for AI in Education.

We have been working with The British Standards Institution (BSI) to scope an AI in Education Standard to fulfil the needs of both business sector (such as entrepreneurs) and education sector (such as High Education) for their regulation and operation and procurement in the new AI era. The development of the British standard is following the system-of-systems approach, which will produce a cluster of components for AI in education. We propose this co-discovery workshop to invite international stakeholders to join us to build an international community in an aim to scope and develop the British standards further into international standards, which will benefit AI in Education globally.

2 RELATED WORK

Since the emergence of Generative AI (GAI) in 2022, all sectors, including education, have been required to adapt to this transformative technology. Many universities and regulatory bodies have begun exploring the impact of GAI on education on an ad hoc basis. However, there is a lack of unified and principled standards to guide practitioners, educators, students, and regulators. These standards are needed to determine which AI products should be used in education, who should use them, and how they should be applied to enhance teaching, learning, and professional development. Additionally, businesses require guidance to develop safe, ethical, and responsible AI products for educational purposes.

Šedlbauer et al. (2024) examined students' opinions on using ChatGPT for learning, finding varying levels of acceptance. Their study also revealed that GAI can enhance critical thinking when integrated with principled technology and learning design in education. Despite efforts to develop guidelines for adapting and integrating AI in education, existing research largely addresses specific, localized issues rather than universal challenges. For instance, Bozkurt (2024) investigated the use of GAI to improve learning outcomes by setting rules and standards for prompts, while Williams (2024) proposed policies and detection technologies to address ethical concerns and uphold academic integrity. Other studies explore regulations from an institutional perspective. Wu et al. (2024) reviewed and summarized generative AI regulations in universities, identifying three prevailing approaches: a) explicit prohibition, b) conditional use (at the teacher's discretion), and c) encouragement. These differing attitudes highlight the challenges of establishing universal standards for GAI use in education. In addition to regulatory research, some studies propose user manuals for AI in education, emphasizing that proper training can enhance AI literacy among students and teachers. This model suggests extending AI-related education to all citizens, making AI literacy a prerequisite skill for effectively using generative AI tools (Chiu, 2023; Rodriguez et al., 2024).

Global efforts to develop GAI standards prioritize ethical, safe, and effective use, particularly in education. UNESCO's 2023 guide provides a broad ethical framework, while the EU's approach combines comprehensive regulations with specific guidance for educators but neglects other key stakeholders (European Commission, 2021). The European Union's Artificial Intelligence Act (AI Act) (2024) introduces a comprehensive regulatory framework that significantly impacts the use of AI in education, classifying it as a high-risk domain for AI applications. The Act underscores the EU's commitment to promoting safe, ethical, and effective AI integration in education, with the aim of enhancing learning outcomes while safeguarding the rights and well-being of students and educators. The UK (UK Government, 2023) provides a theoretical framework for education and governance but lacks mechanisms for practical implementation. The U.S. (NIST, 2024) emphasizes flexible, goal-oriented standards and international cooperation, though it offers limited education-specific details. Meanwhile, China's approach focuses on legality, ethics, and national interests but provides minimal guidance for educational applications (Cyberspace Administration of China, 2023). While these initiatives demonstrate progress, most remain conceptual, lack operational clarity, and fail to address the diverse needs of education stakeholders. More detailed and actionable strategies are urgently required to effectively support the integration of AI in education.

Based on the literature review, we have begun scoping and developing an AI in Education Standard in the UK. Our goal is to advance this British standard into an international framework. By hosting this workshop, we aim to bring together global stakeholders to share knowledge and collaborate on developing international AI in education standards, addressing the pressing global demand for principled guidelines.

3 THE WORKSHOP

3.1 Schedule of the Workshop

The workshop will run as a symposium style which will not be calling for submissions, instead we will run topic-based discussions for the standard scoping.

00:00-00:15 (15 mins) Opening and introduction.

00:15-01:00 (45 mins) Needs and expectations in the HE sectors, Invited talk(s) and discussion.

01:00-01:45 (45 mins) Enterprise Opportunities of AI in Education, Invited talk(s) and discussion.

01:45-02:15 (30 mins) Coffee break

02:15-03:00 (45 mins) Towards developing Interactional standards, Invited talk(s) and discussion.

03:00-03:45 (45 mins) AI in Education Standard scoping exercise

03:45-04:00 (15 mins) Closing

3.2 Outcome of the workshop

The expected outcome of the workshop is to develop a white paper based on the discussion and the result of the scoping exercise, which will be the based of the standard document.

3.3 Workshop organization team

Dr Haiming Liu is an Associate Professor and Director of Centre for Machine Intelligence (CMI) at University of Southampton. Haiming specializes in designing and developing interactive multimodal information seeking, access and retrieval technologies. Her research outcomes have been applied to multidisciplinary domains such as AI in Education. Haiming is active in knowledge exchange and innovation projects that aim to improve the industrial productivity and solve business problems. Haiming is pertinent in developing AI in Education Standards to bridge the gap among key stakeholders.

Prof Kate Borthwick is an experienced and award-winning learning designer and digital educator. She is the Academic Lead for Generative AI in Education at the University of Southampton and Chair of the University of Southampton’s Digital Education Advisory Group, reporting to and advising the Vice-President, Education and Student Experience. She chaired the Generative AI working group July 2023 – Jan 2024 which advised on policy and practice in relation to generative AI in education. Kate is also Director of University open online courses and an enterprise fellow involved in working with external clients to produce CPD and short courses. She is a lecturer in digitally mediated language learning and teaching in the Faculty of Arts and Humanities and her research interests are open educational practices and online and digital education.

Prof. Jize Yan at the University of Southampton holds a BEng from Tsinghua University and a PhD from the University of Cambridge. As a PI and co-I, he's been awarded £28M+ in research grants, published

100+ peer-reviewed publications, and earned multiple best-paper awards. With a portfolio of 10+ patents, Jize co-founded spin-out companies and facilitated technology transfer. In education, he collaborates with leading AI education enterprises to champion AI technologies and promote future AI education standards.

REFERENCES

Bozkurt, A. (2024). 'Tell Me Your Prompts and I Will Make Them True: The Alchemy of Prompt Engineering and Generative AI. Open Praxis, 16(2), p. 111–118.

Chiu, T. (2023). The Impact of Generative AI (GenAI) on practices, Policies and Research Direction in education: a Case of ChatGPT and Midjourney. Interactive Learning Environments, pp. 1-17.

Cyberspace Administration of China (2023). Interim Measures for the Administration of Generative Artificial Intelligence Services. [Online]. Available at: https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm. [Accessed 24 11 2024].

European Commission (2021). Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). [Online]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. [Accessed 24 11 2024].

European Union. (2024). Regulation (EU) of the European Parliament and of the Council of June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Official Journal of the European Union. Retrieved from https://eur-lex.europa.eu.

NIST (2024). A Plan for Global Engagement on AI Standards. [Online]. Available at: https://doi.org/10.6028/NIST.AI.100-5 [Accessed 24 11 2024].

Rodriguez, M. et al. (2024). Introducing Generative Artificial Intelligence Into the MSW Curriculum: A Proposal for the 2029 Educational Policy and Accreditation Standards. Journal of Social Work Education, 60(2), p. 174–182.

Šedlbauer, J., Činčera, J., Slavík, M. & Hartlová, A. (2024). Students' reflections on their experience with ChatGPT. Journal of Computer Assisted Learning. 40. 10.1111/jcal.12967.

UK Government (2023). Generative artificial intelligence (AI) in education. [Online]. Available at: https://www.gov.uk/government/publications/generative-artificial-intelligencein-education/generative-artificial-intelligence-ai-in-education. [Accessed 24 11 2024].

UNESCO (2023). Guidance for generative AI in education and research. [Online]. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000386693?locale=en [Accessed 24 11 2024].

Williams, R. (2024). The ethical implications of using generative chatbots in higher education. Frontiers in Education, Volume 8.

Wu, J., Yang, Z., Wu, S. & Zou, D. (2024). Unveiling the synergy of peer feedback and the Metaverse. Computers & Education: X Reality, Volume 4, p. 100056.

The limits of learning analytics: Introducing perspectives on learning from psychometrics, sociology, and learning theory

Danielle Hagood
University of Copenhagen
dh@ind.ku.dk

Liv Nøhr
University of Copenhagen
lsnl@ind.ku.dk

ABSTRACT: In this workshop we explore the limits of learning analytics. We acknowledge some limits in learning analytics raised in recent scholarly discourse. In response, we reframe limits as the horizons of learning analytics—areas of potential and excitement. We introduce perspectives that may serve as useful guides through new frontiers from psychometrics, sociology and learning theory. This workshop blends collaborative theoretical reflection and practical research design considerations. In this workshop, participants are positioned as collaborators, and the workshop leaders facilitate discussion by highlighting three relevant concepts in theory, providing summaries of research, and designing resources and activities to structure reflection, debate, and envision the methodologies underpinning our work into the future.

Keywords: research design, methodology, theoretical perspectives, validity, rigor

1 WORKSHOP BACKGROUND

The starting point for this workshop—"the limits of learning analytics"—is intentionally provocative. By learning analytics' "limits", we mean both its potential *weaknesses* and the fields' *frontiers*. By introducing core perspectives from three fields, we hope to expand the horizon of learning analytics research as it continues to mature and deepen in terms of methodology and theoretical perspectives (Bergner et al., 2018; Chen & Poquet, 2022). We focus on psychometrics, sociology, and leaning science because each foregrounds a different epistemic perspective at the limits of learning analytics.

From psychometrics we draw a perspective on validity which challenges the existence of "ground truth" in data (Kane 2001). Despite a its 100-year history wrestling with the concept of validity, the analytics techniques in measurement are quite different from those in analytics, but by bridging these differences in techniques, there are core ideas in validity theory that challenge and also enhance machine learning techniques commonly used in learning analytics (Tay et al. 2020; Trognon et al. 2022). From sociology, we draw a perspective emphasizing the situated nature of learning and interconnection between micro-, meso-, and macro- activities (Bergener et at. 2018; Naidoo, 2003). The strong emphasize on prediction and classification from large data in learning analytics sometimes views the abstract and complex sociological forces as "outside" our scope. We raise the question, what are we missing if we don't include these concepts and consider what research approaches can be used to study them in learning analytics. Finally, from learning theory, likely the area most familiar to LAK participants, we draw the perspective that there are different theories of learning (McInerney, 2005). We reflect on when and why learning analytics research sometimes fails to be specific enough in its theoretical grounding.

This workshop joins an ongoing reflection in the field of learning analytics which seeks to establish a rigorous methodological and theoretical foundation that goes beyond simply making interesting models with data (Fergusen, 2012, Prinsloo, 2019, Selwyn, 2020). Recently, in their evaluation of learning analytics, Guzmán-Valenzuela and co-authors (2021) concluded that “learning analytics focused more on analytics than on learning.” We take this assessment as a challenge to reflect on limitations in in the field at present. But also, we take this as an opportunity to for growth. Learning analytics is an interdisciplinary field, with researchers and practitioners from many backgrounds. No one can be an expert in everything. To this end, we hope that this workshop will introduce insights from other fields that prompt discussion and connect people in collaborative research practices (Mackenzie & Knipe, 2006).

This workshop topic is especially relevant this year, in alignment with the main conference theme “Expanding the Horizons of Learning Analytics”, more specifically our workshop focuses in on “expanding our theoretical perspectives” and “expanding our methodological toolbox.” Anyone with an interest in research design and methodology will find this workshop relevant. There is no need for a specific background to benefit from participating. This workshop builds on past LAK and LASI workshops that explored methodology in relation to learning analytics (Hagood et al., 2024).

2 WORKSHOP DETAILS

2.1 Event Type & Structure

We propose a full-day workshop for up to 40 participants. Both newcomers and experts to learning analytics will be able to participate fully. No technical expertise is necessary, but participants should be interested in methodology and research design from their respective disciplinary backgrounds. The workshop will include reflective and hands-on activities through which the participants and workshop leaders will develop a deep understanding and position on research design in learning analytics. These activities will focus on the challenges posed by the three highlighted perspectives, their relevance to our respective research design frameworks, 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 passive lecture event. Despite presenting theoretical topics, we find it important that participants introduce themselves and share their research beliefs and methods knowledge through introducing activities. We have planned several activities that will introduce everyone and start to expose how we think about theory and research in an engaging and (hopefully) fun way. These reflections are relevant to the whole workshop and will be referred to throughout the day. As facilitators, we will present key definitions for ideas and curated selections from research throughout these activities to frame these activities, inform, and promote debate.

2.2.2 *Digging into the Three Perspectives: 2 hours*

In turn, we will introduce the three perspectives from psychometrics, sociology, and learning theory. This section of the workshop will include both short presentations from the organizers (to share theory and relevant studies with the participants) and discussing this information and collaboratively completing reflection activities in small groups (to share perspectives between participants). This section of the workshop aims to raise challenging questions which may not have definitive answers. We have designed conceptual reflection tools to engage the participants and support developing shared perspectives, identifying differences, and representing these perspectives systematically.

2.2.3 *Advancing Methodology in Learning Analytics: 2 hours*

In the second half, the workshop will focus on connecting the conceptual perspectives from the first half to the participants actual research in practice. The participants will be asked to briefly present their work, with an emphasis on the research design choices and challenges. The structure of this phase of the workshop depends on the number of participants and will either be a plenary or divided into smaller groups, so each presenter has 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 to the Methodology in Learning Analytics SIG, as members interests align with this workshop. Additionally, as this workshop builds of two previous workshops, we will invite those participants to join us again, if they wish to go further with these topics. This workshop will also interest those outside learning analytics, in particular researchers in psychometrics, sociology, and learning theory, who are interested in learning analytics. So, we plan to share this workshop with related SIGs and research organizations to disseminate this event widely and beyond the learning analytics community.

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.

3 INTENDED OUTCOMES

First, for participants, we hope to introduce relevant concepts for reflection on research design. Participants will clarify and organize their methodology beliefs and consider how these influence research at different stages. Participants will also contribute with their own knowledge and perspectives to our collective discussion.

Second, for the community at large this workshop provides a venue to assemble careful thought and reflection on methodology in learning analytic. We see the discussions and work from participants as a multidisciplinary thinking space that offers an opportunity to consolidate and advance the ongoing conversation around theory and research rigor in learning analytics. Based on developing this workshop and the resulting discussions it prompts; we plan to write a paper the broader learning analytics and education research audience. This paper as well as the participants experiences contributes to advancing methodology in learning analytics research.

REFERENCES

Ballantine, J., Stuber, J., & Everitt, J. (2021). *The sociology of education: A systematic analysis*. Routledge.

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.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304-317.

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.

Hagood, D. Spikol, D., Nøhr, L. Echeverria, V., Martinez-Maldonado, R., Nieto, G. Srivastava, N., Arslan, A., Curkurova, M., Di Mitri, D., Emerson, A., Giannakos, M., Ochoa, X., & Wang, Y. (2024). Design-based research methodology: Going deeper than methods in multimodal learning analytics. *Companion Proceedings 14th International Conference on Learning Analytics & Knowledge (LAK 24),* 386-390.

'Kane, M. T. (2001). Current concerns in validity theory. *Journal of educational Measurement*, *38*(4), 319-342.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in educational research*, *16*(2), 193–205.

McInerney, D. M. (2005). Educational psychology–Theory, research, and teaching: A 25-year retrospective. *Educational psychology*, *25*(6), 585-599.

Naidoo, R. (2003). Repositioning higher education as a global commodity: Opportunities and challenges for future sociology of education work. *British Journal of Sociology of Education*, *24*(2), 249-259.

Prinsloo, P. (2019). Learning analytics: Mapping a critique and agenda. *Journal of Learning Analytics*, *6*(3), 20-24.

Selwyn, N. (2020). Re-imagining 'learning analytics'… a case for starting again?. *The Internet and Higher Education*, *46*, 100745.

Tay, L., Woo, S. E., Hickman, L., & Saef, R. M. (2020). Psychometric and validity issues in machine learning approaches to personality assessment: A focus on social media text mining. *European Journal of Personality*, *34*(5), 826-844.

Trognon, A., Cherifi, Y. I., Habibi, I., Demange, L., & Prudent, C. (2022). Using machine-learning strategies to solve psychometric problems. *Scientific Reports*, *12*(1), 18922.

How to avoid being LAK-luster: Strategies for impactful presentations of learning analytics research

Linda Corrin
Deakin University, Australia
linda.corrin@deakin.edu.au

Aneesha Bakharia
The University of Queensland
a.bakharia1@uq.edu.au

ABSTRACT: Presenting research in complex interdisciplinary fields such as Learning Analytics (LA) to diverse audiences can often be challenging. Despite the fact the field is full of innovation and ideas that can spark collaboration, new ideas, and development opportunities, sometimes these outcomes are not as impactful as they could have been due to how they were presented to the audience. Opportunities to develop better presentation skills and improve how the story of the research can be constructed and presented are rare in the LA field. Therefore, the aim of this half-day, face-to-face workshop is to inspire researchers and practitioners to rethink how they design and deliver presentations of learning analytics research so that they can better engage their audience and increase the ongoing impact of their work. The interactive session will include creative activities designed to explore a range of strategies for making presentations of LA research more effective, impactful, and engaging. The workshop team will draw on key literature on effective science communication, presentation design, and delivery, to create a safe space for participants to play with different presentation structures, make effective use of visual aids, and deliver their LA research outcomes in ways that can be memorable and impactful.

Keywords: Learning analytics, research, impact, presentation, creativity

1 BACKGROUND

The way that the outcomes of a research study are presented to an audience can embed it in the minds of the attending audience or render it instantly forgettable. Sometimes this is due to the highly technical nature of the research, but it can also be that the audience is left to fill in the gaps that the short conference presentation slot may not allow full detail to be presented. The workshop team, who have been regular attendees and presenters at Learning Analytics and Knowledge (LAK) and other educational conferences over the years, have been concerned for some time about the fact that the impact of interesting and innovative research is sometimes obscured by the presentation approach taken by the presenter. Learning Analytics (LA) research is a complex field with studies ranging from highly quantitative experiments using advanced machine learning techniques, to LA systems implementation studies, to qualitative explorations of LA-generated feedback, to conceptual discussions of social, ethical, and policy implications of LA use. Additionally, the interdisciplinary nature of the LA field means that the audiences this research is being presented to are often diverse in terms of discipline, prior knowledge, and interest (Kitto et al., 2018). This feature of the LA community was recognised as a great asset, but also an interesting challenge, in the very first editorial of the *Journal of Learning Analytics* (Siemens, 2014). Some audience members may have quite specialized, domain specific knowledge without awareness of the broader LA context, while others may have a broader conceptual view of the LA landscape but a weaker understanding of technical and/or methodological approaches.

There are few studies that examine engagement with and impact of conference presentations in an empirical manner. Instead, there are many blog posts, articles, and "how to avoid" resources on bad conference presentations as well as journal editorials and commentary pieces contain advice on presentation skills development (Dolan, 2017), or rules for improving the design of presentation slides (Naegle, 2021). While there have been many studies examining the range of research being conducted in the LA field (e.g., Romero & Ventura, 2020), there are few studies that examine the ways in which learning analytics research is presented via conference presentations and what understanding the audience experiences of this work.

Therefore, we feel that the LAK community would benefit from an opportunity to come together and discuss and explore how LA research can be presented in ways that can increase engagement and impact. Our goal is to build capacity in a creative way using what is known about good presentation approaches that encourage attention and understanding informed by literature on effective science communication, theories of multimedia learning, and theatre studies to help researchers improve the way they tell the story of their research and communicate its outcome to achieve greater impact.

2 OBJECTIVES OF THE WORKSHOP

The main objective of this workshop is to help participants to develop a broader awareness of ways that LA research and practice can be communicated that will result in engagement and lasting impact. The workshop is designed to provide participants with a space to experiment and build confidence in how to present to diverse audiences like the LA community and brings together learnings from existing research with creative activities to inspire participants to try new things and give and receive feedback with peers. The workshop aims align with the 2025 LAK conference themes of "Expanding the Horizons of Learning Analytics" by helping to open up key findings from LA research studies through clear and engaging presentation that can inspire future research and practice in LA.

3 INTENDED AUDIENCE FOR THE WORKSHOP

The workshop is open to anyone in the learning analytics community, including researchers and practitioners of all levels of experience. For PhD researchers and early career academics it will provide guidelines to build foundational confidence and skill in designing and delivering their early ideas to a broad audience for feedback. For mid- or advanced-career researchers/practitioners it will provide a useful opportunity to try new presentation approaches and get feedback from their peers on how to improve previous approaches. Between the workshop facilitators and other participants there will be a mix of different skills and experiences in the room to create a rich environment for the generation of creative ideas, friendly critique, and useful feedback.

4 WORKSHOP PROMOTION AND PREPARATION

Promotion of the workshop will take place across a range of platforms. The workshop team will establish a website where an outline of the purpose and design of the workshop will be presented and which will provide a communication channel through which potential participants can ask questions. This website and the LAK25 registration process will be promoted via a range of social media platforms (e.g., LinkedIn, Bluesky, etc.) and mailing lists (e.g., SoLAR) to potential participants. A few weeks prior to the workshop, the workshop team will contact the participants to introduce themselves and ask them to start to think about the LA research that they might like to bring along to use as part of the workshop activities.

5 WORKSHOP DESIGN

The workshop will start with an icebreaker activity to help participants get to know each other and to seed creative ideas. The workshop is intended to be a safe space to experiment with new ideas and so this will be explored through these initial interactions. We will ask the participants to provide some background on the area of LA research and/or they focus on any particular concerns they may have about their current approaches to presenting their work. The main part of the workshop include:

1. **Structuring the presentation:** This section will focus on the role of storytelling through a series of short activities that explore how a presentation can be structured for the greatest impact. The importance of surfacing the outcomes of the research early so that these are given as much time as the explanation of the background and methodology of the research will be discussed (i.e., how to divide the structure of the presentation so you don't run out of time for the most important part!). Participants will be encouraged to try out different ways of starting and wrapping up a presentation.

2. **Use of visual elements/aides**: The use of visual aides can make or break a presentation. In the second section of the workshop we will begin some of the core principles for using visual elements to enhance a presentation will be introduced. Participants can then decide on the types and number of visuals (or slides) necessary to help deliver the main outcomes of their work. The workshop facilitators have extensive research and practical experience in the visualisation of LA data, and will use this to provide feedback on how visualisations can be accessible and understandable for a wide range of audiences. Resources on tools and visualisation designs will be provided to participants to help inspire new ideas and to use for future reference.

3. **Delivery of the presentation**: The third section of the workshop is inspired by the principles of theatre studies (Cohen & Dreyer-Lude, 2020) and will focus on how the presentation is delivered (whether in person or online). In particular, we will explore how to improve elements such as voice projection, diction, pace, and posture. In this section, participants will be given the opportunity to present parts of their enhanced presentations to their peers to practice the presentation delivery skills featured in this section and receive further feedback.

A sharable Google Slide presentation file will be accessible to all participants which will contain the workshop resources and links to featured research and examples. We will also use this presentation deck to collect together notes from the participants and facilitators to inform the final reflection activity and help to inspire ideas for how the conversation can be continued beyond the workshop. Access to this file will remain open to participants after the workshop for future reference.

The workshop will conclude with an opportunity for participants to reflect on what they have learnt from the workshop activities and can opt to have a peer review of their LAK25 conference presentation conducted by one of the workshop coordinators. Participation in the peer review element is optional, but is offered in the spirit of continuing the conversation and providing further feedback on how the participant has put what they have learnt in the workshop into practice.

6 WORKSHOP OUTCOMES

A key outcome of the workshop will be the consideration of different ways that LA research can be presented to diverse audiences. The workshop activities will enable participants to reflect on their current ways of presenting, explore and play with new approaches, and receive feedback from peers in the field to improve their presentation practice. These outcomes can also be applied beyond the LA space and could inform participants' practice in other elements of their institutional roles where clear communication of complex concepts is necessary. The workshop team plan to write a blog post that could be submitted to the SoLAR Nexus blog to share the discussions and ideas generated throughout the workshop with the wider LA community. The workshop team are also keen to continue the exploration of the elements that make presentations of LA research (in particular) more impactful. Participants will be given an opportunity to help develop a project to explore how LA research is commonly presented, and to look for examples of good practice and areas for further development. The outcomes of this work could be presented at LAK26 to keep the conversation (and learning) going.

REFERENCES

Cohen, I., & Dreyer-Lude, M. (2020). *Finding Your Research Voice: Story Telling and Theatre Skills for Bringing Your Presentation to Life*. Springer Nature.

Dolan, R. (2017). Effective presentation skills. *FEMS microbiology letters, 364*(24), fnx235. https://doi.org/10.1093/femsle/fnx235

Kitto, K., Shum, S. B., & Gibson, A. (2018). Embracing imperfection in learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 451–460). https://doi.org/10.1145/3170358.3170413

Naegle, K. M. (2021). Ten simple rules for effective presentation slides. PLoS computational biology, 17(12), e1009554. https://doi.org/10.1371/journal.pcbi.1009554

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355. https://doi.org/10.1002/widm.1355

Siemens, G. (2014). The journal of learning analytics: Supporting and promoting learning analytics research. *Journal of Learning Analytics, 1*(1), 3-5. https://doi.org/10.18608/jla.2014.11.2