LAK23 Conference Proceedings

Towards Trustworthy Learning Analytics

The Thirteenth International Conference on Learning Analytics & Knowledge

March 13-17, 2023
Hybrid,
Arlington, Texas, USA

Organized by: SOLAR
SOCIETY for LEARNING ANALYTICS RESEARCH

acm *In-Cooperation* sigweb SIGCHI

LAK23 Conference Proceedings

Towards Trustworthy Learning Analytics

The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York New York 10121-0701

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that was previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ACM ISBN: 978-1-4503-9573-1

LAK23 Conference Proceedings

Towards Trustworthy Learning Analytics

LAK23 Program Chairs' Welcome

We are very pleased to welcome you to the Thirteenth International Conference on Learning Analytics and Knowledge (LAK23), organized by the Society for Learning Analytics Research (SoLAR). With the aim of widening participation of the Learning Analytics (LA) community, this year's conference is held in a hybrid format (face to face and online) between March 13th and 15th.

The theme for the 13th annual LAK conference is *"Toward Trustworthy Learning Analytics."* This theme creates the opportunity to discuss several social and educational concerns that emerge from the design and implementation of LA solutions, such as privacy, fairness, and the development of learner autonomy. It invites researchers and practitioners to fully examine unintended consequences of using educational data and algorithms, including potential misuse and mis-interpretation; influence on society and education systems; ethics; privacy; transparency; and accountability. The theme also offers the opportunity for reflection on how the field can move towards a responsible education system that is established on a foundation of trust, reinforcing the use of algorithmic transparency to inform end users on how to interpret and enact LA information and recommendations.

Two excellent keynotes will address this theme across the complementary lenses of education, human-centered design, and data science. Yvonne Rogers is a Professor of Interaction Design, Director of University College London Interaction Centre (UCLIC) and Deputy Head of the Computer Science Department at UCL. Yvonne's keynote will address the theme of interactive technologies that can enhance life by augmenting and extending everyday, learning and work activities. Ken Koedinger is a Professor of Human Computer Interaction and Psychology and Director of LearnLab at Carnegie Mellon University. Ken's keynote will focus on the role of Learning Analytics in promoting Equitable Learning. A debate will also be held for the first time at the LAK conference. The debate will address the role of predictive learning analytics in addressing bias and inequity. The debate will involve a range of engaging and experienced members of our community to raise and challenge current views.

This year, we received a large number of high-quality submissions this year across the Practitioner Track, Posters and Demonstrations, Workshops and Tutorials and to the Doctoral Consortium. After undergoing a rigorous selection process, we were pleased to accept 16 Practitioner Track Papers, 36 Posters, 6 Demos, 25 Workshops (17 to be imparted in-person, 6 Online, 2 offering both online and in-person attendance), and 13 participants to the Doctoral Consortium, each of which is represented in this Companion Proceedings. We are most grateful for all the hard work by the program committee of each one of the tracks, and their insightful and constructive comments and reviews. These proceedings could not have been possible without their generous help and support.

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

We hope that LAK23 participants and other readers of these proceedings will find value in the many varied contributions to the field of LA contained within. Although there is still much to be done to understand human behavior and social values within educational settings, we believe that this conference reinforces a culture that honors the diversity of learners and their need for fair and explainable data-based interventions. We invite both researchers and practitioners to continue a proactive dialogue beyond this conference, reflecting on how LA identifies and breaks down systemic barriers for inclusion by building trust among different educational stakeholders.

Isabel Hilliger
Pontificia Universidad Católica de Chile, Chile

Hassan Khosravi
University of Queensland, Australia

Bart Rienties
Open University, United Kingdom

Shane Dawson
University of South Australia, Australia

LAK23 Conference Organization

Program Chairs
Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile
Hassan Khosravi, University of Queensland, Australia
Bart Rienties, Open University, United Kingdom
Shane Dawson, University of South Australia, Australia

Practitioner Track Chairs
Geraldine Gray, Technological University Dublin, Ireland
Dan Davis, CVS Health/Georgetown University, USA

Workshop Chairs
Antonija Mitrovic, University of Canterbury, New Zealand
Mohammad Khalil, University of Bergen, Norway
Lisa-Angelique Lim, University of Technology Sydney, Australia

Posters and Demonstrations Chairs
Ioana Jivet, DIPF, Germany
Izumi Horikoshi, Kyoto University, Japan

Doctoral Consortium Chairs
Christopher Brooks, University of Michigan, USA
Dragan Gašević , Monash University, Australia
Oleksandra Poquet, Technical University of Munich, Germany
Jelena Jovanovic, University of Belgrade, Serbia

Proceedings Chairs
Srecko Joksimovic, University of South Australia, Australia
Abhinava Barthakur, University of South Australia, Australia

Website Manager
Yancy Vance Paredes, Arizona State University, USA

Organizing Committee
Grace Lynch, Society for Learning Analytics Research, Australia
Justin Dellinger, Texas A&M University, USA
Pete Smith, UTA, USA
George Siemens, University of South Australia, Australia
Nicole Hoover, Society for Learning Analytics Research, USA

LAK22 Companion Proceedings Committee

Practitioner Track Program Committee

Samantha Ahern, University College London, United Kingdom
Theo Bakker, Vrije Universiteit Amsterdam, Netherlands
Abhinava Barthakur, University of South Australia, Australia
Alex Bowers, Columbia University, United States
Andrew Brackett, New York University, United States
Cynthia D'Angelo, University of Illinois at Urbana-Champaign, United States
Dan Davis, Delft University of Technology, Netherlands
Shane Dawson, University of South Australia, Australia
Justin T. Dellinger, Texas A&M University, United States
Ed Foster, Nottingham Trent University, United Kingdom
Stephanie Frost, University of Saskatchewan, Canada
Kerr Gardiner, Jisc, United Kingdom
Geraldine Gray, Technological University Dublin, Ireland
Maureen Guarcello, San Diego State University, United States
Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile
Martin Hlosta, The Swiss Distance University of Applied Sciences, Switzerland
Nicole Hoover, Society for Learning Analytics Research, United States
Hazel Jones, Griffith University, Australia
Hassan Khosravi, The University of Queensland, Australia
Danny Y.T. Liu, The University of Sydney, Australia
Grace Lynch, SoLAR, Australia
Leah Macfadyen, The University of British Columbia, Canada
Alison Myers, The University of British Columbia, Canada
Rachel Niemer, University of Michigan, United States
Joshua Quick, Indiana University Bloomington, United States
Bart Rienties, The Open University, United Kingdom
Daniel Spikol, University of Copenhagen, Denmark
Kp Thai, Age of Learning, United States
Craig Thompson, The University of British Columbia, Canada
Lorenzo Vigentini, The University of New South Wales, Australia
Steven Williams, University of California, Berkeley, United States
Srecko Joksimovic, University of South Australia, Australia

Posters & Demo Program Committee

Abhinava Barthakur, University of South Australia, Australia
Atezaz Ahmad, DIPF, Germany
Laia Albó, Universitat Pompeu Fabra, Spain
Sarah Alcock, The Open University, United Kingdom
Dana Alzoubi, Iowa State University, United States
Alan Berg, University of Amsterdam, Netherlands
Daniel Biedermann, German Institute for International Educational Research, Germany
Fabio Campos, New York University, United States
Bodong Chen, University of Pennsylvania, United States
Guanliang Chen, Monash University, Australia
Heeryung Choi, Massachusetts Institute of Technology, United States
Linda Corrin, Deakin University, Australia

Alexandra Cristea, Durham University, United Kingdom
Mutlu Cukurova, University College London, United Kingdom
Yiling Dai, Kyoto University, Japan
Shane Dawson, University of South Australia, Australia
Lars de Vreugd, University Medical Centre Utrecht, Netherlands
Justin T. Dellinger, Texas A&M University, United States
Yannis Dimitriadis, University of Valladolid, Spain
Vanessa Echeverria, Monash University, Australia
Erkan Er, Middle East Technical University, Spain
Elaine Farrow, The University of Edinburgh, United Kingdom
Rafael Ferreira Mello, Federal Rural University of Pernambuco, Brazil
Stephanie Frost, University of Saskatchewan, Canada
Sebastian Gombert, DIPF | Leibniz Institute for Research and Information in Education, Germany
Geraldine Gray, Technological University Dublin, Ireland
Selina Griffin, The Open University, United Kingdom
Dai Griffiths, UNIR-iTED, Spain
Julio Guerra, Universidad Austral de Chile, Chile
Areum Han, University of Luxembourg, Luxembourg
Birte Heinemann, RWTH Aachen University, Germany
Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile
Martin Hlosta, The Swiss Distance University of Applied Sciences, Switzerland
Nicole Hoover, Society for Learning Analytics Research, United States
Izumi Horikoshi, Kyoto University, Japan
Xiao Hu, The University of Hong Kong, Hong Kong
Ioana Jivet, DIPF, Germany
Srecko Joksimovic, University of South Australia, Australia
Yeonji Jung, New York University, United States
Rogers Kaliisa, University of Oslo, Norway
Onur Karademir, DIPF | Leibniz Institute for Research and Information in Education, Germany
Mohammad Khalil, University of Bergen, Norway
Hassan Khosravi, The University of Queensland, Australia
Rene Kizilcec, Cornell University, United States
Elizabeth Koh, Nanyang Technological University, Singapore
Malgorzata Korolkiewicz, University of South Australia, Australia
Vitomir Kovanovic,University of South Australia, Australia
Jakub Kužílek, CIIRC, CTU in Prague, Czechia
Charles Lang, Columbia University, United States
Alwyn Vwen Yen Lee, Nanyang Technological University, Singapore
Huiyong Li, Kyoto University, Japan
Lisa Lim, University of Technology Sydney, Australia
Grace Lynch, SoLAR, Australia
Rwitajit Majumdar, Kyoto University, Japan
Jorge Maldonado-Mahauad, Universidad de Cuenca, Pontificia Universidad Católica de, Chile
Mirko Marras, University of Cagliari, Italy
Alejandra Martínez-Monés, Universidad de Valladolid, Spain
Wannisa Matcha, Prince of Sonkla University Pattani Campus, Thailand
Lukas Menzel, "studiumdigitale, Goethe University, Frankfurt am Main", Germany
Konstantinos Michos, University of Zurich, Switzerland
Negin Mirriahi, University of South Australia, Australia
Pedro Manuel Moreno-Marcos, Universidad Carlos III de Madrid, Spain
Xavier Ochoa, New York University, United States

Alejandro Ortega-Arranz, Universidad de Valladolid, Spain
Zacharoula Papamitsiou, SINTEF Digital, Norway
Ward Peeters, Kanda University of International Studies, Japan
Malte Persike, RWTH Aachen University, Germany
Oleksandra Poquet, Technical Univesity of Munich, Germany
Sambit Praharaj, CAIS & Ruhr University Bochum, Germany
Paul Prinsloo, University of South Africa, South Africa
Eyal Rabin, The Open University of Israel, Israel
Bart Rienties, The Open University, United Kingdom
Fernando Rodriguez, University of California, Irvine, United States
José A. Ruipérez Valiente, University of Murcia, Spain
Marcel Schmitz, Zuyd Hogeschool, Netherlands
Clara Schumacher, Humboldt Universität zu Berlin, Germany
Niels Seidel, FernUniversität in Hagen, Germany
Antonette Shibani, University of Technology, Sydney, Australia
Kyosuke Takami, Kyoto University, Japan
Michelle Taub, University of Central Florida, United States
Craig Thompson, The University of British Columbia, Canada
Olga Viberg, KTH Royal Institute of Technology, Sweden
Joshua Weidlich, Deutsches Institut für Internationale Pädagogische Forschung, Germany
Sebastian Wollny, DIPF - German Institute for International Educational Research, Germany
Lixiang Yan, Monash University, Australia
Jane Yau, University of Mannheim, Germany
Gaoxia Zhu, Nanyang Technological University, Singapore

LAK23 Conference Proceedings

Towards Trustworthy Learning Analytics

Table of Contents

Practitioners Report

Selecting Distractors for Automatically Generated MCQs

Sean Shiverick[1], Clarence Dillon[1], Steve Hookway[2]
[1]ICF, [2]Charles River Analytics
sean.shiverick@icf.com, clarence.dillon@icf.com, shookway@cra.com

ABSTRACT: Selecting distractors at an appropriate level of difficulty is necessary for effective automated item generation (AIG). Measuring the similarity of key and distractors is the prevailing method for controlling the difficulty of MCQs (Kurdi et al., 2019). Representing the text of the training manual as a graph helped to identify suitable distractors by their connections to the item key. This study examined three approaches for approximating key-distractor similarity: class-based, text-based, and graph-based. Distractors in the same section of the document as the key were scored as more closely related to key than distractors in sections far from key. Conceptual structure in the source document used to generate the items provided information for mapping ontological relations between keywords and concepts.

Keywords: Assessment, Difficulty Metrics, Distractor Selection, MCQs

1 BACKGROUND

Technologies for automated item generation (AIG) require methods for selecting suitable distractors at an appropriate level of difficulty. The prevailing method for controlling MCQ difficulty is based on measuring the similarity between an item key and distractors using different measures of similarity (Kurdi et al., 2019; Liang, et al. 2018). Distractors that are semantically similar to the key are more difficult to differentiate from the key, thus increasing item difficulty. The relationships between entities and concepts can be represented as a hierarchy of classes at different levels of abstraction, with general concepts represented in higher-level categories extending to more specific concepts at lower levels (Stasaski & Hearst, 2017). The current study considered stems and distractors generated from semantic relationships (e.g., Tool-Purpose) identified in a field radio training manual. Table 1 shows a subset of chapter sections and entities that describe the functional organization of the radio components and operation. The source document structure revealed conceptual classes and subclass relations to estimate similarity without an ontology. A basic strategy for distractor selection is to choose responses from the same class or subclass as the key (e.g., siblings, cousins). Representing key-distractor relations as a graph helped identify suitable distractors by their connections to the key.

Table 1: Example User Manual Chapter Sections and Entities for Class-Based Distractor Selection

Chapter 2: Operating Instructions								
2.1 Controls, Indicators, Connectors					2.2 Operating Procedures			
2.1.1 Controls		2.1.2 Connectors			2.2.3 Programming Menu			
2.1.1.1 Keypad		2.1.3.2 Audio Key Fill Connector			2.2.3.1 Key Fill			2.2.3.2 Zeroize
2.1.1.1 Control Key	2.1.1.1 Auxiliary Control key	2.1.3.2 Audio Key Fill Connector	2.2.3.1.2 Storage Key	2.2.3.1.5 Frequency Hopset	2.2.3.1.7 Time of Day	2.2.3.1.9 Word of Day Screen	2.2.3.1.12 Remote Fill	2.2.3.2 Firmware

2 GRAPH REPRESENTATION OF DISTRACTORS

2.1 Implementation

Prominent keywords or 'n–grams' (i.e., sequence of n words) in the user manual were identified using the rapid automated keyword extraction (RAKE) algorithm (Rose et al., 2010). The document produced 3,845 keywords: 131 were identified as parts of the Tool-Purpose semantic relationship; 38 keywords were Tools. Connections between keywords in the text were graphed in Neo4J. Figure 1 presents part of the graph arranged as a dendrogram. The grey node at the top represents the source document; the orange nodes represent chapters 1 and 2. Green nodes represent chapter sections, subsections, and sub-subsections, etc. (labeled). The grey, orange, and green nodes, together, represent the structure of the information contained in the source document. Blue nodes represent sentences contained under a subsection (numbered). Yellow nodes represent response options contained within the sentence. Subsection headings provided class and subclass relations among keywords in the text.

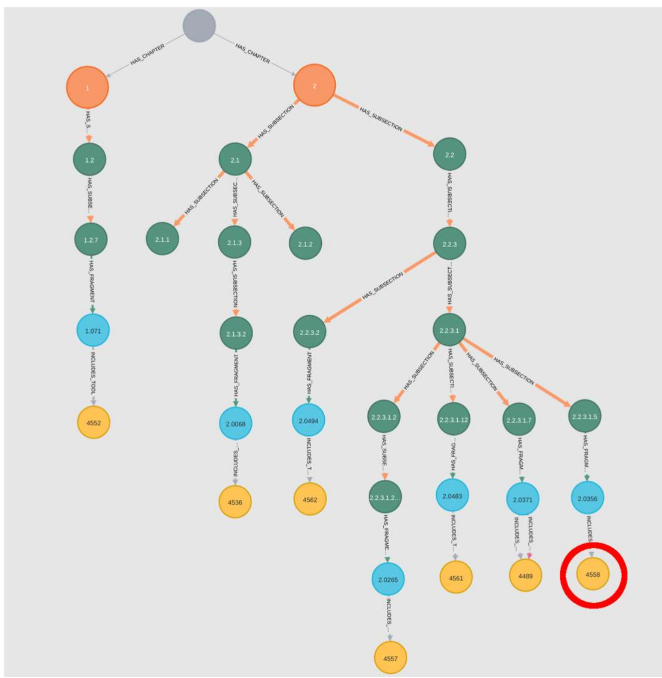


Figure 1: Dendrogram Depicting Structure Between Response Options (Key Highlighted in Red)

2.2 Findings

Distance between nodes in the graph provided a measure of conceptual relatedness among keywords. The example in Figure 1 shows an item key (circled in red) and six other Tools as distractor candidates. By "walking the graph" and counting subsection nodes, the ontological distance (i.e., similarity) between the key and potential distractors to be selected for an item was estimated by the distance between structural nodes in the graph. For example, the two distractors (yellow nodes) immediately left of the key are both three structural nodes away from the key. The third distractor lowest in the dendrogram is four structural nodes from the key. The other three distractors (right to left) are four, eight, and nine nodes away from the key. Keywords with the fewest hops between them based on the subsections (i.e., shortest distance) were more similar than keywords with more hops between them.

3 DISTRACTOR METRICS AND RATINGS

3.1 Implementation

This study tested three metrics for approximating key-distractor similarity: (1) Class relations based on chapter sections in the user manual, (2) Sentence distance between the key and each distractor, and (3) Difference in relative importance scores measured using the RAKE algorithm (Rose et al., 2010). Distractors were selected for 16 MCQs automatically generated from the user manual. The stems each stated a Purpose and the keys all described a Tool that achieved the purpose. Distractors were selected from a set of 39 tools or related concepts based on estimated similarity to the key using the three metrics. Sentence distance (i.e., number of sentences) and RAKE difference scores were calculated between all key-distractor pairs and represented in separate matrices. Distance and difference scores, ranked from smallest to largest, were used to select 5 distractors nearest to the key, and 5 distractors farthest from the key, for each stem. A subject matter expert (SME) rated 30 distractors for each stem according to, "how closely the distractor is related to key", on a 3-point scale (0=unrelated, 2=highly related). Ratings were summed for distractors nearest to the key and distractors farthest from the key, creating aggregated scores on 10-point scale. It was hypothesized that distractors nearest to key, as measured by chapter section, sentence distance, or RAKE difference, would be rated as more closely related to the key than distractors farthest from key. Figure 2 shows a boxplot of the aggregated scores for the nearest and farthest distractors identified by these metrics.

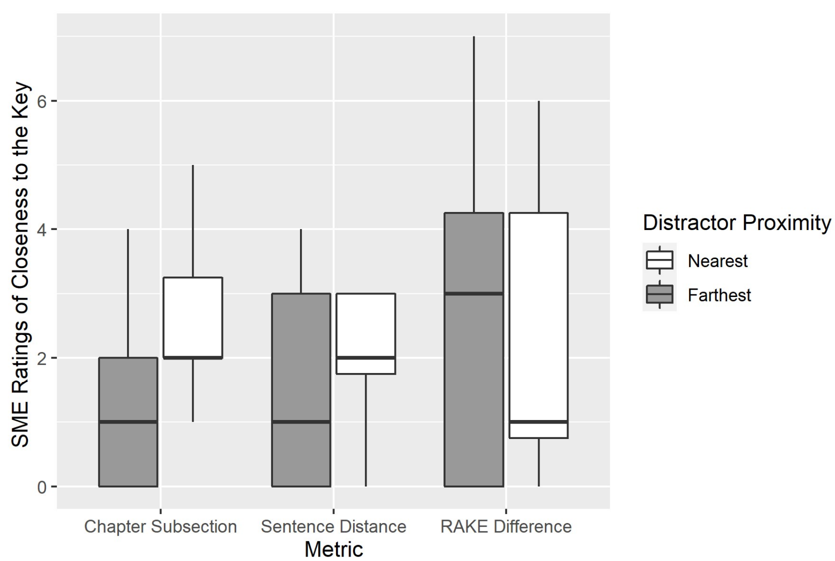


Figure 2: Boxplot of SME Ratings of Distractor Relation to the Key by Distance and Metric

3.2 Findings

A mixed ANOVA conducted on the distractor rating scores yielded a Distance by Metric interaction, $F(2, 60) = 3.93$, $p = .025$ ($\eta^2 = 0.04$). The effect of Distance was significant for distractors selected using the class-based approach ($p = .02$); as predicted, the SME scored the distractors identified as nearest to the key by document section as more closely related to the key than distractors farthest from key. The effect of Distance was marginally significant for distractors selected by sentence distance ($p = .06$).

The SME also scored some distractors farthest from the key in text distance as being related to the key, which was expected because some concepts were described in different sections of the manual. Third, there was no effect of Distance on ratings for the RAKE difference scores (p = .61), which showed the SME did not differentiate the key and distractors by the relative importance of keywords.

4 CONCLUSIONS

Automated technologies capable of generating large numbers of stems and response options necessitate a method for selecting distractors at an appropriate difficulty level. Findings from the graph representation and distractor ratings indicated that the structure of the source document provided information about class and subclass relations between keywords in the document that were used to approximate key-distractor similarity. Representing the text of the user manual as a graph revealed connections to the key to inform the selection of distractors at varying levels of difficulty. Distractors from the same document section as the key were more closely related to the key than distractors in more distant sections. Furthermore, the relations among concepts in the user manual were somewhat analogous to relationships in a domain ontology. The authors previously reported that graph representations can be used to estimate characteristics of an ontology (Shiverick, Dillon, Smith & Harvey, 2021). The estimation of ontological relationships yielded insights about distractors that were conceptually closer to the key, and therefore more difficult to differentiate from the key when selected for use in an assessment item. This approach may be useful for developing knowledge assessments in focused training courses when an existing ontology is not available.

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

REFERENCES

Kurdi, G., Leo, J., Matentzoglu, N., Parsia, B., Sattler, U., Al-Emari, S. (2019). A systematic review of automatic question generation for educational purposes. *Int J of AI in Ed, 30*; 121-204.

Liang, C., Yang, X., Neisarg D., Wham D., Pursel, B., & Giles C. L. (2018). Distractor generation for multiple choice questions using learning to rank. *In Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 284–290.

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In M.W. Berry and J. Kogan (Eds.), *Text Mining: Applications and Theory*, pp. 1-20. Wiley Online Library. https://doi.org/10.1002/9780470689646.ch1.

Shiverick, S., Dillon, C., Smith, M., & Harvey, J. (2021). Estimating ontology-based difficulty metrics for automated item generation. In *Companion Proceedings of the 11th International Conference on Learning Analytics* (LAK21), pp. 27-30. Society for Learning Analytics Research (SoLAR).

Stasaski, K., & Hearst, M. A. (2017). Multiple choice question generation utilizing an ontology. In *12th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 303–312.

3 Years of GOAL project in Public School: Leveraging Learning & Smartwatch Logs for Self-directed Learning

Rwitajit Majumdar, Yuanyuan Yang, Huiyong Li, Brendan Flanagan & Hiroaki Ogata
Kyoto University, Japan
majumdar.rwitajit.4a@kyoto-u.ac.jp

ABSTRACT: The GOAL project aimed to collect and synchronize learners' data from physical activity sensors as well as online learning tools to design data-driven services. We extend the potential of learning tools interoperability (LTI) protocol to link physical activity and sensor data from smartwatch platforms. Our primary purpose is to provide this synchronized self-data to the learners for reflection and promoting self-directed learning habits. The project is partially supported by multiple national funding and implemented at scale at a combined public junior high and high school since the summer of 2019. Across the three years more than 1300 users have used the different services built on GOAL. We collected 5,92,599 daily learning and physical activity logs. Further, 1,72,674 logs of user interaction within the GOAL application were collected to identify self-directed behaviors. This paper overviews the research journey of GOAL over the last three years highlighting the implementation challenges and how they were overcome. As an ongoing project it discusses the potential of anonymous yet linked multi-attribute learner data and its implication for research and development in the field of learning analytics.

Keywords: GOAL, Smartwatch, Learning Logs, Self-directed Learning, DAPER model, LEAF

1 BACKGROUND

Collecting and synchronizing multiple attributes of learners remains a challenge in the field of learning analytics. On the other hand, off-the-shelf wearable technologies such as smart bands and watches have made automatic logging of users' physical activities and physiological data more affordable and easier. However, the datasets from multiple sensors and the learning logs from the e-learning systems are often collected in separate data silos and limited to use for research purposes only. While earlier research has discussed the potential of reflection in learning using quantified self-approaches [1,2], the technology infrastructure for that was still rare. The GOAL project aims to synthesize multi-source data of learners and create services for learners and teachers as the end-users. The data-driven services aim to introduce a paradigm of supporting executing self-direction skills (SDS) of analyzing, planning, monitoring, and reflecting on practice from daily learning lifestyle logs.

This project is partially supported by multiple national funding and will continue till April 2025. First, we piloted a developed mobile application in a university course in 2018. Based on that, a web version was developed to be rolled out at scale in the school context in 2019. We collaborated with a public city school and linked their learning system data to the GOAL infrastructure. Smartwatches were set up and distributed in the junior high grades in that school, and students could freely use the device. Currently, all three grades of junior high school and two grades of high school use the GOAL system.

2 GOAL TECHNOLOGY MODEL

The GOAL technology architecture follows Learning Tools Interoperability (LTI) protocol to link to the existing learning management system (LMS). The server component synchronizes the data from multiple systems through APIs. The client can be accessed through a web browser that is also packaged in iOS and Android apps. The server-client architecture is linked through REST API. The interactions on the client end are also tracked and stored in the GOAL database. The activity and interaction data are processed to create GOAL's user model. Figure 1 presents the system architecture and the landing page interface in GOAL that supports a five-phase process model, DAPER [4].

Figure 1a: The GOAL system architecture 1b: User interface and activities in GOAL

3 CHALLENGES & LEARNINGS ACROSS PHASES OF IMPLEMENTATION

Designing daily learning support with multi-source data: For a regular learner at a public school, daily self-directed learning activities are in different contexts, such as extensive reading practice in English, solving practice problems for weekly mathematics quizzes, etc. These learning episodes are distributed in space (within a classroom or outside), time (happening synchronously or asynchronously with other classmates), and medium (can be tracked online or happens offline). Planning for holistic support in such a context is challenging due to the lack of integrated data. We bridge that gap by utilizing tools linked with LTI to collect the online data. For offline data, GOAL provides forms to collect the data. We found starting with the learning context, where data is automatically synchronized in the system, and the teachers can guide the activity in a synchronized classroom setting, helped the junior high school students to get familiar with the DAPER process. Hence instead of a generic reading activity tracking, setting up specific extensive reading (ER) tasks in the e-reader as part of an English language course helps to set the context of automatically synchronizing reading behavior data in GOAL. Recommendation modules such as eBook recommender for ER being a part of LEAF could be connected to the GOAL client within the ER dashboard. It provided the students' scope of executing their self-direction skills supported with the GOAL system in that specific context both within the class and during the vacation period. The longitudinal study indicated a positive effect on student's motivation and performance outcomes [3].

Expanding Self-direction skill practices to daily lifestyle: To initiate SDS in the daily lifestyle, we set up 390 smartwatches for the students and synchronized their accounts to the tablet used in their classroom. At the end of each year, the graduating batch returned the device, and it was made sure the data and the accounts were deleted from the service providers and freshly set up for the incoming batch. A step-wise protocol for setting up smartwatch devices and synchronizing the application to the students' tablets made the process smoother for the school authorities to prepare for the setup from the second year, along with the assistance of the GOAL team members. A targeted user manual in print and video format also helped the students and teachers in onboarding. Further specific campaigns were introduced where the students participated in executing their self-direction skills in their daily steps taken and sleep. Introducing activity specific dashboard helped to aggregate the tasks that the users need to execute in the DAPER flow for a specific activity.

Adaptive scaffolds for SDL skill acquisition: While the basic workflow of the GOAL system is based on the DAPER model, we built the system in modular form with a scope of augmenting additional data-driven services. For instance, we added a process recommendation function based on the learner model created in the system. Figure 2 provides a data flow of the adaptive support strategy used in the GOAL system. Following a standard protocol such as xAPI to log interaction and synchronize data within GOAL helped it maintain interoperability with other LEAF components.

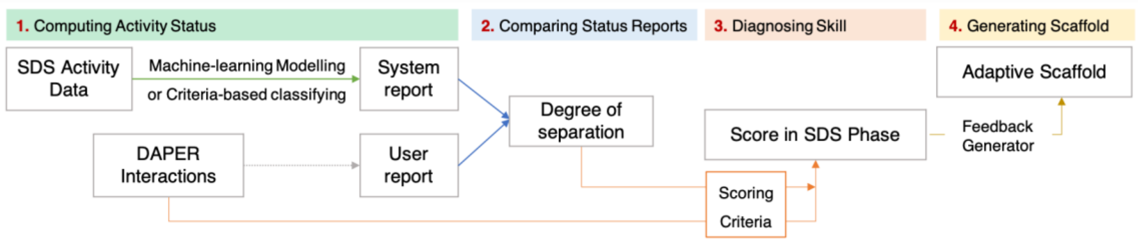


Figure 2: A model for adaptive scaffolding in GOAL system

Cooperation of the teachers and co-design efforts: The cooperation of the teachers, the school management, and the education board was crucial at every implementation step. At the onset, they had to approve the ethical implementation plan of the project. The school's coordinating teacher shared the project information with the parents of the students. After being aware of the project's scope, the type of data collected, and the functions created with them, the parents had to consent for their ward to participate in the studies and use the results for academic reporting. It was also essential to discuss with the teachers the actual students' context in which they can be supported for self-directed activities. Over this period, we co-designed activities with teachers for English, mathematics, and physical education courses at school.

4 USAGE TILL NOW AND THE FUTURE PLANS

Figure 3 presents the accumulated activity logs and the GOAL system's self-direction skill interaction logs. We can see that most interactions are still in the learning context. While data are collected from the smartwatch activity context, it is still underutilized by the students. One of the reasons is the limited school hours during the covid-19 period, which did not allow coordinating physical activity events for developing self-directed skills. We plan to explore how GOAL can assist teachers in tracking students' self-directed competency, which is now part of the national educational policy.

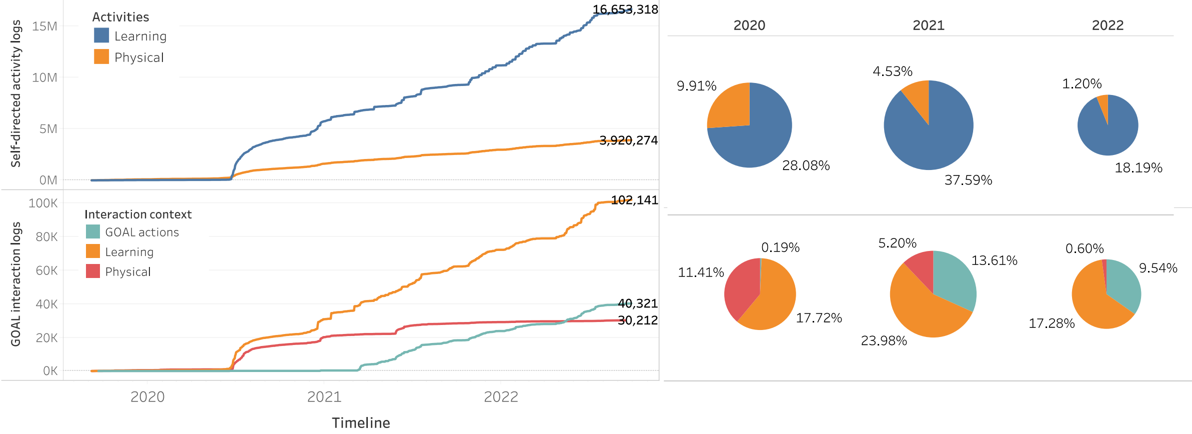


Figure 3: GOAL accumulated logs (from Sep 6, 2019 to Sep 28, 2022)

ACKNOWLEDGEMENTS

This research was partially supported by JSPS KAKENHI 20K20131, 22H03902, 20H01722 and NEDO JPNP18013, JPNP20006.

REFERENCES

[1] Rivera-Pelayo, V., Zacharias, V., Müller, L., & Braun, S. (2012). Applying quantified self-approaches to support reflective learning. 2nd LAK pp. 111-114. doi.org/10.1145/2330601.2330631

[2] D. Boud, R. Keogh, and D. Walker. Reflection: Turning Experience into Learning, chapter Promoting Reflection in Learning: a Model., pages 18–40 Routledge Falmer, New York, 1985.

[3] Li H., Majumdar R., Chen M.R.A. and Ogata H. (2021) Goal-Oriented Active Learning (GOAL) System to Promote Reading Engagement, Self-Directed Learning Behavior, and Motivation in Extensive Reading. Computers and Education. doi.org/10.1016/j.compedu.2021.104239

[4] Majumdar R., Li H., Yang Y.Y., Akçapınar G., Flanagan B. and Ogata H. (2019) Adaptive Support for Acquisition of Self-direction Skills using Learning and Health Data, Procs. of 19th IEEE ICALT pp.54-56, Maceió, Brazil July 2019.

Orchestrating Virtual Reality Simulations in Undergraduate Nursing Education

Mamta Shah, PhD
Elsevier
m.shah@elsevier.com

Jennifer Wills-Savoia, DNP
University of St. Francis
JWillsSavoia@stfrancis.edu

ABSTRACT: Nursing faculty complement clinical experiences with simulations to expose undergraduate students to a variety of clinical situations, facilitate theory-practice integration, and cultivate a range of competencies. The adoption of immersive virtual reality (VR) simulations in nursing programs has risen recently. However, little is known about the design of learning analytics tools in VR systems for nursing education, and how faculty utilize them to meet their clinical education goals. Thus, in this practitioner-corporate track report, the authors unpack the design and implementation of the SimX moderator tool–a multimodal interface used to select and facilitate scenarios–in Simulation Learning System with Virtual Reality (SLS with VR) using categories for technologies for classroom orchestration (Dillenbourg & Jermann, 2010). Insights from Fall 2020 and 2021 indicate that the moderator tool affords nursing faculty the agency to prioritize student learning goals, personalize instruction within a certain range, adjust their scaffolding just-in-time, and maintain the realism of clinical settings.

Keywords: virtual reality, learning analytics tool, nursing education, classroom orchestration, simulation

1. Introduction and Background

The growing number of nursing schools in the United States and increasing student enrollment has left many institutions struggling to obtain clinical partnerships. This issue is juxtaposed with a rapid shift in patient demographics and social determinants of health, digital transformation in healthcare and education, and reports indicating that new graduates are not being sufficiently prepared for participation in clinical settings (Kavanagh & Sharpnack, 2021). Over the last two decades, nursing programs have reliably used simulations to complement clinical experiences. More recently, the adoption of immersive virtual reality (VR) simulations has gained momentum. Studies are increasingly demonstrating the positive impact of this high-fidelity modality on nursing students' cognition and psychomotor skills (Choi et al., 2022). Furthermore, reports illustrating the technological affordances of VR (e.g., immersion, cost-effectiveness) have catalyzed acceptance among nursing faculty and students. However, there is a dearth of reports that unpack the design and implementation of learning analytics tools that faculty use to facilitate simulation experiences in nursing programs (Fernandez-Nieto et al., 2022). Focusing on this technological-pedagogical gap is the goal of this practitioner-corporate track paper.

In what follows, we (*Wills-Savoia*, practitioner author-director of clinical and simulation learning at University of St. Francis and *Shah*, corporate author- learning scientist at Elsevier) report findings from a descriptive case study capturing nursing faculty insights about the moderator tool (developed by SimX) in Elsevier's Simulation Learning System with Virtual Reality (SLS with VR). We use Dillenbourg and Jermann's (2010) design categories for technologies for classroom orchestration. As a design metaphor, orchestration provides a lens to understand the effectiveness of learning analytics tools from teachers' perspectives. Thus, to situate the reader, first we introduce SLS with VR and the anatomy of its moderator tool. This is followed by a description of SLS with VR implementation at a private nursing college in mid-western United States. Results are organized by the nursing faculty (practitioner author and partnering faculty) insights about the moderator tool's affordance for teacher-centrism, cross-plane integration, sequentiality, time management and physicality (Dillenbourg & Jermann, 2010). We discuss findings in the context of extant literature and conclude with implications for future inquiry.

2. SLS with VR

SLS with VR enables nursing schools to provide undergraduate students with immersive clinical experiences alongside traditional simulation experiences. Faculty have a choice of 100 scenarios and associated student-facing activities and faculty resources across multiple content areas in nursing. The moderator tool enables faculty to (a) select scenarios, (b) orient students to the clinical environment and possible actions in VR; (c) introduce virtual characters, situational distractions, control patient and other virtual character speech and actions, (d) monitor student participation and patient health, (e) provide just-in-time support; and (f) obtain an end-of-scenario report of interventions performed by the learners. Nursing faculty navigate and choose from the following features in the moderator tool while facilitating scenarios - 1. Orders & actions pane, 2. Dialog tab, 3. Monitor tab, 4. Required actions pane, 5. VR view pane, 6. State map tab, 7. Description tab, 8. Settings button, 9. Screen recording button. This video provides a brief demonstration of SLS with VR, including the use of the moderator tool and the instructor view it affords.

3. Description of Implementation

Two faculty (practitioner author whose expertise is in Pediatrics, and an Obstetrics and Fundamentals expert) utilized SLS with VR with a group of 50 undergraduate prelicensure nursing students initially during Fundamentals–their first clinical course (Fall 2020) and then in Obstetrics and Pediatrics (Fall 2021)–their third clinical course. Both implementations were 8 weeks long. Typically, in each simulation session, two students participated in the patient scenario while two others observed. At the same time, four students completed pre-simulation activities such as concept maps in the waiting room before it was their turn to role-play. The second implementation (Obstetrics and Pediatrics) included the same fifty students, plus two additional students who had no prior experience with VR. Faculty that led SLS with VR simulations with the first group also led them with the second; this aided in consistency in pre-briefing, facilitation and debriefing. During both implementations, the corporate author and her team provided onboarding and technical support, and engaged in remote observations of the SLS with VR sessions.

4. Results

In Fall 2020, faculty used one scenario featuring a patient with diabetes and a cellulitis wound. Students had to assess the wound, engage in empathetic communication and provide education to the patient and their family, report to the provider, and administer medications. However, at that time, many of the students had not yet gained practical experience in the clinical setting. This inexperience was reflected in students' struggle with performing most tasks in the simulation and completing interventions. Instead, the participants focused on communicating with the patient and family. In Fall 2021, this same group used two scenarios. The first was a patient with preeclampsia and the second was a pediatric patient with sickle cell disease. At this time, the students had spent approximately 150 hours in the clinical setting. Their growing competence was reflected in their ability to complete their assessments, determine necessary interventions, implement those interventions and reevaluate their care, while they communicated with the patient and family. The implementation of SLS with VR in the two clinical courses helped the nursing faculty complement students' clinical experiences. The moderator tool provided them the means to orchestrate scenarios for nurturing students' practice readiness and observing clinical judgment improvements during the simulation sessions.

Dillenbourg and Jermann's (2010) first design category states that technologies designed for classroom orchestration should be teacher centric. They should grant teachers leadership, flexibility and control in order to meet their instructional goals. Overall, the moderator tool allowed the nursing faculty to "drive the bus" and choose specific milestones. Although multiple tabs and panes were available to provide a real-time view of how a scenario was unfolding, the faculty could decide what they wanted to prioritize for their students for each simulation session and scenario. The practitioner author and her colleague noted using the dialog tab frequently to prompt and respond to students through different characters in the scenario, steer their attention towards specific aspects, and encourage critical thinking about patient-centered communication, teamwork and collaboration. Using the moderator tool, faculty maintained a certain level of control; however, as in the real-world, patient outcomes in the virtual world depended on student interventions.

The second category is cross-plane integration; Dillenbourg and Jermann (2010) suggest that tools should facilitate students' engagement in the curricula at multiple levels. The moderator tool provided an array of functions and multi-modal feedback mechanisms to deepen (individual and dyad) students' participation in a scenario. For instance, as students progressed in their program, nursing faculty prioritized multiple learning goals including peer collaboration, assessing patient condition, performing interventions, demonstrating cultural sensitivity and adopting safety measures. Sequentiality, the third design category, is characterized by the extent to which a tool allows teachers to expect a degree of linearity and continuity, and introduce drama in a learning situation when needed. Across scenarios and semesters, the moderator tool allowed nursing faculty to guide student participation through experiences of patient assessment, intervention, and communication. This consistency allowed them to observe students' growing competency and knowledge gaps. Introducing characters (e.g. a call from a provider seeking an assessment report) and situations (e.g., making the virtual parent walk up to the student role playing nurse and ask them why the child is hurting) provided a way to make a scenario mimic the characteristics of a dynamic clinical setting.

Time management and physicality are the final design categories; they are self-explanatory. Nursing faculty believed the moderator tool was most useful in these categories. Simulation sessions were preceded by lengthy and large group lectures on specific content. SLS with VR scenarios also complemented the lectures allowing students to apply theoretical knowledge in short durations and smaller groups, and allowing faculty to facilitate reflection in and on action during the preparation/pre-briefing, scenario and debriefing phases. The moderator tool enabled the faculty to maintain a participant observer-like presence during the simulation; they were able to watch and scaffold their students' communications and actions in simulated clinical settings and most importantly get a first person view of what the students were seeing too. Nursing faculty rarely get this perspective in clinical settings.

5. Discussion and Implications

“Orchestration tools are based on the idea of capturing, analyzing, and visualizing student activities during class time and feeding them back to teachers to facilitate real time monitoring and support of students” (van Leeuwen et al., 2018, p. 1227). The moderator tool in Simulation Learning System with Virtual Reality (SLS with VR) affords these technological and pedagogical functions for nursing educators interested in using VR simulations to facilitate clinical readiness (Kavanagh & Sharpnack, 2021). According to Dillenbourg and Jermann (2010), teachers translate the design of orchestration technologies in the context of their practice (Dillenbourg & Jermann, 2010). In this study, nursing faculty reported using functions of the moderator tool to prioritize specific clinical competencies and personalize instruction based on students’ clinical experiences over two semesters (Fall 2020 and Fall 2021).

Future work should continue examining how nursing faculty orchestrate simulation experiences using learning analytics tools in VR systems in a variety of programs and disciplines. Pursuing this direction should include identifying best practices and challenges, generating opportunities for design enhancements, and assessing impact on students’ preparedness for clinical practice. An endeavor of this nature would be of mutual benefit to practitioners, researchers, and industry.

REFERENCES

Choi, J., Thompson, C. E., Choi, J., Waddill, C. B., & Choi, S. (2022). Effectiveness of immersive virtual reality in nursing education: Systematic review. *Nurse Educator*, *47*(3), E57-E61.

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

Fernandez-Nieto, G. M., Kitto, K., Buckingham Shum, S., & Martinez-Maldonado, R. (2022). Beyond the learning analytics dashboard: Alternative ways to communicate student data insights combining visualization, narrative and storytelling. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 219-229).

Kavanagh, J. M., & Sharpnack, P. A. (2021). Crisis in competency: A defining moment in nursing education. *Online J Issues Nurs*, *26*(1).

van Leeuwen, A., Rummel, N...et al. (2018). Orchestration tools for teachers in the context of individual and collaborative learning: What information do teachers need and what do they do with it?. In *13th International Conference of the Learning Sciences (ICLS) 2018, Volume 2. London, UK: International Society of the Learning Sciences* (1227-1234).

Learning Analytics Dashboards for Professional Academic Advisors

Juan-Claude Lemmens, PhD
University of Pretoria
jlemmens@up.ac.za

Hestie Byles, PhD
University of Pretoria
hestie.byles@up.ac.za

ABSTRACT: The format of the presentation is a practitioner's presentation on the effective use of a Learning Analytics Dashboard (LAD) by professional academic advisors to support undergraduate students at a large research-intensive university in South Africa. Academic advisors provide multifaceted support in the life-cycle of students. This includes support in both academic and non-academic domains. In the past, these advisors had minimal information about the students that consulted them. Moreover, the advisors were not in a position to proactively identify students in need of academic support early in the semester. Previous LADs provided aggregated risk indicators that were not very useful to identify the challenges students may face with specific subjects or courses. The new LAD was created using student activity data from Blackboard Learn as well as their formative assessment results per month. Students' demographic data, intervention attendance, and final semester results are also included. This was done to provide advisors with a profile of students' engagement and academic performance over time, as this information is not presented in any of Blackboard Learn standard reports. The new LAD is an attempt to provide the advisors with the necessary information to proactively support students academically through various interventions aimed at assisting students to manage their academic careers.

Keywords: Learning Analytics Dashboards, Student Support, Academic Advisors, Decision Support.

1 BACKGROUND

Learning Analytics in South Africa is an underdeveloped research field and the application of student data is mainly focused on Academic Analytics for reporting and strategic planning purposes (Lemmens and Henn, 2013, Prinsloo and Kaliisa, 2022). While there is a large-scale adoption of learning management systems across the 26 public higher education institutions in South Africa, each institution differs in its use of the LMS' analytics and reporting tools, and using these tools to improve learning in the classroom is often limited to the monitoring of clickstream activity.

At one large research-intensive public university in South Africa, the adoption and use of Blackboard Learn has a long history, including the use of Blackboard Analytics for Learn and recently the introduction of Blackboard Engage and other LMS reporting tools. While these reports and analytics are available for teaching staff to improve their teaching and to facilitate student learning in a single course, there are varying degrees of use among academic teaching staff. On the other hand, Blackboard Engage provides a view of students' academic performance across their subjects and is available to our teaching staff and Academic Advisors to identify and support students that have challenges over a number of subjects in a semester (2018-2019). Albeit, the academic teaching staff were not interested in these dashboards because they were not directly concerned about the academic performance of students outside of the subject/s they teach. The information provided in these dashboards were, unfortunately, not of much help to the academic advisors either.

Academic advisors perform multifaceted functions in the experiences of students, including functions in both academic and non-academic domains, which are broader than the role of tutors and/or mentors (Medernach, 2018). The nature of this direction might be to inform, suggest, counsel, discipline, coach, mentor, or even teach (Kuhn, 2008). The purpose of academic advising was to broaden the support to students and improve their learning through targeted interventions, which place them in an ideal position to support students across subjects.

2 DESCRIPTION OF IMPLEMENTATION

Initially, Pyramid Analytics was chosen for the development of LADs because it was integrated with Blackboard Learn data. Licenses were obtained for 400 staff members, including academic management staff, heads of departments, and academic advisors. The academic advisors, however, experienced technical difficulties with some functionalities on their dashboards. It became clear that advanced programming skills, which were not available at the time, would be needed. Consequently, in 2021, the decision was made to move to Tableau with the aim of implementation in 2022.

There is a total of 26 academic advisors for a total undergraduate student body of over thirty-five thousand students spread across nine faculties. A faculty in the South African context is similar to a College or School in some higher education institutions. These academic advisors are assigned to a specific Faculty where they provide support to students. This assignment of advisors amounts to roughly three academic advisors for large faculties and one academic advisor for the smaller ones.

The unit for Higher Education Research at the center for teaching and learning at the university is responsible for harvesting the student information from three different warehouses, joining the data, and updating the dashboards on a monthly basis.

3 RESULTS

An evaluation of the support from Academic Advisors was conducted using data from the 2020 and 2021 first-year student cohort. The evaluation showed that students who attended at least one or two academic advising appointments with an academic advisor after indicating a need for advising from the onset of the first academic year, performed academically better than those students that indicated a need for advising but did not attend an appointment. This finding was proved to be significant, using multinomial regression analysis, with Levene’s Test for Equality of Variance.

Early intervention in the form of academic advising seems to have positive academic outcomes for students, when taking the above results into consideration. Drake (2011) agrees that student retention can be linked to “solid” academic advising. This underscores the need for timely and comprehensive data of students’ engagement and academic progress in subjects over the duration of a semester and for this data to be available to academic advisors.

The academic advisors were included in the design process of the new Tableau dashboard from the onset. The academic advisors requested the integration of the student case management system into the Tableau dashboards. This gave advisors a data of the students they consulted, as they only need to have the dashboard open and not the case management system too. In addition, and unique to universities in South Africa, the advisors also requested data to identify floundering students within challenging subjects (i.e. gateway courses). Thus, identifying and supporting students that are registered for challenging subjects based on their formative assessment marks in the course. The inclusion of the academic advisors in the design approach proved to be an important factor for the utilization of the dashboards and user satisfaction.

4 DISCUSSION

The key challenge with the implementation of the Tableau Learning Analytics dashboard for the academic advisors was to join the data from various data warehouses. Among others, the Blackboard Learn data showing students' use of the tools and formative assessment had to be extracted via Pyramid Analytics on a monthly basis and joined using a Tableau Prep file. Although it is not highly technical it is time-consuming to make the joins with new data and is error-prone. This is especially prevalent when multiple joins and various table calculations and transformations are used. While the initial Tableau dashboard included the additional data mentioned above, it still mimicked the Pyramid Analytics Dashboard. Feedback from the advisors led to further customizations in the design and elaboration of the dashboard. In addition, it is also necessary to include regular training and to demonstrate any new developments to advisors, to ensure the effective use of a system such as the LAD.

Below is feedback from our Academic Advisors on the Tableau Learning Analytics Dashboard, which included the key challenges, notes for practice/ use of the system and the need for further developments.

4.1 Key challenges advisors experienced with the Tableau Dashboard.

> "I do not have access to the current (real time) semester mark, this only becomes available after semester." The main concern when dealing with student success, is being proactive and reaching out to students before they become at risk of failing. Currently, the LAD provides early alert data based on activity on the LMS and formative marks, where available. It is updated monthly and therefore does not provide real time data. Even though the activity and formative assessment data is made available earlier, formative data remains a challenge as it gets processed centrally first and is not available in Blackboard in real time".

> "Getting through all the information it [LAD] provides to only the information I need can be a challenge, but not an insurmountable one." The LAD houses data of all 36000 undergraduate students and advisors had to be trained to navigate through the vast amount of data to drill down to their own group of students. This is still a learning curve for many advisors who do not feel comfortable using data and do not know how to navigate dashboards".

4.2 What do advisors use the dashboard for.

As mentioned earlier, the main goal with the LAD is to act like an early warning system. As such, advisors use the LAD to track student performance with the aim of reaching out to students and offering support as proactively as is possible. From the feedback of the advisors, it appears that they work to achieve the main goal:

> "For tracking at-risk students and supporting them in terms of modules [subjects] that they struggle with and checking performance of students who want to transfer to the Faculty of Education".

> "I used the dashboard to first identify at-risk students based on their first semester exam academic performance. Secondly, I used the dashboard to follow up on their performance throughout the semester. Lastly I used the dashboard to check the performance of walk-in students and students who have been referred by other staff".

> "I used Tableau to track students who were at risk of failing some modules. The following month I would check if there is improvement in those particular modules, if not, I then would invite those students for face-to-face discussions".

4.3 Notes for practice going forward

Advisors are still finding their feet, so to speak, on the dashboards and are still trying to unearth all the potential it has. Below are some of the things they would still like to try or would want to do differently in the future:

> "Performance tracking remain key for now. In[the] future [it] maybe built up [may have] enough data to influence student degree choice based on matric [Grade 12] math marks in particular."

> "If it can be a reflection of the students current academic standing, then it would be more useful for early intervention."

> "I want to attempt to monitor the progress of students who fell in the range of 0-49% … using the Dashboard."

REFERENCES

Drake, J. K. (2011). The role of academic advising in student retention and persistence. *About Campus,* 16(3), 8-12.

Kuhn, T. (2008). Historical Foundations of Academic Advising. In Gordon, Habley and Grites. Academic Advising: A Comprehensive Campus Process. San Francisco: Jossey-Bass.

Lemmens, J. C., & Henn, M. (2016). Learning analytics: A South African higher education perspective. In J. Botha & N. J.Muller (Eds.). Institutional Research in South African Higher Education (pp. 231–253). Stellenbosch: SUN PReSS. https://doi.org/10.18820/9781928357186/12

Medernach, R. (2018). Intrusive Advising and Student Success. Advising and Student Affairs, Learning. https://richardmedernach.wordpress.com/2018/04/08/intrusive-advising-and-student-success/

Prinsloo, Pl & Kaliisa, R. (2022). Learning Analytics on the African Continent: An Emerging Research Focus and Practice. Journal of Learning Analytics. 1-18. https://doi.org/10.18608/jla.2022.7539

Pause for effect: academic staff survey results of a long-standing learning analytics implementation and implications for future practice

Author: David Fulcher
University of Wollongong
dfulcher@uow.edu.au

Author: Juliana Cavalieri Goncalves Peloche
University of Wollongong
jpeloche@uow.edu.au

Author: Soroush Homayounpour
University of Wollongong
soroush@uow.edu.au

ABSTRACT: Top-down approaches to Learning Analytics (LA) in higher education often seek to "democratize the data" and thus maximize its utility across the institution. Without close collaboration with stakeholders these initiatives run the risk of being disconnected from everyday teaching and learning practices. This paper provides a summary of responses for an operational survey of academics run across multiple academic sessions at a university with a long running LA implementation. Considerations for future practice arising from these results are consistent with the themes of trust, validity, and transparency that emerge from the literature, especially when ambitious plans are in place to widen and deepen both teacher and student use of LA to help drive student success.

Keywords: "learning analytics", "higher education", "implementation"

1 BACKGROUND

Two key promises LA makes are consistent with those made under the broad umbrella of Artificial Intelligence (AI): solve problems and save time (Kandlhofer et al., 2016; Nazaretsky et al., 2022). LA intends to provide an "understanding [of] the complex learning processes and learning outputs using a multi-disciplinary combination of computer-science, educational psychology, engineering, and learning sciences" (Rienties et al., 2020). In reality, there is often an enduring pressure on higher education institutions to harness LA as a tool for reducing student attrition (Shah et al., 2021). The promise of using LA to enhance the teaching and learning process is challenged by teachers due to issues of trust, validity, and transparency (Nazaretsky et al., 2022; Tsai et al., 2021; van Leeuwen et al., 2022). To understand how LA can be further integrated and increase the likelihood of delivering on the potential benefits it is necessary to reflect on how the systems have been used and perceived.

2 DESCRIPTION OF IMPLEMENTATION

The site of this implementation is a regional university with over 35,000 students representing more than 175 nationalities. The multi-faceted approach to the LA implementation at this institution commenced in 2014. A key piece of this work is a series of reports provided by the central LA team during the academic session to unit coordinators (academic staff with specialist knowledge responsible for the delivery of the unit). These reports help to identify students who may be at risk of failure or withdrawal from enrolled units and provide insights related to unit engagement, progress, and assessments. 98% of undergraduates are in at least one unit receiving this form of support.

A survey is sent to unit coordinators at the end of the academic session seeking feedback on this experience. The offer to participate in the survey accompanies the final report sent to unit coordinators following the release of academic results to students. The survey instrument consists of seven questions, four of which use a five-point Likert scale. A fifth question targets the use of analytics tools within the Learning Management System (LMS). Two open-ended questions round out the survey that ask about suggestions for improvement and any other comments respondents want to make. The intent with the design of this survey is to minimize disruption to academics and thus try and maximize survey response rates.

3 SURVEY RESULTS

The survey has now run across seven semesters and the response rate hovers around 15%. Overall, most survey respondents agreed that the LA reports provided them with useful information to support students. As more units receive these reports, the agreement rate has stabilized for this question (strongly agree, somewhat agree) to 83% in Semester 2 2022. The agreement rate on LA saving time in unit delivery has remained consistent over the past few academic sessions, with a result of 45% in Semester 2 2022. There was a moderate agreement rate (65% for Semester 2 2022) hat new insights were found from the LA reports unit coordinators would otherwise not be aware of. A similar result was found for LA reports providing clear and helpful guidance for how to act with identified students (68% for Semester 2 2022). Most respondents (81% for Semester 2 2022) indicated familiarity with the built-in analytics features of the LMS.

4 DISCUSSION

While efforts are made to minimize disruption to academics in completing the survey, responses suggest that not all staff that receive these reports use them to an extent that generates conviction in their opinions about the experience. At least enough to want to fill out the survey. 22% of academics receiving these reports have done so for at least two consecutive years. So, one possible explanation for the low response rate could be that people do not feel the need to complete it each semester. Other explanations for this are being explored, including the issues of transparency and trust as found in the literature (Nazaretsky et al., 2022; Tsai et al., 2021; van Leeuwen et al., 2022). The LA reports are sent via a central unit and are designed with certain assumptions in mind. Co-design approaches with academic stakeholders are underway to address potential drawbacks with such a top-down LA operating model, which may also help address possible issues of transparency and trust (Kaliisa et al., 2021; Li et al., 2021). An implication here is that what matters most for one unit coordinator may not be of much interest to another unit coordinator. So, it needs to extend beyond co-design to include collaborative priority setting. This way, a “test and learn” approach can be rolled out for new capabilities, establishing success criteria with regular checkpoints for measuring performance against these criteria. Work has commenced at this site using this approach for a new interactive dashboard tool for unit coordinators, with the design refined over multiple iterations based on both anecdotal and focus group feedback.

Positive survey responses about how LA provides useful information to support students can be connected to users' trust in the tool to understand students’ situation and make informed decisions. And yet, the standard nature of the current LA reports makes it difficult to identify patterns relevant to all unit settings. As the rollout of LA has expanded at this university, a long tail of units with relatively small enrolment numbers are receiving this form of support for the first time. The LA value proposition may not be as apparent in these units. Chances are these academics are already familiar with the patterns identified in the LA reports and so they are less likely to be useful for them. It also challenges the notion implicit in top-down LA implementations that see pervasive analytics use as the natural outcome. It is not just about “building it and they will come”. When student retention is a key driver for LA there should be targeted initiatives with a variety of stakeholders to identify and act on areas of concern. Different stakeholders' participation in the process leads to building trust and in turn implementing evidence-based solutions that meet student needs (Gray et al., 2022; Ifenthaler & Yau, 2022; Shah et al., 2021). Cultivating academic use of analytics tools embedded in the LMS is also a needed action to respond to the issues related to validity already outlined.

Although users say they are familiar with the built-in analytics features of the LMS, survey results suggest they are not confident in its use and so the main use of the data (promise to solve problems) has not taken place as expected. This suggests further opportunities exist for professional development activities to support the integration of built-in analytics features of the LMS into academic staff practices. Given that staff already use the LMS for teaching purposes, it seems reasonable to also encourage using it for thoughtful analysis and reflection on the data to make responsive changes during the teaching session. Furthermore, these results also indicate the need to investigate enhancements to built-in analytics features of the LMS that make it easier and more compelling for academics to incorporate LA into their teaching practice. It is, however, essential to include different stakeholders in such systems development to increase understanding of the tools available and thus address the three crucial factors: trust, validity, and transparency.

5 CONCLUSION

A review of these survey results at a long running LA implementation highlights both the gains and the pain points. A key implication here is that the participation of different stakeholders in the design of LA can be a solution to increase the number of academics that integrate it into their teaching practice (Tsai et al., 2021). But it is not just about increasing the numbers. It is more about utility optimization rather than utility maximization. By being part of the design (e.g. Buckingham Shum & Luckin, 2019) stakeholders are invited to share their experiences and expectations. In addition to the opportunity to understand how systems are built, these experiences can reduce the opaqueness of LA and create trust in those expected to harness it for improved student outcomes. All the skills required for analytics success do not solely reside within one individual. The survey results outlined here suggest co-design efforts may help increase user’s trust through the increased agency of more participatory system development. The rapid pace of technological change requires LA to be treated as a living ecosystem to effectively address the issues outlined here and ensure responsible use of data that protects students and teachers.

REFERENCES

Buckingham Shum, S. J., & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. *British Journal of Educational Technology*, *50*(6), 2785–2793. https://doi.org/10.1111/bjet.12880

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

Ifenthaler, D., & Yau, J. Y.-K. (2022). Analytics for Supporting Teaching Success in Higher Education: A Systematic Review. *2022 IEEE Global Engineering Education Conference (EDUCON)*, 1721–1727. https://doi.org/10.1109/EDUCON52537.2022.9766734

Kaliisa, R., Mørch, A. I., & Kluge, A. (2021). 'My Point of Departure for Analytics is Extreme Skepticism': Implications Derived from An Investigation of University Teachers' Learning Analytics Perspectives and Design Practices. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-020-09488-w

Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., & Huber, P. (2016). Artificial intelligence and computer science in education: From kindergarten to university. *2016 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE.2016.7757570

Li, Q., Jung, Y., & Friend Wise, A. (2021). Beyond first encounters with analytics: Questions, techniques and challenges in instructors' sensemaking. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 344–353. https://doi.org/10.1145/3448139.3448172

Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022). An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 56–66. https://doi.org/10.1145/3506860.3506866

Rienties, B., Køhler Simonsen, H., & Herodotou, C. (2020). Defining the Boundaries Between Artificial Intelligence in Education, Computer-Supported Collaborative Learning, Educational Data Mining, and Learning Analytics: A Need for Coherence. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.00128

Shah, M., Kift, S., & Thomas, L. (2021). *Student Retention and Success in Higher Education: Institutional Change for the 21st Century*. Springer International Publishing AG. https://doi.org/10.1007/978-3-030-80045-1

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2021). More Than Figures on Your Laptop: (Dis)trustful Implementation of Learning Analytics. *Journal of Learning Analytics*, *8*(3), Article 3. https://doi.org/10.18608/jla.2021.7379

van Leeuwen, A., Wise, A. F., & Teasley, S. D. (2022). Teacher and student facing learning analytics. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *The handbook of learning analytics* (2nd ed., pp. 130–140). SoLAR. https://doi.org/10.18608/hla22.013

In-Course Social Interaction Associated with High Performance in Active, Social Online Business Courses

Brent Benson
Harvard Business School Online
bbenson@hbs.edu

ABSTRACT: While research has shown that educational methods that encourage student activity and social interaction result in better outcomes, asynchronous online courses often default to lecture videos with accompanying slides and multiple-choice assessments. This work examines a particular technique for encouraging active and social learning in the context of online business courses where students are regularly asked to reflect on an aspect of what is being learned in the form of textual responses, which are then shown to others in a social cohort where they can be liked, commented upon, and viewed alongside the cohort's other reflections. Our initial finding is that the amount of interaction around reflections is correlated with course grade, with large effect sizes for top grade earners and to lesser extent for other grade levels. Initial steps are being taken to evaluate the quality and authenticity of the interactions, rather than a simple measure of interaction counts, and work is being done to provide feedback to students around their social activity in the courses.

Keywords: active learning, social learning, social interaction metrics, online learning

1 ACTIVE AND SOCIAL STRATEGIES IN ONLINE LEARNING CONTEXTS

It is hard to overstate the importance of active and social teaching and learning techniques as a way of improving outcomes for students in general (Freeman et al., 2014) and also reducing achievement gaps for underrepresented student groups (Theobald et al., 2020). However, it can be challenging to facilitate active and social learning experiences in online learning contexts, especially asynchronous online learning contexts where students may be working on course materials on different schedules, at different rates, and from geographically distant locations.

Many efforts around building social connections in asynchronous online classes have used discussion forums (Adraoui et al., 2017), blackboards, and external networking tools (Kilgore, 2016). While these tools can be valuable by providing opportunities for students to ask questions, provide each other with feedback, and create peer relationships, they are often only used by a fraction of the students and can stray from direct engagement on topic (Husssin et al., 2019), limiting their effectiveness.

The Harvard Business School Course Platform considered in this research was built by an in-house technology team to support an asynchronous online version of the case method (Barnes, 1994), encouraging active, social, case-based learning in an asynchronous online context as described in Benson & Houtti, 2022. While there are many ways for students to engage with each other in this platform, the primary social affordances are social teaching elements called shared reflections in which students are presented with a reflective prompt at key learning junctures to which they must provide a textual response. Once the student has provided the textual response, they are presented with their peers' responses which they are encouraged to read, mark those they particularly like with a star, and engage in nested comment conversations around the responses.

In order to try and understand the effectiveness of the individual pedagogical feature of shared reflections we developed interaction metrics and measured their associations with positive course outcomes.

2 SOCIAL INTERACTION METRICS AND OUTCOMES

The simplest measures of social interaction around shared reflections are counts of comments given, stars given, and number of views of other students' reflections. We are using the number of views metric when comparing to course grade, as number of comments and number of stars are used as a very small input into the grading process, while number of views is an independent measure.

It is important to normalize or scale the social interaction metrics by the number of opportunities to interact, i.e., the number of shared reflection teaching elements to which the student provided a response, as described by (Benson & Houtti, 2022). Without normalization, students who didn't finish the course would not be judged equally on the parts of the course they did complete. The resulting metric is called *views per shared reflection*.

The data for this research was collected across 35 offerings of an intensive three course introduction to business program from 2018 to 2021 encompassing 18,267 students. The findings also hold across subsets of these courses by time period and in individual program offerings.

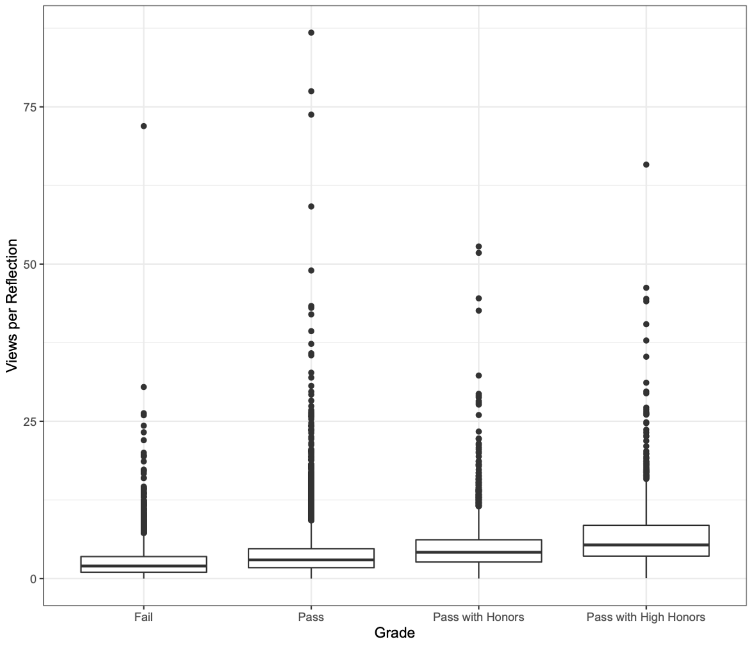


Figure 1: Distribution of views per reflection by course grade

Computing the views per shared reflection for each student and then looking at the distribution of those values by a comprehensive grade assessment shows a relationship between higher reflection views and higher grades as see in Figure 1. While there are many students who have high participation counts and did not achieve a high grade, the median and average values of participation are higher for each successive grade level.

In order to understand the significance of the relationship between views per shared reflection and grade we used *Cohen's d* to measure the effect size on transitions from fail to any level of passing, from failing or passing to passing with some level of honors, and from any other level to passing with high honors as seen in Table 1.

Table 1: Effect size of views per reflection between groups

Group 1	Group 2	Effect Size
Failed (n=4,441)	Passed/Honors/High Honors (n=13,826)	0.44 (small)
Failed/Passed (n=14,881)	Honors/High Honors (n=3,386)	0.66 (medium)
Failed/Passed/Honors (n=16,968)	High Honors (n=1,299)	0.92 (large)

The most significant effect was in students achieving the highest grade vs. others. The amount of participation, along with other indicators like amount of time on the platform, and quiz scores are being used to understand which students are at risk for falling behind or missing a deadline, with the goal of eventually providing actionable feedback to students during the course.

3 INTERACTION QUALITY RATHER THAN QUANTITY

When examining the interactions between students in shared reflections more closely, it becomes clear that not all social interactions are equal. Giving a star to a comment, reading over a reflection, or responding with a comment of "I agree" is not the same as a comment that synthesizes ideas from another's reflection around a unifying principle, or a response to another's point relating it to a specific personal experience.

The work of Marti and Smith (Marti & Smith, 2022) describes a social participation approach designed to encourage deep engagement with course material. They have developed a set of standards around the quality of social engagement and students are asked to keep a participation portfolio containing examples of their own good participation, as defined by the standards as interpreted by the student. At certain intervals, the students are asked to choose three of their best examples of participation for that period and to reflect on how the participation exhibited their intellectual growth, or contributed to the intellectual growth of others.

We have done some initial work on training an NLP model to classify textual participation based on human-classified examples based on standards similar to those in Marti and Smith using Transformer-based deep learning models using open-source HuggingFace libraries (*Distilbert-Base-Uncased · Hugging Face*, 2022) although more labeling will be needed to achieve acceptable accuracy. We would also like to implement platform-based approaches around self- and peer-assessment of participation quality, which would have great ability to scale across large, asynchronous online courses.

4 SUMMARY

We have shown a relationship between the amount of social participation around shared reflections and course grade, especially for top grade earners in a rigorous set of asynchronous, online business courses. We are working towards incorporating measures of participation quality as defined by a participation rubric, and measure through self-assessment, peer-assessment, and/or trained machine learning models.

REFERENCES

Adraoui, M., Retbi, A., Idrissi, M. K., & Bennani, S. (2017). Social learning analytics to describe the learners' interaction in online discussion forum in Moodle. *2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–6. https://doi.org/10.1109/ITHET.2017.8067817

Barnes, L. B. (1994). *Teaching and the case method: Text, cases, and readings* (3rd ed.). Harvard Business School Press.

Benson, B., & Houtti, M. (2022). Developing Social Interaction Metrics for an Active, Social, and Case-Based Online Learning Platform. In Y. "Elle" Wang, S. Joksimović, M. O. Z. San Pedro, J. D. Way, & J. Whitmer (Eds.), *Social and Emotional Learning and Complex Skills Assessment: An Inclusive Learning Analytics Perspective* (pp. 299–310). Springer International Publishing. https://doi.org/10.1007/978-3-031-06333-6_14

Distilbert-base-uncased · Hugging Face. (2022, November 3). https://huggingface.co/distilbert-base-uncased

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Husssin, W. T. T. W., Harun, J., & Shukor, N. A. (2019). Online interaction in social learning environment towards critical thinking skill: A framework. *Journal of Technology and Science Education*, *9*(1), Article 1. https://doi.org/10.3926/jotse.544

Kilgore, W. (2016). *Social Learning in Online Environments*. https://humanmooc.pressbooks.com/chapter/social-learning-in-online-environments/

Marti, D., & Smith, M. D. (2022). *Motivating Data Science Students to Participate and Learn* (arXiv:2204.14108). arXiv. https://doi.org/10.48550/arXiv.2204.14108

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Flexible Coupling of Learning Analytics Research and Practice in the University: A Collective Strengths Approach

Alyssa Friend Wise, Andrew Brackett, Ben Maddox
New York University, New York, USA
[alyssa.wise, andrew.brackett, ben.maddox] @nyu.edu

ABSTRACT: This paper describes the relationship, characterized as a flexible coupling, between a mission-aligned learning analytics research center and service unit within New York University (NYU) over the last six years. This is done to highlight a collective strengths approach where complimentary areas of focus reinforce shared goals (Kim, et al, 2022). The flexible and collaborative model endeavors to best utilize the unique skills and strengths different organizational units can bring to the development of a learning analytics program, while not disrupting activities that are not necessarily overlapping in nature or focus, in order to initiate investment to develop a culture of research-informed LA at scale. Through this presentation we offer an additional model to that of the innovation center (Buckingham Shum & McKay, 2018) for how learning analytics research and practice can be positioned to be mutually supportive and growth-oriented within the university context.

Keywords: Institutional architecture, practical adoption, strategies for scalability

1 BACKGROUND & RATIONALE

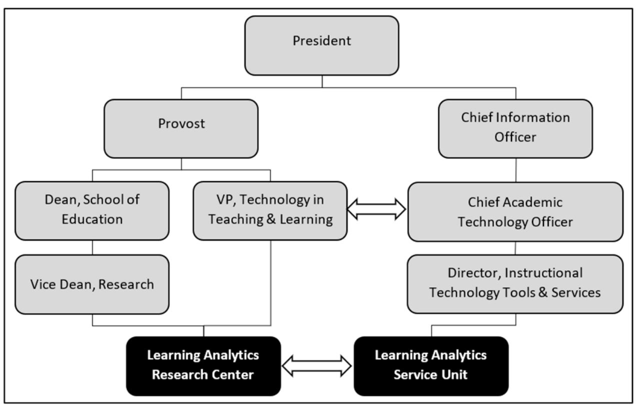
Learning analytics (LA) is a field of research and practice concerned with leveraging insight from educational data on a timely basis to improve teaching and learning. It's applied character has been a key feature of the field from the start and can be seen in its ongoing attention to use and systems (Lang et al., 2022) as well as a track devoted to reports of practice both at *LAK* and in the *Journal of Learning Analytics*. Yet while both LA research and practice continue to grow, their relationship is often tenuous, with advances in each relatively disconnected from the other (Ferguson et al., 2016). In the university context this may be attributed in part to the institutional architecture within which LA work takes place. As described by Buckingham Shum and McKay (2018): (1) Faculty/researcher-led LA efforts may be highly innovative and generate rigorous empirical evidence but often lack the infrastructure to scale and/or fail to respond to widely recognized needs; (2) IT services team-led LA efforts have the resources and skills to deliver analytics at scale, but can be limited by pre-defined data provided by product vendors and/or lack of engagement in participatory design practices. Buckingham Shum and McKay (2018) offer one way to overcome this dichotomy through a model that is currently implemented in each of their institutions: (3) The innovation center as an autonomous LA unit housing both academic and professional services staff together. In the current paper we present an additional model that emerged over the last six years at our institution: (4) Flexible coupling of mission-aligned LA research and service units in a collective strengths approach.

This model emerged to meet the challenge of initiating investment to develop a culture of research-informed LA at scale. By 2014, instruction at NYU, even when fully "in person," had a large digital component but no ways to examine this from a research standpoint or to support faculty in finding insights into what was happening in virtual learning spaces. The faculty advisement committee on teaching and learning technologies thus recommended both the convening of a working group to understand needs and opportunities for a LA service unit and the creation of a complementary research center to strengthen technology-enhanced learning innovation across the university. As learning analytics was a relatively new technology, this strategy solicitated relatively modest investments on each of the two sides to get the ball rolling quickly while generating additional buy-in.

2 THE LA SERVICE UNIT: ORIGIN AND ORGANIZATION

The working group launched in 2015, led by a post-doctoral scholar with faculty and IT staff from central university and individual school administrations. The group identified unique attributes that NYU might bring to the research and practice of LA: (1) The vast number of students and unique global reach suggested a LA-at-scale strategy as a distinguishing feature; (2) The breadth of academic disciplines and decentralized organization pointed to the need for ongoing consultation to understand diverse learning practices and needs. Over 2015-2016, the working group's recommendations, with support from expert and early adopter faculty, IT leadership and the Office of the Provost, led to the establishment of a multi-year capital project to fund and staff a nascent LA technical and support program housed within the IT academic technology division (Fig 1a). As of Fall 2022, the Learning Analytics Service is now available to all faculty, which is over 6,000 individuals.

(a)



(b)

The LA Research Center's mission… fuses service to the university (by supporting improved teaching and learning) with research (by increasing our understanding of how data can inform decision making). The synergy of research and service in learning analytics is dramatically more powerful than if either were to be pursued in isolation: research without a clear teaching or learning need is not impactful, service activities that do not draw on or contribute back to the underlying knowledge base are less effective.

Figure 1: (a) Organizational Position of the Units; (b) Research Center Founding Document Excerpt

3 THE LA RESEARCH CENTER: ORIGIN AND ORGANIZATION

2016 marked the start of a two-year visioning and consultation process, led by senior leadership and research faculty in the School of Education. A university-wide perspective was infused through ongoing consultation with the university VP for Technology in Teaching and Learning and the Chief Academic Technology Officer. This led to the launch of the Learning Analytics Research Network (NYU-LEARN) in 2018 as a research center focused specifically on LA, with several distinctive characteristics. First, it had a mandate to fuse attention to research and scholarly concerns with practical service to the university (Fig 1b). In support of this, the center was designed to be team-based (rather than driven by a single faculty's research agenda) and grounded in collaborations with partners across the university (e.g., Arts & Sciences, Dentistry, Engineering, Professional Studies). Finally, the NYU-LEARN received initial support and had accountability directly to the university-wide Provost's Office (Fig 1a), leading to its representation not as a boutique research lab but as "a university-wide research center housed in the School of Education."

4 FLEXIBLE COUPLING AS A COLLECTIVE STRENGTHS APPROACH

The relationship between the LA Service Unit and Research Center dates back to 2016 when the planning for both was underway. These processes were formally independent (and staggered) but had two important parallels. Both (1) were highly collaborative, widely consultative and took place over an extended period of time; (2) included balancing input (the post-doc brought a research lens to Service Unit planning; Provostial and IT leadership brought practice lens to Research Center planning). Another important alignment emphasized by the leaders of both units was (3) orientation towards human-centeredness: involving the intended users of LA in the process of their creation (Buckingham Shum et al., 2019) through participatory methods (Sarmiento & Wise, 2022). These three alignments created need and desire for the units to engage the complementary expertise of the other to enrich their efforts on particular projects. Equally importantly, a trustful relationship has developed that fundamentally undergirds work across the units. Without a formal organizational relationship, regular conversations have been critical to create shared awareness of efforts and identify opportunities to leverage collective strengths. Some key examples of joint work resulting from this follow below.

Instructional Dashboard. Early years of the LA Service Unit's work involved consolidation of data streams from instructional tools and developing use cases for common teaching challenges in consultation with faculty. As initial dashboard versions were built to address these challenges, the Research Center studied how faculty worked with the data to inform their teaching. This both fed back into future tool development and built scholarly knowledge about analytics use (Wise & Jung, 2019). In addition, it drove dashboard support efforts in which an NYU-LEARN doctoral student worked with the LA Service Unit to develop training workshops and resource materials for faculty. University-wide availability of analytics came after ~5 years of effort in Fall 2022. Launching a program of this complexity across multiple schools (and global locations) required a collective commitment that LA is both a time and resource priority across academic, technical, instructional and research communities.

Rapid Response to Remote Instruction. A significant challenge faced by many universities in 2020 was the shift to remote instruction that occurred in response to the COVID 19 pandemic. In addition to scrambling to quickly modify teaching methods and materials to an online format, many faculty felt that they were "driving blind" without the opportunity to observe and reflect on student engagement in the classroom. In response, the two units came together quickly to offer faculty simple reporting on student interactions with online platforms. Expertise from NYU-LEARN helped shape report design and guiding language while the LA Service Unit focused on scalable reporting via existing information channels. The result was delivered through university-embedded Google Suite tools and early positive feedback during Summer 2020 led to revision and expansion to thousands of faculty in Fall 2020.

Disciplinary Driven Analytics. In addition to university-wide projects described above, the LA Service Unit and Research Center also receive requests from disciplinary faculty whose needs are best met with involvement from both units. For example, in 2019 a faculty member in Professional Studies approached NYU-LEARN about developing analytics to support students in cultivating their online teamwork skills (Li et al., 2021). This project required working with specific data streams with which the Service Unit was intimately familiar. Conversely, a group of STEM faculty recently approached the Service Center about improving student success tools for faculty in large format courses. To take a research-informed approach to the development of classification and prediction tools, a doctoral student has been jointly appointed to the Service Unit and Research Center.

5 REFLECTIONS

The above examples and our experiences over the last six years have revealed both advantages and drawbacks to the flexible coupling approach compared to other research-service configurations. For advantages, flexible coupling may be a model that is easier to adopt from a leadership and fundraising perspective, while at the same time allowing service and research units their own "lanes" of production that produce benefits more aligned with the "worlds" each inhabits. From a financial perspective, this means advocating for smaller investments in learning analytics across different and non-competing funding sources that are easier to justify compared to a larger new unit. The parallel investment on the two sides also can help lessen concerns of tension or territory over a single unit's agenda for learning analytics at the institution, with the Service Unit being concerned with infrastructure, data governance, and providing service at scale to constituents, and the Research Center focused on knowledge generation, stakeholder partnerships and basic research. One central drawback to flexible coupling is the continual effort needed for the two units to stay aware of and coordinated with each other's projects. Even with well-aligned missions and orientation, the lack of combined leadership or portfolio means the coupling's success is only as strong as the commitment to communication and coordination of the individuals involved at an operational level. There may also be moments where timelines do not align, leading the partnership to take place in drawn out stages, which can delay research/service development and outputs. Considering the flexible coupling and innovation center models together, we see several similarities. Both involve (1) non-traditional positioning of a research functionality that has a university-wide purview with the ability to develop strategic relationships with senior university leaders; (2) dedicated staff with LA expertise, distinct from the general IT team; and (3) engagement with faculty clients / champions as a key additional player (which may be extended to students in the future). Perhaps these three characteristics, regardless of the specific model of connection between academic and service personnel, create the necessary conditions for LA research and practice to be mutually supportive in the university context.

REFERENCES

Buckingham Shum, S. & McKay, T. (2018). Architecting for learning analytics: Innovating for sustainable impact. *Educause Review, 53*(2), 25-37.

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

Ferguson, R., et al., (2016). *Research evidence on the use of learning analytics: Implications for education policy*. Seville, Spain: Joint Research Centre. doi.org/10.2791/955210

Kim, S. Y., et al. (2022). Updating the institutional collective action framework. *Policy Studies Journal, 50*(1), 9-34.

Lang, C., Siemens, G., Wise, A. F., Gašević, D. & Merceron, A. (Eds.) (2022). *Handbook of learning analytics (2nd ed)*. Edmonton, AB: SOLAR.

Li, Q., Jung, Y., Wise, A. F., Sommer, S., & Axelrod, V. G. (2021). Designing analytics to support team learning. *Perspectives on Digitally-Mediated Team Learning* (pp. 147-165). Springer.

Sarmiento, J. P., & Wise, A. F. (2022). Participatory and co-design of learning analytics: An initial review of the literature. *Proceedings of LAK'22* (pp. 535-541). New York: ACM.

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.

Exploring the potential of NLP for automated assessment of creativity

Antonella Strambi
University of South Australia
antonella.strambi@unisa.edu.au

David H. Cropley
University of South Australia
david.cropley@unisa.edu.au

David Caldwell
University of South Australia
david.caldwell@unisa.edu.au

ABSTRACT: This Practitioner Report discusses a new approach to assessing creativity, operationalized as the ability to produce a useful and novel solution to a problem. A creativity test was designed that involves presenting participants with a visual stimulus and asking them to generate a caption. This method enables observation of the product of a creative process, rather than potential creative ability, while limiting variation in participants' responses, thus facilitating automated scoring. Using Natural Language Processing (NLP) to measure similarity between captions generated by participants and a low-novelty model caption suggests that this approach approximates human judgements of creativity, especially for those captions that were judged by humans as non-creative. This has the potential to increase efficiency in creativity assessment at scale, in that it allows early identification of captions that are almost certainly non-creative and can therefore be excluded from further examination. To arrive at an automated scoring system that approximates human judgement for highly creative responses, however, a model that combines similarity with the detection of humour and figurative language may be required.

Keywords: Creativity, Natural Language Processing, NLP, Semantic similarity, Assessment

1 BACKGROUND: MEASURING CREATIVITY

Creativity is highly valued by employers (Deloitte, 2016). Therefore, assessing students' creativity is an important concern for educational institutions (Marrone & Cropley, 2022). While creativity is often associated with artistic expression, the type of creativity we deal with in this report concerns problem solving. From this perspective, we follow Plucker et al.'s definition of creativity as "the interaction among aptitude, process and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context" (2004: 90).

Current approaches to assessing creativity often operationalise creativity in terms of ideational fluency and divergent thinking. For example, the Alternative Uses Test -- a classical test of ideational fluency -- involves naming the greatest possible number of uses for a common object, such as a paper clip (Snyder et al., 2004). Other tests, like the Divergent Association Task (Olson et al., 2021), ask participants to generate lists of words associated with a verbal stimulus and measure semantic distance (i.e., divergence) between these words as an index of creativity. Classic divergence tests have traditionally relied on human judgement for scoring, although attempts have been made to automate the scoring process in recent years (e.g., Beaty & Johnson, 2021).

Existing tests have been found to produce valid and reliable measures of an individual's potential to engage in creative problem solving (Runco & Acar, 2012). However, it can be argued that word lists provide limited information on a persons' actual creative ability; tests that generate more complex and informative participants' responses, on the other hand, are time consuming to score and potentially less reliable, as they require human judgement. To address these shortcomings, the second author designed a test that asks participants to generate a caption in response to a visual stimulus (i.e., a photograph). This limits output variation, facilitating scoring and reliability, while at the same time enabling observation of an authentically creative product. We explored the possibility of automating at least part of the scoring, working toward reliable and efficient assessment of creativity at scale. In this report, we present the results of one of our explorations, which involves computing similarity between participants' responses and a low-creativity model caption.

2 DATA AND DESIGN

As mentioned above, a creativity test was designed that involves presenting participants with a randomly selected image out of four possible options. The image was displayed on a computer screen and the participants, who were all students at an Australian secondary school, were asked to type a creative caption into a text field. We then generated 'literal' descriptions of each image as a basis for comparison. The model captions meet the 'usefulness' criterion of the creativity construct, but fail on the 'novelty' one, hence exemplifying low-creativity responses:

Image 1 ('Elephant'): Flying elephant plays basketball
Image 2 ('Car'): Two persons on the road look at a dent on the back of a white car
Image 3 ('Bulldozer'): Person standing outside a building site looks at a bulldozer beyond a fence
Image 4 ('Stairs'): Person sits at the bottom of the stairs

Our dataset consists of 1418 participant-generated captions, including: 222 captions for Image 1; 450 for Image 2; 439 for Image 3; and 307 for Image 4. Human judges were asked to score each caption in terms of its creativity on a scale from 1 = lowest creativity to 5 = highest creativity. Mean scores for each image are presented in Table 1.

Table 1: Mean and Standard Deviation of Human Judges Scores

Image	N of captions	Mean Judge Score	StDev
1 ('Elephant')	222	3.07	0.99
2 ('Car')	450	2.66	0.90
3 ('Bulldozer')	439	2.92	0.98
4 ('Stairs')	307	2.58	0.97

To measure divergent thinking in the participants' captions, we tested for similarity with the relevant low-creativity model captions. We generated sentence embeddings using sBERT SentenceTransformers (https://www.sbert.net/) and a high-performing pre-trained model, paraphrase-MiniLM-L12-v2, which is designed specifically to detect paraphrasing. We then computed cosine similarity between embeddings for participants' and model captions. Finally, we calculated Pearson's correlation between our similarity scores and human judges' scores for each caption, to explore the relationship between these two measures.

3 FINDINGS

On the whole corpus comprising four sets of image captions, a moderate, statistically significant inverse correlation was found between similarity with the model captions and mean judges scores, at r=-0.32 (N=1418; p<0.00001). The strength of the relationship varied across datasets: 'Stairs' returned the strongest association, with -0.5505 (N = 307; p<0.00001), whereas the 'Car' dataset returned the weakest, at -0.188768 (N = 450; p<0.0001). Results for each image dataset are shown in Table 2.

Table 2: Correlation between Similarity with Model Caption and Human Judges Scores

Image	*N of captions*	*Mean similarity with Model Caption*	*Pearson's Correlation with Mean Judge Scores **p<0.00001, *p<0.0001*
1 ('Elephant')	222	0.3818	r = -0.41**
2 ('Car')	450	0.2204	r = -0.19*
3 ('Bulldozer')	439	0.2598	r = -0.37**
4 ('Stairs')	307	0.3056	r = -0.55**
TOTAL	1418	0.2851	r = -0.32**

Based on these results, computing similarity with a surface-level description of the image prompt appears useful, given that there is some correlation with human judges' scores. This is an interesting finding, as it suggests that divergence from a surface description of the image prompt is a relevant criterion for human judges assessing caption creativity. In the 'Stairs' set, for example, up to 30% of variance in judges' scores can be explained by similarity ($R^2 = -.5505^2 * 100$). Variance across datasets, however, suggests that image prompt selection may play an important role: when images offer less room for creative interpretation, as is the case of the 'Stairs' and 'Elephant' datasets, a greater proportion of captions show high similarity with the model, increasing correlation with human judges' scores. Similarity scores may also have been influenced by the way model captions were phrased, as correlations are higher for model captions containing fewer words. These observations will guide future refinements in test design to improve validity and reliability.

Visual inspection of participants' captions also suggests that our approach may be especially useful in identifying low-creativity responses, as correlations appear stronger when similarity is higher. When similarity with the model caption is low, on the other hand, additional criteria may need to be considered. Participant captions that were rated as highly creative by human judges tended to evidence linguistic creativity through: (a) use of humour – especially puns, satire, and parody. This included intertextual references to well-known cultural tropes and artefacts, such as film titles and characters, or exploitation of polysemy for wordplay (see Skalicky, 2018 for similar observations); (b) metaphors and 'headlinese'. Several captions that received high creativity scores from human judges were in a 'headline' form and contained 'serious' (i.e., non-humorous) metaphors, which demonstrates the participants' ability to link distant concepts through abstract thinking (see also Skalicky, 2018). A valid and reliable model for automated scoring of creativity would therefore need to account for these elements, in addition to measuring distance from literal image descriptions.

In conclusion, our preliminary exploration shows great potential for Natural Language Processing (NLP) to assist in automatic scoring of creativity, operationalized as individuals' ability to generate image captions that are both effective and novel. Similarity with effective but low-novelty model captions correlated significantly with human judges' scores, suggesting that this measure may be employed to discriminate between participants' captions that are almost certainly not creative and those that require human judgement to reliably evaluate creativity. To approximate human judgement of highly creative responses, however, a model that combines similarity with the detection of humour and figurative language may be required. While there have been advancements in algorithm development for this purpose (e.g., Gong et al., 2020), considerable work remains to be done to overcome existing limitations. Once developed, such model could be applied to texts beyond simple captions -- for example, students' essays. For now, our approach could be viewed as a simple, easy-to-implement method for conducting preliminary screening and identifying low-creativity responses, building some efficiency in the assessment of students' creative potential at scale.

REFERENCES

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior research methods*, *53*(2), 757-780.

Deloitte. (2016). Talent for survival: Essential skills for humans working in the machine age. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/Growth/deloitte-uk- talent- for- survival- report.pdf

Gong, H., Gupta, K., Jain, A., & Bhat, S. (2020). Illinimet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing* (pp. 146-153).

Marrone, R.L., Cropley, D.H. (2022). The Role of Learning Analytics in Developing Creativity. In: Wang, Y., Joksimović, S., San Pedro, M.O.Z., Way, J.D., Whitmer, J. (eds) *Social and Emotional Learning and Complex Skills Assessment. Advances in Analytics for Learning and Teaching*. Springer, Cham. https://doi.org/10.1007/978-3-031-06333-6_5

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39(2), 83–96.

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, *118*(25), e2022340118.

Runco, M.A. & Acar, S. (2012) Divergent Thinking as an Indicator of Creative Potential, *Creativity Research Journal*, 24:1, 66-75, DOI: 10.1080/10400419.2012.652929

Skalicky, S. (2018). Beyond Literal Meaning: Linguistic and Cognitive Features of Figurative Language Processing and Production. [Doctoral thesis, Georgia State University] https://scholarworks.gsu.edu/alesl_diss/44/

Snyder, A., Mitchell, J., Bossomaier, T., & Pallier, G. (2004). The creativity quotient: an objective scoring of ideational fluency. *Creativity Research Journal*, 16(4), 415-419.

An Exploratory Analysis Investigating Determinants of Learning Performance in the Workplace

WeiHan Lin[1], **Xingheng Wang**[1], **Aarren Minneyfield**[2], **Mengyu Liang**[1], **Yihua Wu**[3]
[1.] Novartis Innovative Medicine, China; [2.] Southern Illinois University, Carbondale; [3.] Shanghai Delightgo Internet Technology Co,. Ltd.
harrylin@umich.edu

ABSTRACT: The widespread adoption of online learning in response to the pandemic has left organizations with copious amounts of learning data that organizations can use to justify the optimization of online learning and learning effectiveness overall; the question is how? Considering the unpredictable, and uncertain marketplace, organizations seek to identify the potential of new hires by assessing their learning capabilities. Therefore, the present study sought to use learning data to identify predictors of effective learning among workers, which is invaluable to any corporation that wants to increase its competitive advantage on talent management and development. A pretest-posttest quasi-experimental design was used to evaluate the learning outcomes of medical representatives in a multinational pharmaceutical company. Participants were new-hired medical representatives that were required to finish a self-paced, online professional training program (N = 156). A cluster analysis was conducted to categorize participants into three performance groups (promotion-focused, prevention-focused, and balanced) based on eight predictor variables. The study's exploratory findings provide some evidence that learning outcomes can be predicted with behavioral measures of learning. Implications and practical significance are also discussed.

Keywords: Organizational learning, Learning outcomes, Employee Performance,

1 INTRODUCTION

As it stands, the rapid development of online learning during the pandemic has resulted in a large amount of learning behavior data for organizations. The question is, how do we use this data to justify the optimization of online learning and testing to improve learning effectiveness. Given the present organizational climate characterized by technological innovations, constant changes in organizational structures and processes, the criteria for determining and evaluating work performance are constantly changing. Therefore, it would be prudent for training managers and commercial training heads to utilize this behavioral data to identify the factors that distinguish effective learners from less effective ones. However, there is little research regarding methods in identifying effective learners, especially when identifying individuals who can effectively acquire and apply knowledge in a workplace context. As such, the aim of the study is three-fold. First, we examine whether employees can be categorized into clusters based on variables of learning behavior. Second, we examine whether there are variables that significantly correlate with the measures of employees' learning effectiveness and assess which variables could be used to develop a predictive model. Finally, we perform an exploratory data analysis (EDA) from sequential data to see if there are any difference in the learning behavior patterns between clusters. Afterwards we discussed the characteristics of each cluster to explore the implications of these differences.

2 METHODS

2.1 Participants

The study consisted of 166 medical representatives from a multinational pharmaceutical company in China who were required to take part in a self-paced, online professional sales training program. The final sample population was 156 participants after 10 trainees were dropped for failing to complete at least 90% of the online training and take the posttest.

2.2 Procedure

The study had a quasi-experimental design conducted during the employee's training. The pretest was assigned a week before the first day of the training. Afterwards, the participants were given 12 days to complete the online professional training. The training was divided into 12 subsections to encourage learners to finish one subsection each day. The online professional training course consisted of 79 learning items including 4 different formats: video, quiz, practice assignment and reflection question. All participants were required to take a pretest and a posttest at the beginning and end of the training respectively. At the end of the training, all participants were required to reflect on the training and provide qualitative responses on the value of the professional training on their professional development.

2.3 Measures

Table 1: Variables and their definitions

Variable	Definition
Pretest Score	The participant's standardized test score before the training began.
Posttest Score	The participant's standardized test scores a week after training ended.
Improvement Rate	The difference between the pretest and posttest scores expressed as a percentage of the pretest.
Time Per Assignment	The time each learner spent completing assignments at the end of each training subsection. Measured in seconds.
Time Per Quiz Section	The time at which each learner spent on each quiz section including responses time and review time. Measured in seconds.
Time Per Quiz	The time at which each learner provided responses to each quiz. It was measured in seconds.
Quiz Score	The score that a learner got from quiz section during the training.
Quiz Rank	The average rank by learner's quiz scores that learner got for each quiz. A high rank indicates poor performance and vice versa.

2.4 Analyses

Firstly, a cluster analysis was conducted because we were interested in seeing whether the variables of learning behavior could produce distinct subgroups of learners. ANOVAs were used to compare the final clusters on the different variables of learners. Pearson's correlations were used to explore the relationships between these variables and an exploratory data analysis was conducted to explore participates' learning behaviors, using temporal sequence data. a sequence data framework was used to analyze temporal sequence data from learner-centric perspective (Johnston et al., 2021). We calculated the probability of each learning item that was completed in each possible sequence. We denoted $P_{(i, j)}$ with i representing the sequence number of the learning item (1-79) and j representing the order number in which it was completed. For instance, $P_{(1,79)}$ denotes the probability of the first learning item being completed as the 79th task. If 50% of learners complete the first learning item as the 79th task, then $P_{(1,79)}$ is 0.5. Therefore, we calculated a 79 x79 probability matrix to explore the difference between learning item sequence for each learner cluster.

3 RESULTS

A cluster analysis was conducted on the learners in order to identify learner subgroups based on learning effectiveness. The analysis used k-means clustering, which is a partitioning clustering algorithm whereby the numbers of clusters are automatically specified a priori which resulted in a three-cluster solution. The resultant clusters were labeled promotion-focused (n = 85), prevention-focused (n = 52) and balanced (n = 19). Promotion-focused learners were characterized by their high scores on the pretest, low posttest scores, and fast completion times on the assignments and quizzes. Prevention-focused learners, in contrast, had the lowest scores on the pretest and moderately high scores on the posttest, as well as the slowest completion times on their quizzes and assignments. The Balanced group had moderately high scores on the pretest, highest on the posttest and had moderately quick completion times on their assignments and quizzes. Uncorrected univariate ANOVAs for the three-cluster solution revealed significant differences on 7 of the 8 variables of learning behavior. Although pretest scores were found to be the same between all three clusters, the ANOVA revealed significant differences between the 3 clusters for average posttest scores ($F(2,154) = 190.273$, $p = 0.000$), reflection score ($F(2,154) = 15.096$, $p = 0.000$), average quiz score ($F(2,154) = 4.280$, $p = 0.016$), average quiz rank ($F(2,154) = 4.583$, $p = 0.012$), average time spent per quiz section($F(2,154) = 10.194$, $p = 0.000$), average time spent per assignment ($F(2,154) = 3.998$, $p = 0.020$), and average time spent per quiz ($F(2,154) = 5.080$, $p = 0.007$). Multiple Pearson product-moment correlation coefficients were computed to assess the relationship between learners' posttest scores and variables of learning behavior. Posttest scores were found to be weakly positively correlated with average time spent per assignment [$r(154) = 0.198$, $p < 0.05$], average time spent per quiz [$r(154) = 0.292$, $p < 0.01$], delay [$r(154) = 0.212$, $p < 0.05$], and average quiz score [$r(154) = 0.255$, $p < 0.01$].

Table 2: Means, standard deviations and ANOVA results for learning variables

	Mean ± standard deviation			F	p
	Promotion (n=85)	Prevention (n=52)	Balanced (n=19)		
Pretest Score	66.96±4.56	65.42±4.07	66.30±4.51	1.996	0.139
Posttest Score	65.03±3.59	71.94±2.52	80.05±3.68	190.273	0.000**
Reflection Score	1.37±1.02	2.27±1.33	2.84±1.80	15.096	0.000**
Quiz Score	46.15±22.86	52.50±15.91	60.09±18.42	4.28	0.016*
Rank Score	100.60±43.82	86.87±35.29	72.38±36.53	4.583	0.012*
Time per Quiz Section	229.34±134.27	327.15±138.35	329.08±132.60	10.194	0.000**
Time per Assignment	96.87±58.42	128.38±76.35	119.93±61.67	3.998	0.020*
Time per Quiz	175.32±109.00	228.12±97.07	226.71±82.63	5.08	0.007**

*p<0.05; **p<0.01

To better understand the learning sequence of each cluster, we made a visualization for 79 x 79 matrix of probability. The x-axis displays the order of being completed (1-79) and the y-axis displays the sequence number of the learning item (1-79). The mark is encoded with 2 channels: color for the types of learning items (blue for video, green for quiz, purple for assignment, orange for reflection), saturation for the portion of total learners (0% saturation for 0% of total learners, and vice versa). If all learners of the cluster strictly follow the desired learning sequence, there will be a clear diagonal with high saturation from the top-left to the bottom-right. The first graph in Figure 1 (leftmost), the learning sequence of the promotion-focused cluster is much more divergent than the others. Learners with early learning behaviors are shown below the diagonal, with fewer appearing over time. There are marks with less saturation near the diagonal as well which indicates that learners did not follow the desired learning sequence even at the subsection level. Within Figure 1's middle graph, we observe there are two clear deviations from the diagonal. A certain portion of learners from prevention-focused cluster completed video contents (blue mark) in advance and delayed assignments (purple mark). In Figure 1's rightmost graph explains how most learners from balanced cluster strictly follow the desired learning sequence to complete the training program. There are small groups of learners who deviated from diagonal on some videos (advance than the desired sequence) and some assignments (delay than the desired sequence).

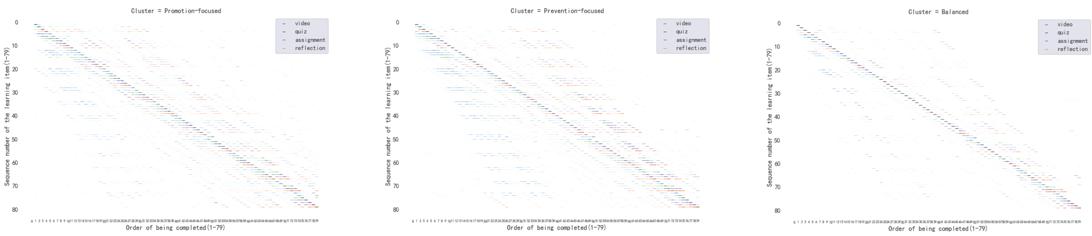


Figure 1: Learning sequence from the resultant clusters

4 DISCUSSION

The present study demonstrated how learning behaviors can be used to form clusters based on different levels of learning performance, which can be used to identify effective learners. Our sample of learners were partitioned into three groups using eight variables. Learners' scores on the posttest were associated with how much time they spent over the time allotted to their training, their average quiz rank, average quiz scores, as well as the average time they spent on their assignments and quiz. This suggests that the participants who scored high on their quiz scores and spent more time on their assignments and quizzes were more likely to have high posttest scores. These findings imply that potential employees who spend more time learning work relevant information are more likely to perform better, which is conducive to literature on effective learning strategies (Tagg, 2018). As for the sequence data framework, we can observe the obvious learning sequence pattern between the three clusters. This can be part of the explanation for the differences in learning outcomes. The learners who followed the designed learning sequence can learn most effectively, which also indicates the effectiveness of pedagogy (learn, test and practice for every content). In sum, the variables for clustering and the sequence learning behavior can be utilized as the indicators for learning management in practice for workplace learning. Future studies should adopt this framework and determine if there are specific learning behaviors associated with performance as well as to assess the reliability of the training timeline given the high standard deviation values for the time spent measures.

REFERENCES

Johnston, I. G., Greenbury, S., Peach, R., Lefèvre, D., & Barahona, M. (2021). Understanding learner behaviour in online courses with Bayesian modelling and time series characterisation.

Tagg, J. (2018). Rome Wasn't Built in a Day: Why Transformative Learning Takes Time. *Journal of Transformative Learning*, *5*(1).

Building an Infrastructure for A/B Experiments at Scale: The Challenges, Opportunities, and Lessons for the Learning Analytics Community

Sreecharan Sankaranarayanan, Chellie Harrison, Sarita Kumari, Kim Larson, Robert Lemiesz, Nicolas Mesa, Ryan Mitts, Ishita Verma, Dawn Zimmaro
Talent Development, Amazon.com, Inc.
{sreeis, chelliec, saritku, kilarson, lemiesz, nicfern, rmitts, eshaverm, dzimmaro}@amazon.com

ABSTRACT: A/B testing at scale provides opportunities for learning analytics researchers to learn from large sample sizes. Deploying and running live intervention experiments with such large samples, however, raises infrastructural challenges. This paper discusses some of those challenges, and reports on two possible implementations that address those challenges in a workforce learning context at a large technology company. The pros and cons of the alternatives are discussed with the help of a specific use-case, an A/B test comparing personalized feedback on open-ended responses to static feedback. Along the way, the paper discusses idiosyncrasies that have to be kept in mind while conducting and evaluating learning experiments in the industry.

Keywords: A/B Testing, Workforce Learning, Learning at Scale, Experimentation at Scale

1 MOTIVATION

A/B testing (terminology emerging from user experience research) is a “between-subjects” experimental method where an individual is assigned to one of two (or more) experimental conditions (often control, and one or more treatment conditions). Provided the baseline characteristics of these groups are identical across conditions, any observed effect can be attributed to the treatment. While a random assignment of individuals to conditions is often used to create groups with similar baseline characteristics, a small sample size reduces the statistical power of generalizing the findings to a broader population outside of the study (Faber & Fonseca, 2014). This is the promise of A/B testing at scale. A large sample size increases the likelihood of observed effects generalizing to the population.

The promise of A/B testing at scale is, of course, complemented by several challenges. First is the challenge of finding an appropriate use-case – one that measurably and quickly advances business and/or learning goals while also satisfying the dual methodological requirements of limiting possible harm or inconvenience from a treatment, as well as preventing a beneficial treatment from being withheld for too long. Second is the technological challenge of developing, testing, deploying, and maintaining an infrastructure that enables A/B testing experiments to be conducted at scale. Last is the technological challenge of collecting the data, quickly analyzing it, and, preferably automatically, declaring a “winning” condition either to limit harm or inconvenience, or to multiply the benefit.

Both large samples and the associated infrastructure for A/B testing at scale have previously been relegated primarily to large-scale collaborative initiatives such as the Pittsburgh Science of Learning

DataShop (Koedinger et al., 2010), the Super Experiment Framework under the Next Generation Learning Challenge (Stamper et al., 2012), E-TRIALS from ASSISTments (Krichevsky et al., 2020), and MOOCs (Reich, 2015), although noteworthy examples from single research groups do exist (Mostow et al., 2003). This paper describes two kinds of implementations built to address the challenges described above. The first uses custom-built or equivalent open-source technologies, and the second uses Amazon Web Services (AWS) technologies. The pros and cons of the two implementations are discussed with the help of an example use-case comparing personalized, AI-generated feedback on open-ended responses to static, hard-coded feedback. The discussion should provide research groups with the information needed to build scalable A/B testing infrastructures of their own.

2 THE USE-CASE

The industry context provides us access to a large sample but requires additional considerations. An ideal use-case will be one that produces actionable results in the short-term. That is not to say that long time-horizons are not possible, but that they need to be coupled with delivering results in the short-term. One such use-case was identified in an eLearning course for learning designers on writing learning outcomes. An open-ended question asked learners to rephrase and improve on a sub-par learning outcome. The control condition provided learners with an exemplar that they can, on their own, compare against their response. Building on prior work (Zhao et al., 2021), the treatment condition provided learners with personalized feedback on their response using an algorithm that sits atop a pretrained language model. Given a set of learner answers and exemplar answers, the algorithm identifies key phrases present in the model answers that are missing from the learner answer, allowing those missing key phrases to be delivered as personalized feedback to the learner. The model backend additionally determines a score for the learner answer based on this comparison between their answer and the model answers. This score is not displayed to the learner and is only used for hypothesis testing. In either condition, learners can submit responses as many times as they want, although only the treatment condition would provide personalized feedback each time. Four hypotheses were tested across the conditions. First, we hypothesized that learners who received personalized feedback would be more likely to edit their responses and resubmit. This is fairly trivial; however, it enables us to test if the additional time spent revising the responses was worth it for learning. Second, as a result of the personalized feedback, the quality of the terminal submission, as measured by the model-generated score, ought to be higher in the treatment. Third, the combination of the personalized feedback, and an increased number of attempts to refine their submission based on it, should lead to improved learning in the treatment. Finally, the model-generated score should be higher in the treatment condition for equivalent attempt numbers (second attempts compared across conditions, for example). The final hypothesis allows us to understand the mechanism of learning i.e., whether learning was due to personalized feedback or due to editing and refining one's responses. Lacking a proximal measure of learning such as a pre- and post-test, we used performance on the set of assessments tied to the learning outcome associated with the open-ended response question as our (distal) measure of learning. Since it is a high bar to expect a single question to produce a significant impact on the learning outcome as a whole, we reserve the last two hypothesis as exploratory. The results of the first two hypotheses, therefore, determines the "winner" of the A/B test, with the experimental condition being declared as such only if both of the first two hypotheses are satisfied.

3 THE INFRASTRUCTURE

Open-source frameworks such as Upgrade allow augmenting existing educational applications with A/B testing capabilities through Learning Tools Interoperability (LTI). Nevertheless, they are unsuited for use in corporate contexts for various reasons – scalability, security, and lack of customizability/extensibility being chief among them. As our first solution, therefore, we built a custom solution that is similar to Upgrade. Two copies of the learning experience, one with the desired treatment, are made, and the system redirects learners to one or the other based on the rules provided through an authoring interface. Note that the authoring interface needs to be implemented in addition to the redirection logic. The authoring interface can specify the percentage of learners in each condition, defined start and end conditions, and the post-experiment actions among other parameters. The redirection logic is not only a one-time redirection of a learner to an experimental condition but has to persist that assignment across learner sessions through drops in connection, log-outs, and breaks. Data from each condition must then be collected, and correctly and consistently mapped to the learners assigned to those conditions. If data-processing rules are set, the evaluation metrics can be automatically calculated to determine the outcome of the experiment and the "winning" condition that subsequent learners will be assigned to. Otherwise, duration or number of learners can also be used as terminating conditions. In our case, the number of learners (200-400) is used as the termination condition while we wait for the automatic evaluation to be implemented. The evaluation is conducted offline, and learners are reverted to the control condition in the meantime.

Our second implementation leverages Amazon Cloudwatch Evidently. Features for both the control and treatment are implemented, but hidden behind 'if' statements on the front-end interface. At runtime, the application code queries a remote service. The service decides the percentage of users who are exposed to the new feature, i.e., the treatment, and returns information about whether a specific specific user is in the treatment or control condition allowing the appropriate front-end components to be rendered. Evidently additionally allows statistical analysis to be performed at run-time and decisions made based on the results of those statistical analyses.

Each approach has its pros and cons. The former learning experience-level redirection approach allows more complex A/B comparisons such as different sequences of instruction and assessment, while the latter real-time rendering approach is ideally scoped to the level of a page, or even one learning object on the page. On the other hand, the experience-level redirection approach causes duplication of assets resulting in an increase in the burden of reporting on metrics such as completion, and an increased likelihood of learners chancing directly upon the URL for the control or treatment learning experiences in case the owner omits updating the link in all the places it is referenced (email blasts, wiki links, course catalogs etc.). The real-time rendering implementation does not require a redistribution of links and allows for a smoother post-experiment transition since it is all handled within the same learning experience. The determination for which infrastructure to use will depend on the requirements of the experiment. In our case, since the experiment focused on rendering a single assessment within the learning experience, the real-time rendering approach was the best suited.

Another element of this experiment is the machine learning model consuming learner answers and providing real-time personalized feedback. Not only does this involve the backend infrastructure necessary to consume learner answers, return the model output, and fashion that into a response to learners, it also involves the frontend infrastructure elements of rendering the personalized feedback, refreshing the interface to allow learners to respond again, and re-rendering feedback. All of this has to, of course, happen in real-time to allow for a reasonable experience for learners. The model inference code is easy to deploy, packaged along with all the requirements files into a Docker container that is then deployed using Amazon SageMaker. The front-end component development is much more effortful.

4 ETHICS OF A/B

The A/B infrastructure has interesting implications for the ethics of A/B experiments, which we wanted to make explicit. A well-developed infrastructure will allow for quickly resolving in favor of the experiment or treatment conditions, thus preventing harm or inconvenience, or making the beneficial treatment available. Of course, this is ultimately down to the experimental design and the experimenter. Nevertheless, the needed ease of doing so has implications for the design of the A/B infrastructure, as outlined in earlier sections.

5 CONCLUSION

We present two possible infrastructures for A/B testing learning experiences at scale, discuss their pros and cons, and demonstrate making a choice between them with a use-case. We expect research groups to be able to use this information to build scalable A/B testing infrastructures of their own.

REFERENCES

Faber, J., & Fonseca, L.M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics, 19*, 27 - 29.

Glaser, A. (2021, July 30). *Amazon now employs almost 1 million people in the U.S. - or 1 in every 169 workers*. NBCNews.com. Retrieved December 8, 2022, from https://www.nbcnews.com/business/business-news/amazon-now-employs-almost-1-million-people-u-s-or-n1275539

Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. *Handbook of educational data mining*, *43*, 43-56.

Krichevsky, N., Spinelli, K., Heffernan, N., Ostrow, K., & Emberling, M. R. (2020). *E-TRIALS* (Doctoral dissertation, Doctoral dissertation, Worcester Polytechnic Institute).

Mostow, J., Beck, J. E., & Valeri, J. (2003). Can automated emotional scaffolding affect student persistence? A Baseline Experiment. In *Proceedings of the Workshop on" Assessing and Adapting to User Attitudes and Affect: Why, When and How?" at the 9th International Conference on User Modeling (UM'03)* (pp. 61-64).

Reich, J. (2015). Rebooting MOOC research. *Science*, *347*(6217), 34-35.

Stamper, J. C., Lomas, D., Ching, D., Ritter, S., Koedinger, K. R., & Steinhart, J. (2012). The Rise of the Super Experiment. *International Educational Data Mining Society*.

Zhao, J., Larson, K., Xu, W., Gattani, N., & Thille, C. (2021). Targeted feedback generation for constructed-response questions. *AAAI-2021 Workshop on AI Education*.

A Rapid Approach to Learning Analytics: A Distance-Based Matching Program Assessment Tool

Gina Deom
Indiana University
gdeom@iu.edu

Stefano Fiorini
Indiana University
sfiorini@indiana.edu

ABSTRACT: Some of the most common and effective methodologies used in program assessment are matching methods, which simulate a randomized control experiment as closely as possible by balancing characteristics between a treatment and control group through some distance measure or score (Stuart 2010). In this **practitioner paper**, we present 1) a distance measure that has served well for program assessment analyses at our institution and 2) an internally built RShiny application that provides a user-friendly interface for matching and streamlining program evaluation analyses. This tool has been effective in a variety of applications at our institution, including assessing the impact of instructional changes in the classroom. The code base for the tool, installation instructions, and user guides for various applications are available for download via a public GitHub repository.

Keywords: program assessment, distance-based matching algorithms, RShiny

1 PROJECT BACKGROUND

One of the core roles of any learning analytics researcher is program assessment – gathering and analyzing program data to improve student learning and success. To that end, researchers are called upon to evaluate the impact of campus programs, student interventions, and instructional changes using institutional data. Methods and tools are needed to deliver meaningful and timely evaluation analyses to program practitioners, instructors, and campus leaders.

Some of the most common and effective methodologies used in this space are matching methods (Stuart 2010), and this paper describes our institution's approach to streamlining such analyses for program assessment and learning analytics applications. This process includes choosing a versatile distance measure for the matching process and building a distance-based matching program assessment tool.

2 DISTANCE MEASURE

The foundation of any matched analysis is a distance measure, which quantifies how similar two units of analysis (i.e. students) are to each other. The smaller the distance between the two units, the more similar they are and the greater likelihood that the two units should be paired together in a matched analysis. There are a variety of distance measures available (propensity score, Mahalanobis distance, etc.), but the quantity of requests and types of analyses called for a distance measure that could 1) handle a variety of data types and data challenges (i.e missing data) and 2) force matches on key student subpopulations to account for the nested structure of various aspects of the educational environment (cohorts, academic terms, courses, etc.).

Following a review of various matching methods and distance measures available, we found that a variation of the Gower distance formula (D'Orazio 2021) provides the flexibility and configuration capabilities to serve most of our assessment needs. The figure, table, and footnotes below represent the distance measure in algebraic form and describe the detail of its components. The distance measure is highly intuitive as it quantifies the average distance between any pair of treatment and control units across all matching variables, weighted by the importance of the individual matching variables as set by the researcher. The weighting component of the distance measure is not only effective at creating nested subpopulations in the matched datasets, but it is also valuable for generating contoured analysis datasets that focus on student populations most served by the program or intervention being studied.

$$G_{ij} = \frac{\sum_{v=1}^{p} w_v d_{ijv}}{\sum_{v=1}^{p} w_v}$$

Figure 1: Distance Measure Formula

Table 1: Distance Measure Components

Component	Description
G_{ij}	The combined, calculated distance between the ith treatment unit and the jth control unit
w_v	The weight of the vth matching variable
d_{ijv}	The distance[1] calculated on the vth matching variable for the ith treatment unit and jth control unit treatment pair. More details are provided in the footnote.

[1] For numeric variables, the distance is calculated as the absolute value of the difference between the two units on the variable. Prior to calculating that difference, numerical variables are normalized between 0 and 1 by dividing each value by the maximum value of the variable in the dataset. For categorical variables, the distance is set to 0 if the two units match exactly on the level of the variable and the distance is set to 1 if the two units do not match exactly on the variable. If either the ith treatment or jth control unit have a missing value on the variable, the distance is set to 1 to reflect the uncertainty of that variable in the matching process for the potential matched pair.

3 DISTANCE-BASED MATCHING PROGRAM ASSESSMENT TOOL

Figures 2 and 3 display portions of the distance-based matching program assessment tool that was built to support analytics projects at our institution. This tool is a RShiny application that was built using the statistical software R. It leverages R's MatchIt package (Ho et al. 2011) to select matched samples of treated and control units that are similar with respect to the Gower distance measure described in the previous section. The application's code base, which is downloadable via GitHub (https://github.com/iu-ia-research-analytics/distance-based-matching-program-assessment-tool), only requires installation of R and RStudio and preparation of a CSV file for matching analysis. The specification details of the CSV file are also provided in the documentation on the GitHub site.

After installing R and RStudio and downloading the application code base, users launch the user interface in the RStudio environment. Users load the prepared CSV file into the tool, select matching and treatment variables, and set various aspects of the matching process (variables weights, matching algorithm, the number of controls to match to a treatment unit, etc.). The application then computes the pairwise distances between the treated and control units and runs the matching algorithm to create a matched data set that can be downloaded for further analysis. The application also provides capabilities to check balance on each variable before and after matching. For demonstration purposes, the example provided in the figures uses a subset of a larger variable list that was used to assess the impact of instructional changes in a specific course and academic term (ACAD_TERM_CD) on student outcomes in the subsequent course; the example shows students being matched on gender, their grade in the preceding course (CRS_OFCL_GRD_NBR) and their grade point average in all other courses (GPAO) (Matz et al. 2017) prior to the subsequent course.

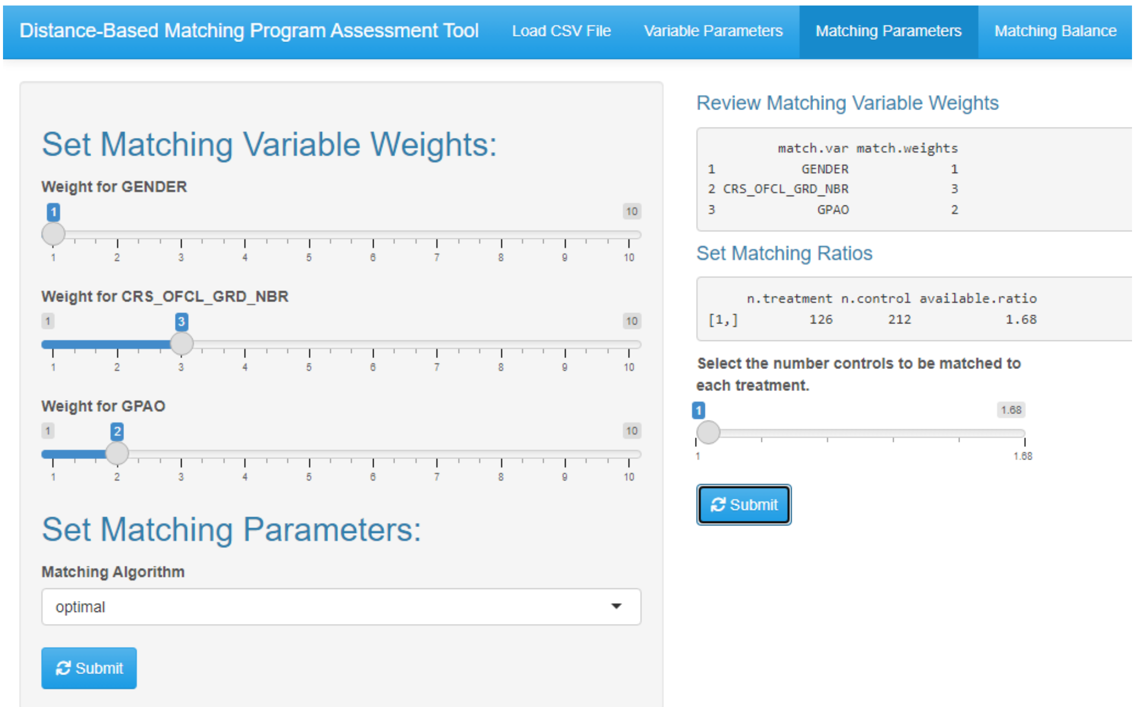


Figure 2: Assessment Tool – Set Matching Parameters

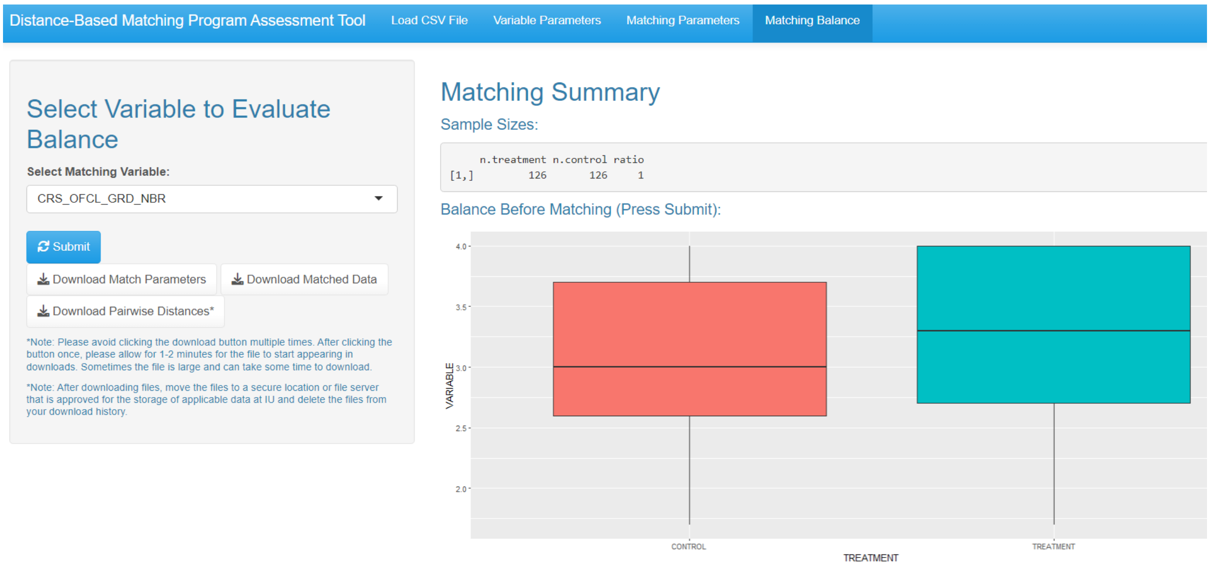


Figure 3: Assessment Tool – Check Matching Balance and Download Analysis Files

4 CONCLUSIONS AND NEXT STEPS

This distance measure and interactive, distance-based matching program assessment tool have been highly useful for our institution and have improved our capacity to deliver analytics to improve campus programs and instruction. The institution is also using the tool to create lists of peer institutions for benchmarking analysis. We are actively expanding the tool to include statistical analysis of selected outcome variables, and we are providing self-guided trainings to various stakeholder groups to maximize the tool's use. These trainings include handouts and user guides to help users install software, prepare data files, and set parameters in the matching application.

REFERENCES

D'Orazio, Marcello. (2021). Distances with Mixed-Type Variables, Some Modified Gower's Coefficients. https://www.researchgate.net/publication/348320805_Distances_with_mixed_type_variables_some_modified_Gower's_coefficients

Ho DE, Imai K, King G, Stuart EA (2011). "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, 42(8), 1–28. doi:10.18637/jss.v042.i08

Indiana University Institutional Analytics Research and Analytics. (2023). Distance-based-matching-program-assessment-tool. Available from https://github.com/iu-ia-research-analytics/distance-based-matching-program-assessment-tool

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of Gendered Performance Differences in Large Introductory Courses at Five Research Universities. *AERA Open*, 3(4). https://doi.org/10.1177/2332858417743754

Package 'MatchIt'. 2022. https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf

Stuart, Elizabeth A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statist. Sci.* 25(1) 1-21. https://doi.org/10.1214/09-STS313

A Peer in the Loop: The Human Touch that Analytics Needs

Jenna Matthews[1,2], Hillary Swanson[1], Morgan Slugantz[3], Kausha Miller[3]
[1]Utah State University, [2]Pearson Education, [3]Bluegrass Community & Technical College
jenna.matthews@pearson.com, hillary.swanson@usu.edu, mslugantz0004@kctcs.edu, kausha.miller@kctcs.edu

ABSTRACT: Efforts to put humans in the analytics loop focus on getting instructor feedback and perspectives. While this is beneficial, there's a human touch that is lacking from both the interpretation and use of learning analytics insights. We discuss the process of identifying a new behavioral metric and involving the voice of a peer in the interpretation and (most importantly) use of the metric for student support. We show preliminary results indicating that both the new metric and peer voice contributed to improved success in a general education math course.

Keywords: Peer mentoring; community college; mathematics; reengagement

1 MOVING FROM A GOAL TO A METRIC

As a group, students attending community college are among those most in need of support. They can be especially isolated, making the connection with peers increasingly important as they often lack the on-campus support of residential students (Crisp, 2010). For older, non-traditional students, difficulty in key courses such as Math, combined with a lack of support and sense of community can lead to withdrawal from school (Bahr et al., 2022). In recent research, Guo et al. pointed out that frequently the withdrawal is not tied to cognitive difficulties, but rather to affective elements, such as relationships with other individuals at the institution, personal factors (such as self-efficacy and mindset), and academic support (2022).

1.1 The Context

This paper reports on work taking place within a long-term research-practice partnership involving a community college in the southeastern United States. The goal of the partnership was to improve student success within a general education mathematics course. This course had historically poor completion rates - made worse by the COVID-19 pandemic - and was required for students completing non-STEM degrees. In many instances, this course stood between students and graduation.

Earlier work had included a variety of efforts - including a partial redesign of the course - without achieving the desired results. So, as a team, we returned to raw activity data to find something which could be a useful early indicator of problems.

Part of the course design involved students working on math problems in an online learning environment. Students were able to work on assignments and quizzes with flexibility - and could stop and return later if needed. Data collected in this environment included a variety of student actions - such as checking their answer, asking for a hint, or working through a guided solution - as well as flags to indicate whether the student got a specific answer correct.

1.1.1 *Defining "Giving Up"*

Using these actions captured in the online environment, we created a metric called "Giving Up" using the following process:

- We started by grouping all student activity into sessions, using the Google Analytics (Google) standard window of 30 minutes of inactivity to split working time into working sessions.

- Within a session, we identified each problem with which the student interacted. These interactions could include checking their answer, asking for a hint, or working through a guided solution. In order to capture those problems where students were working, not just looking (a single interaction), we limited the data set to those problems with multiple interactions.

- Using the problems with which the students interacted from step #2 as the denominator and the number of solved problems as the numerator, we came up with a percentage of problems which were solved. The inverse of solved problems represents those with which students interacted, but which they did not solve before ending their session. That inverse was the session-level "Giving Up" percentage.

In this way, we had a percentage score which could be tracked across time, allowing us to see both fluctuations (where a score would alternate between climbing and falling) and trends (where a score would change in the same direction – climbing or falling – across several sessions).

We reviewed previous semesters of the course using this metric and saw significant differences between successful (A/B/C grades) and failing (D/F/W grades) students in terms of their average change between sessions – whether a subsequent session had a higher, lower, or unchanged "Giving Up" percentage.

We also reviewed the "Giving Up" behavior over time for students who ultimately withdrew from the course. We found that individual sessions with high rates of "Giving Up" were unimportant - and nearly universal. Fluctuations from one session to the next were too noisy to be useful for flagging students in need of support. However, a trend of increased "Giving Up" across three sessions worked as an early indicator – giving support staff weeks of time before students withdrew from the course. Based on these findings, we used this trend as a flag for student outreach.

2 A PEER IN THE LOOP

For the semester, we focused on a single course on a traditional 16-week schedule with a total of 267 students. No other changes were made to either the course content or delivery during the semester.

In previous research covering a range of both ages and disciplines, mentorship by near-peer mentors - fellow students who have slightly more experience or knowledge - has been shown to be a sustainable option for improving student engagement, interest, and academic outcomes (Clarke-Midura et al., 2018; Pluth et al., 2015; Tenenbaum et al., 2014). This research provided our rationale for including a peer guide on the project team, and more importantly having the peer guide conduct the outreach. Our project team included a researcher/data scientist, an assistant dean/coordinator at the college campus, and a peer guide - a student employee at that college campus who met with our team weekly and handled the student outreach. For both the data scientist and coordinator, this project was part of their regular work, requiring no additional funding. During weekly meetings, the researcher would present/explain data and the instructor would lead the prioritization process and give the “teacher version” of an outreach message for students.

The peer guide participated in the meetings, gave feedback on the metrics, did the message translation (going from a “teacher voice” to a “student voice”), and conducted the actual student outreach. During the week, the peer guide reached out to students via email, text, and phone. While there was some prioritization of methods based on the severity of need, the peer guide was encouraged to adapt the mode, method, and message as preferred. Outreach to students was very open-ended, offering support to help students re-engage and finish the course successfully.

3 REVIEW & RESULTS

At the end of the semester, we reviewed data for outreach, activity, and outcomes. During the semester, 159 distinct students received a total of 331 outreach messages. Even though most students never responded, we did see an impact in both individual student behavior and course success rates related to the outreach. Individual student behavior, measured as a reduction in “Giving Up” (or working on problems but leaving them unsolved) for the session after outreach, was especially meaningful for several student groups as shown in Table 1 below.

Table 1: Reduction in “Giving Up” behavior – average percentage change in subsequent session.

	Post-Outreach Records			Not Post-Outreach Records			
Group	Improvement	Students	Records	Improvement	Students	Records	p-value
First-Generation	6.01%	46	72	0.91%	108	5059	0.1314
African American	7.83%	14	33	-1.00%	37	1755	<0.0001***
Pell-Eligible	8.09%	83	132	0.96%	186	8573	0.0091***

We also saw a large improvement in the course-wide success rate (% of students receiving an A/B/C grade in the course) as shown in Table 2 below:

Table 2: Difference in course success rates.

Category	Fall 2019	Fall 2020	Fall 2021
Students	201	242	267
Success Rate	45.27%	37.19%	46.07%
Change Year-over-Year		-8.08%	+8.88%

The success rate for Fall 2021 – the semester during which we ran this project – put this course slightly above their pre-COVID numbers. These results led to an expansion of the work with additional courses and campuses joining the ongoing project.

4 DISCUSSION

As in previous learning analytics studies we are tracking a behavior ("Giving Up") rather than performance or outcomes, to identify students who may need help. It is a behavior which was defined as part of ongoing research-practice partnerships, and which has not before been available to faculty, advisors, or other members of the student support teams.

Perhaps more importantly, however, is that the outreach based on this behavior is coming from a peer, in a student voice. We have seen results which suggest that peer outreach may have some of the same benefits as those already identified in peer mentoring (Clarke-Midura et al., 2018; Pluth et al., 2015; Tenenbaum et al., 2014). Not only was the peer guide able to translate from a teacher to a student voice, but the outreach was open-ended. One of the things we found during our pilot was that students' struggle was not necessarily academic but frequently related to their life outside of school, including issues related to employment, illness, or family problems. In these instances especially, arbitrarily sending students to a tutor would not have been helpful because it wasn't the math causing the problems.

We have since expanded the usage of this metric to additional courses and campuses where we are continuing to see similar improvements in behavior and course outcomes. Ongoing and planned analysis includes understanding the impact of different methods of student contact, comparing behavior across a larger time frame, and identifying the most effective window for peer outreach.

REFERENCES

Bahr, P. R., Boeck, C. A., & Cummins, P. A. (2022). Is Age Just a Number? A Statewide Investigation of Community College Students' Age, Classroom Context, and Course Outcomes in College Math and English. *Research in Higher Education*, *63*(4), 631–671. https://doi.org/10.1007/s11162-021-09660-w

Clarke-Midura, J., Poole, F., Pantic, K., Hamilton, M., Sun, C., & Allan, V. (2018). How Near Peer Mentoring Affects Middle School Mentees. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 664–669. https://doi.org/10.1145/3159450.3159525

Crisp, G. (2010). The impact of mentoring on the success of community college students. *The Review of Higher Education*, *34*(1), 39–60.

Google. (n.d.). *How a web session is defined in Universal Analytics*. Google Support. https://support.google.com/analytics/answer/2731565?hl=en#zippy=%2Cin-this-article.

Guo, Y., O'Halloran, K. P., Eaker, R. M., Anfuso, C. L., Kirberger, M., & Gluick, T. (2022). Affective Elements of the Student Experience That Contribute to Withdrawal Rates in the General Chemistry Sequence: A Multimethod Study. *Journal of Chemical Education*, *99*(6), 2217–2230. https://doi.org/10.1021/acs.jchemed.1c01227

Pluth, M. D., Boettcher, S. W., Nazin, G. V., Greenaway, A. L., & Hartle, M. D. (2015). Collaboration and Near-Peer Mentoring as a Platform for Sustainable Science Education Outreach. *Journal of Chemical Education*, *92*(4), 625–630. https://doi.org/10.1021/ed500377m

Tenenbaum, L. S., Anderson, M. K., Jett, M., & Yourick, D. L. (2014). An Innovative Near-Peer Mentoring Model for Undergraduate and Secondary Students: STEM Focus. *Innovative Higher Education*, *39*(5), 375–385. https://doi.org/10.1007/s10755-014-9286-3

Connecting Pedagogical Design to Intervention: A Large Scale Implementation of Embedded Learning Analytics

Amir Winer
The Open University of Israel
amirwi@openu.ac.il

Nitza Geri
The Open University of Israel
nitzage@openu.ac.il

ABSTRACT (Presentation): In most institutions Learning Analytics (LA) initiatives emerge out of small-scale research projects. We present a practitioner-led large-scale implementation of embedding LA within a MOODLE Virtual Learning Environment (VLE) across all 800 courses. Our presentation shows the Human-Centered approach to both the conceptual design and the developmental stages of the LA solution at The Open University of Israel. We specify how embedded LA can better connect the pedagogical design of online courses to timely interventions throughout the semester. We begin with the pedagogical guidelines for our course design. We then outline the ethical concerns, constraints, and opportunities that were created by an LA committee that included student representatives and university stakeholders. We then present the five developmental stages of our LA solution: (1) visually presenting the pedagogical intent of course developers(2) instructors' ability to seamlessly create course-level interventions from within the learning analytics layer; (3) instructors' ability to identify students at risk and approach them with personalized support; (4) generating automatic interventions for faculty and students based on pre-defined criteria; (5) future efforts to use machine learning for identifying and promoting effective interventions. We conclude with adoption, challenges, limitations, and future directions.

Keywords: Embedded Learning Analytics, Pedagogical Design, Intervention, Human-Centered Learning Analytics (HCLA), Virtual Learning Environment (VLE)

1 THE PEDAGOGICAL DESIGN OF DIGITAL COURSES

The foundational triadic connection between a Virtual Learning Environment (VLE), the pedagogical design of courses, and learning analytics is the focus of this presentation. Learning Management Systems (LMS) or Virtual Learning Environments (VLEs) are platforms that allow institutions to present the learning materials of digital courses and capture students' learning activities. The core development of VLEs is usually external whereas instructors can make some adjustments and configurations according to their pedagogical preferences. Lockyer, Heathcote, and Dawson (2013, p. 1,439) define learning design as a "form of documentation of pedagogical intent that can provide the context for making sense of diverse sets of analytic data." They specify that the course design usually includes resources, tasks, and support mechanisms to assist in the evaluation and completion of the learning tasks. Learning Analytics can be defined as the "measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Gašević, 2012, p. 1). The strategic nature of the LA initiative led to the adoption of an organizational approach to this triadic challenge.

Linking pedagogical intentions to the measurement of expected learning outcomes was at the core of our organizational LA initiative. We began with internal development of a new MOODLE course format that captures and presents a digital version of all the learning materials. Sequencing the courseware into bite-size content chunks allowed for a modern presentation that mimics MOOC (Massive Open Online Courses) course design. A formative assessment layer was added to ensure that tracking was directed toward completion and actual understanding of content rather than technical student engagements. Instructors could now define completion criteria for every sub-topic and provide automatic feedback on students' learning progress.

2 FROM ETHICAL DILEMMAS TO EMBEDDED LA

The dangers of taking a reductive view of learning analytics are well documented in current research (Shum, 2019). To avoid the known concerns (Macfadyen et al., 2014), while focusing strategically on fulfilling the goal of the LA initiative, a LA committee was assigned by the university's rector. Its members included various stakeholders, both academic and administrative, as well as student representatives. The first task was to choose the relevant data for analysis. There are two types of student data: static data which is not expected to significantly change over time, and dynamic data which reflects current learning. Static data includes admission records such as high school grades, financial support, and basic demographic data such as age, gender, and address. Whereas static data is mostly associated with admission data, VLE is mainly used to store dynamic learning data that reflects traces of actual learning. We acknowledged the ethical concerns raised by Scholes (2016) about using static data to flag students as “high risk” regardless of their actual progress, even before they began their studies. Therefore, we limited the LA initiative to dynamic data (Winer & Geri, 2019).

Ransbotham et al. (2015) outline three types of analytics: Descriptive, Predictive, and Prescriptive. We defined descriptive LA as visual manifestation of the pedagogical intent that is captured in the VLE. The predictive perspective of LA can provide a future estimation of quantifiable goals based on current measurements. Finally, we used prescriptive LA as support measures or as a pedagogical call-to-action to proactively address deviations from the course design. Concerning predictive analytics, Pasquale (2015) warns about the rise of a black-box society. He shows that algorithms are indeed able to predict individuals’ behavioral patterns, however, they remain unintelligible. In the context of higher education, this might impair the ability to identify the exact source of risk. Moreover, for instructors, predictive analytics poses two additional sources of resistance to assimilating AI technologies into current teaching practices: Ethical concerns (Winer & Geri, 2019) and a general humanistic concern that focuses on the human-machine imbalance (Dimitriadis et al., 2021; Shneiderman, 2020). Therefore, we limited our initial offering to descriptive and prescriptive LA.

Macfadyen et al. (2014) emphasized the need for policy and strategy as a means for embracing the complexity of educational systems that plan to adopt LA. However, along with organizational commitment to change, we also noticed the usability and user experience drawbacks of current external dashboards. Herodotou et al. (2019) indicate that lecturers find it hard to filter relevant information from their analytics tools and access relevant data about each student. Moreover, despite the perceived usefulness of LA visualizations, lecturers have difficulties in connecting LA information to concrete interventions. Therefore, we embeded LA practices into the current VLE processes. Specifically, we created LA functionality as an integral part of the existing teaching practices, rather than an external isolated system. We called this approach "Embedded LA".

The LA committee serves as a change facilitator and as a hub for resolving technical political and pedagogical dilemmas. For instance, when we debated about in-house sourcing versus adopting an external analytics tool, we chose to distinguish between Tableau as a long-term research tool for senior management and administrative purposes and MOODLE LA as a small-scale tool for instructors' pedagogical interventions. The LA committee generated broader circles of commitment and participation. We formed a team of early adopters to pilot new LA functionalities and experiment with course designs. Multiple professional development seminars were crafted to train and assimilate the new LA practices and policies.

3 THE FIVE DEVELOPMENTAL STAGES OF THE LA SOLUTION

The first stage of the LA model was named "visual reflections". Its main purpose was to facilitate a visual comparison between pedagogical intentions of the course design and actual performance of students. We used human-centered design practices to explore the user experience of instructors. The challenges of data literacy were quickly singled out as the key barrier to large-scale adoption. Instructors wanted to teach rather than be taught how to become data scientists. Therefore, to make the LA as seamless as possible, we made an effort to make the extra layer of data almost imperceptible and fully embedded into familiar teaching practices. Consequently, the new visual modes of data presentation, were instantly regarded as familiar and understandable.

The second stage, "course level support", focused on instructor ability to quickly identify engagement difficulties with asynchronous courseware and apply appropriate course-wide interventions. Instructors use these features to improve their formative assessments and offer content-driven support on difficult issues. Instructors raised concerns about LA turning into a "big brother". Hence, we limited the LA design to an aggregated view of students' progress and difficulties. Additionally, only dynamic data was presented to avoid instructor's ability to filter student demographics. We then agreed on a limited set of student dynamic data that would be transferred into the organizational Business Intelligence (BI) platform for management analysis and decision-making. Finally, we assured the teaching staff that their LA activities would not be monitored by senior management.

In the third stage, "Instructor's Personalized Support", we identified four dimensions that could serve as early indications of students at risk: student engagement with the VLE, content progress, assignment submissions, and participation in tutoring sessions. For each dimension, we designed dashboards, and timely tailored interventions to address these issues as soon as possible. The fourth stage, "Automated Personalized Support", is used to replace instructors' interventions with automatic support. Given the superior effect of the instructor's interventions, we limited automatic support to address technical issues, such as assignment submission reminders. To avoid any confusion, we clearly stated that in these cases, a bot was now approaching students rather than their instructor. We are currently evaluating the effectiveness of automatic alerts for instructors. The fifth stage, "Explainable Machine Learning support", is still facing strong resistance given the current difficulties to overcome the inherent bias of these technologies and the fear of machine-based categorizations of our students.

4 ADOPTION, CHALLENGES, LIMITATIONS, AND FUTURE PLANS

The development and adoption of LA is still raising periodic tides of concern and resistance. However, the new course format in the familiar VLE along with a modern pedagogical approach to digital course design and embedded LA were accepted as an unbreakable whole that is backed by the university leadership. The triadic connection of VLE, pedagogical design, and LA was gradually rolled out to all the instructors in a centralized manner. Early adopters provided initial feedback and supported the agile development methodology. Once the functionalities of the new course format were set in motion the piloting instructors changed their face-to-face teaching practices into hybrid formats allowing for a better balance between synchronous and asynchronous teaching. Currently, about 25% of our 800 courses have been changed to the new course format. These courses belong to diverse disciplines: Computer science and mathematics, humanities, natural and life sciences, and social sciences. The substantial increase in student course engagement and satisfaction is a promising indication for our long-term mission to migrate all courses to this new format. The LA capabilities are present in all the university courses. However, only courses that migrated to the new course design can benefit from the alignment between their pedagogical intent to the course's LA. Finally, we are one stage before explainable machine-learning LA. These technologies are only applicable to very large courses and cannot be used for most courses. Moreover, machine-learning LA reflect the known dilemmas of the digital revolution in higher education. These changes are set to radically modify the current human-machine balance in the coming years. Therefore, until a strong use case is available, we cautiously present the fifth stage as an organizational call for further exploration.

REFERENCES

Dimitriadis, Y., Martínez-Maldonado, R., & Wiley, K. 2021. Human-centered design principles for actionable learning analytics. In T. Tsiatsos, S. Demetriadis, A. Mikropoulos, & V. Dagdilelis (Eds.), *Research on e-learning and ICT in education* (pp. 277-296). Springer. https://doi.org/10.1007/978-3-030-64363-8_15

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational Technology Research and Development, 67*(5), 1273-1306. https://doi.org/10.1007/s11423-019-09685-0

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

Pasquale, F. (2015). The black box society. Harvard University Press.

Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). *Minding the analytics gap*. MIT Sloan Management Review, *56*, 63-68.

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development, 64*(5), 939-955.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495-504. https://doi.org/10.1080/10447318.2020.1741118

Shum, S. B. (2019). Critical Data Studies, Abstraction & Learning Analytics: Editorial to Selwyn's LAK keynote and invited commentaries. *Journal of Learning Analytics*, *6*(3), 5-10.

Siemens, G., & Gašević, D. (2012). Special Issue on Learning and Knowledge Analytics. *Educational Technology & Society, 15*(3), 1–163.

Winer, A., & Geri, N. (2019). Learning analytics performance improvement design (LAPID) in higher education: Framework and concerns. *Online Journal of Applied Knowledge Management*, 7(2), 41-55. https://doi.org/10.36965/OJAKM.2019.7(2)41-55

Integrating Competency-Based Education in Interactive Learning Systems

Maximilian Sölch, Moritz Aberle, Stephan Krusche
Technical University Munich
maximilian.soelch@tum.de, moritz.aberle@tum.de, krusche@tum.de

ABSTRACT: Artemis is an interactive learning system that organizes courses, hosts lecture content and interactive exercises, conducts exams, and creates automatic assessments with individual feedback. Research shows that students have unique capabilities, previous experiences, and expectations. However, the course content on current learning systems, including Artemis, is not tailored to a student's competencies. The main goal of this paper is to describe how to make Artemis capable of competency-based education and provide individual course content based on the unique characteristics of every student. We show how instructors can define relations between competencies to create a competency relation graph, how Artemis measures and visualizes the student's progress toward mastering a competency, and how the progress can generate a personalized learning path for students that recommends relevant learning resources. Finally, we present the results of a user study regarding the usability of the newly designed competency visualization and give an outlook on possible improvements and future visions.

Keywords: competency-based education, adaptive learning, learning analytics

1 INTRODUCTION

A successful and individual learning experience requires that students receive formative feedback and insights into their learning progress in a course (Scheffel et al., 2014). Students should be able to adapt the course content based on their competencies and needs. Instructors want to know whether their students gained specific competencies while attending their course. However, such functionality is unavailable or works poorly in existing learning management systems.

Artemis allows instructors to create a scalable and productive student-instructor setting, even for extensive or remote-only courses. The platform aspires to reduce the overall workload for instructors with large audiences without compromising the benefits of an interactive learning environment (Krusche et al., 2017; Krusche & Seitz, 2018). It focuses on active student participation using automated or manual software-assisted grading and iterative feedback. In addition, Artemis supports lecture units, such as video, text, file, online, and exercise units. More than 10,000 students from several universities across Germany and Austria actively use Artemis (Krusche, 2021).

2 APPROACH

Instructors should be able to teach their courses using a fundamentally different approach. With competency-based education, they can outline the desired student abilities at the end of the course and then provide content to help students reach these learning objectives (Curry & Docherty, 2017). A learning system must track students' progress and allow instructors to define relations between different competencies. Instructors can create a graph network of learning objectives by marking a competency as a prerequisite or adding more detailed sub-topics to a competency. These dependencies between competencies allow students to find a suitable and personalized learning path throughout the defined competencies to progress in the course individually. In addition, the learning system needs to visualize the progress toward mastering a competency for the students and the course instructors.

2.1 Competency Relations

To allow instructors to define relations between different competencies, we designed and implemented the user interface shown in Figure 1. Instructors can create directed relations between exactly two competencies of the same course by specifying the relation's head and tail competency as well as the relation's type. A relation can have one of the following types: *assumes*, because a competency can require that another competency is mastered first, *extends*, because a competency can add new aspects to another competency, *relates*, because a competency can be connected to another competency, and *matches*, because a competency can be identical to another competency. A relation cannot be reflexive and no more than one relation of each type can exist between two competencies. To get a better overview of all the existing relations between a course's competencies, Artemis generates a competency relation graph. This graph updates when an instructor adds or removes a relation.

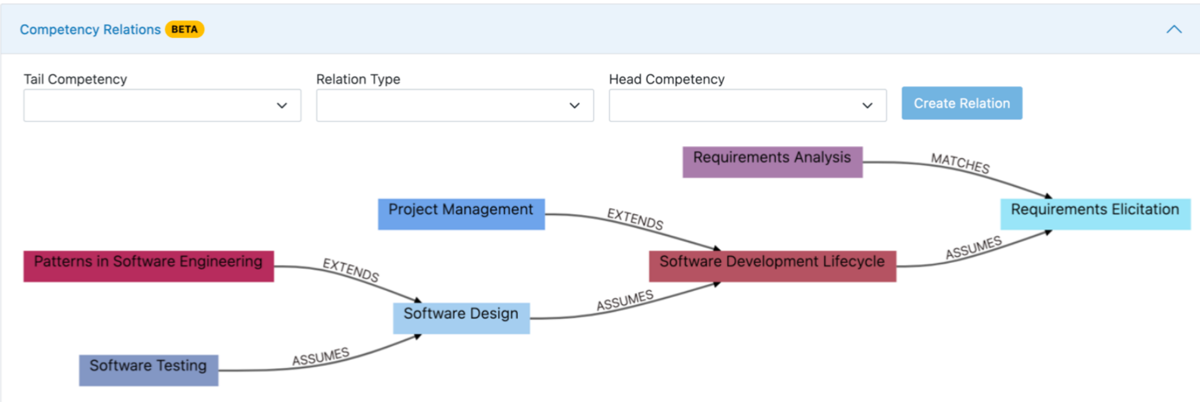


Figure 1: Defining relations between competencies in Artemis using a competency graph

2.2 Learning Analytics

To measure the student's progress in a course, Artemis tracks the active participation in exercise units and uses a checkbox for visualizing the completion status of a lecture unit. Besides the visual indication, the checkbox functions as a button for the user to manually toggle the completion status at the same time. Artemis also marks lecture units as completed automatically depending on their type when the following conditions are met: Once the user clicks the button to download an attachment file, the corresponding file unit is considered complete. Text units are completed as soon as the student clicks on the unit to open the collapsed text. Students automatically complete an online unit when they click on the link forwarding them to the external website. Video units are marked as completed five minutes after clicking on the unit to reveal the embedded video. In addition to the completion status, Artemis also provides different exercise statistics for students to track their performance and compare themselves to the course average and for instructors to get insights into their course's performance.

2.3 Different Metrics for Competencies

Artemis calculates different metrics and displays them to the students allowing them to get a quick overview of their learning progress regarding a competency. They can track their mastery advancement for each competency by viewing the progress value, calculated from the number of completed learning resources linked to the competency, and the confidence value, composed of the average score for all exercises linked to the competency. A competency is considered mastered if the progress is at 100% and the confidence value is equal to or higher than the defined mastery threshold of the competency. Inspired by the Apple Watch's fitness rings, Artemis shows the user three individual progress bars in a circle depicted in Figure 2.

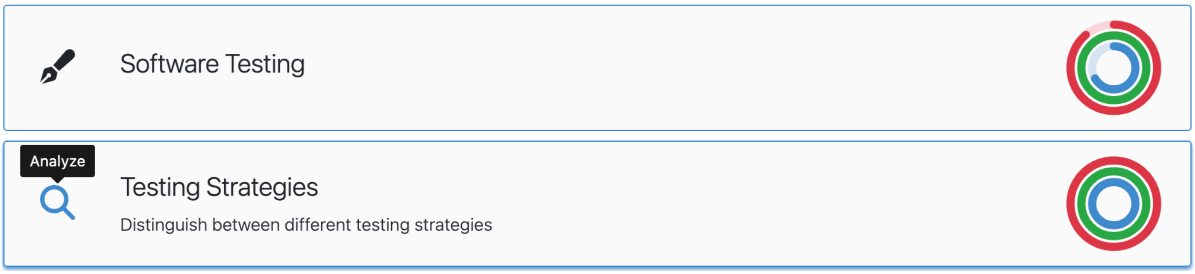


Figure 2: Visualization of competencies and a student's progress in Artemis

The innermost blue ring visualizes the overall progress P, which is the percentage of completed lecture units and participated exercises linked to the competency. The second, green progress bar shows the confidence level C of the competency in proportion to the threshold value T (mathematically $C * \frac{1}{T}$). As an example, if the student's latest confidence level equals 60% and the mastery threshold is set to 80%, the ring would be 75% full. The outermost red ring represents the advancement toward mastery as a weighted function (with $w = \frac{2}{3}$) of progress and confidence defined as $(1 - w) * P + w * C * \frac{1}{T}$. If a student has mastered the competency without completing all linked learning resources, we override the progress of the mastery ring to 100% nevertheless. Competencies can be categorized according to Bloom's revised taxonomy (Krathwohl, 2002). In the visualization we use different icons that support a quick identification of a competency according to the taxonomy. Instructors can link competencies to learning resources as prerequisites or as learning goals. In combination with the competency relation graph and the student's competency mastery, Artemis uses these links to recommend suitable learning resources and generate a personalized learning path for students.

3 EVALUATION

To evaluate the newly designed visualization of competencies and a student's progress in Artemis, we conducted the short version of the User Experience Questionnaire (UEQ-S) with 7 participants (Schrepp et al., 2017). All of them are students randomly invited from a TUM university course and are familiar with Artemis. During the UEQ-S the participants rate pairs of terms with opposite meanings on a 7-point Likert scale and their answers are scaled from -3 to +3.

The results of the UEQ-S attest the proposed visualization a good usability overall with a mean of 1.37. In particular, the results show an excellent hedonic quality with a mean value of 2.06 indicating that the user interface is appealing and not boring. The pragmatic quality, which relates to the practicality and functionality of a user interface, is slightly lower with a mean value of 0.714 and shows potential for additional improvements. One possible improvement could be to add a legend explaining the meaning of the rings as well as the icons. During the evaluation, most of the participants mentioned that this improvement would have made the visualization clearer and easier to understand. Therefore, we plan to add an informative legend to the visualization before we release it to the public.

4 CONCLUSION

We presented an approach to integrating competency-based education into learning management systems, specifically for Artemis, including a new and innovative visualization of a student's progress toward mastery of a competency. To evaluate this new visualization, we conducted a UEQ-S that showed promising results for the user experience.

Future work includes improving the calculation of metrics for competencies. For example, mastery of competencies might slowly decline after the final exam. Artemis could account for this fact and model knowledge decay within the adaptive learning system (Bergeron, 2014). Another potential area of improvement is the visualization of learning paths for students and instructors. Artemis uses a graphics-based interface to show an instructor the relationships between all competencies in the proposed system. However, it is difficult for the user to derive all possible learning paths from this representation. Artemis does not yet represent the learning paths chosen by students in the course. Visualizing how certain groups are progressing in the competencies would be a fascinating insight for the instructor in the context of learning analytics.

REFERENCES

Bergeron (2014). Application of learning record stores and other forms of electronic competency records in modeling competency degradation. *Studies in Health Technology and Informatics*, *196*, 23–28.

Curry & Docherty (2017). Implementing Competency-Based Education. *Collected Essays on Learning and Teaching*, *10*, 61.

Krathwohl (2002). A Revision of Bloom’s Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218.

Krusche (2021). *Interactive learning - A scalable and adaptive learning approach for large courses.* [Habilitation, Technical University Munich].

Krusche & Seitz (2018). ArTEMiS: An Automatic Assessment Management System for Interactive Learning. *49th ACM Technical Symposium on Computer Science Education*, 284–289.

Krusche, Seitz, Börstler & Bruegge (2017). Interactive Learning: Increasing Student Participation through Shorter Exercise Cycles. *19th Australasian Computing Education Conference*, 17–26.

Scheffel, Drachsler, Stoyanov & Specht (2014). Quality Indicators for Learning Analytics. *Journal of Educational Technology & Society*, *17*(4), 117–132.

Schrepp, Hinderks & Thomaschewski (2017). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *Int. Journal of Interactive Multimedia and Artificial Intelligence*, *4*, 103.

OLI Torus: a next-generation, open platform for adaptive courseware development, delivery, and research

Norman Bier[1], John Stamper[1], Steven Moore[1], Darren Siegel[1], Ariel Anbar[2]
Carnegie Mellon University[1], Arizona State University[2]
{nbier, jstamper, stevenmo, darrensiegel}@cmu.edu, anbar@asu.edu

ABSTRACT: Torus is the next-generation of the Open Learning Initiative (OLI) platform, providing an open, community-based system for designing, delivering, improving, and researching adaptive courseware. The project has attracted a diverse array of active partners, users, and contributors, notably ASU's Center for Education Through Exploration (ETX). Current capabilities support and interweave OLI and ETX approaches, with expanded features in development. This report discusses the Torus platform and community.

Keywords: Adaptive Courseware, Learning Analytics, Learning Engineering, OER, OSS

1 INTRODUCTION

For two decades, Carnegie Mellon University's OLI has been a leader in science-informed adaptive courseware, developing learning environments designed to improve outcomes while providing a testbed for learning research. Multiple, rigorous studies demonstrate the potential of OLI courseware to improve student outcomes while enacting research (e.g. Kaufman et al., 2013; Bie et al., 2019; What Works Clearinghouse, 2020). The OLI platform has been a key element in this success. This tool for creating and delivering online instruction embeds core learning science principles into the system's design. The platform ensures an instrumented experience, capturing rich learner interaction data in conjunction with semantic context to support feedback for learners, educators, authors, and researchers (Moore et al., 2020a). The richness of this data stream is foundational for OLI's pioneering work in learning analytics, best seen in the Learning Dashboard (Lovett, 2012; Bier et al., 2014).[1] This instructional intelligence system provides educators with per-learning-objective measurements of learning as well as more detailed information, including sub-skill summaries, individual student learning estimates, and guidance on student struggles.

The platform has seen expansive use, with enrollments from more than 5 million independent learners and over 750,000 enrollments from hundreds of academic institutions. 6,180 educators have instructor accounts; authoring and improvement tools host 900 users, who include faculty, learning engineers, authors, and instructional designers, developing more than 300 courses. The platform also supports a large research community via seamless integration with DataShop (Koedinger et al., 2013) and LearnSphere (Liu et al., 2017); 700 researchers access OLI datasets for primary and secondary analysis. With this use has come many lessons learned about the system's limitations, particularly architectural assumptions; alternate R&D models; community; outreach, and licensing. As the legacy

[1] Lead by Marsha Lovett, the Learning Dashboard was funded by the Spencer Foundation and developed at CMU by Judy Brooks, Bill Jerome, John Rinderle, Ross Strader, and Candace Thille. See oli.cmu.edu/educators/the-learning-dashboard/

technology has grown more dated and new technologies have emerged, the legacy platform's limitations have become even more glaring.

To meet these challenges, OLI has launched Torus as its next-generation platform, built on OLI's successes and integrated with CMU's larger learning engineering ecosystem. Launched in 2020, and developed under an open, business-friendly MIT license, the effort aspires to attract a broad coalition of post-secondary and industry participation, maintaining essential elements of OLI's success while expanding options for delivering and researching learning. Key Torus goals include:

- Addressing lessons learned from the legacy platform, across multiple dimensions, including cost, effectiveness, ease-of-development, and community engagement.
- Delivering on OLI's commitment to grow as a "community-based research activity," broadening participation in learning engineering (Thille & Smith, 2011).
- Promoting active exploration of new approaches to pedagogy, learning, and analytics.
- Supporting the next two decades of OLI success.

This report to LAK reflects on Torus progress, previews plans and describes ways to engage.

2 OLI TORUS

Torus is the latest iteration of the OLI platform, updating and expanding capabilities for developing, delivering, and improving adaptive courseware while providing a workbench for learning science research. Launched in 2020, the effort is informed by a number of goals. Foundational to the project is OLI's need to replace its legacy system; while the platform has effectively served OLI since its first use in 2006, it has become increasingly restrictive. Torus development was launched as an open effort, reflecting OLI's open philosophy and also as an attempt to build trust among users who had suffered from prior vendor lock-in and to invite broader participation. Torus has also been architected to be pedagogically agnostic. OLI's scientific agenda demands humility and an acknowledgment of how much is still unknown about human learning, creating clear requirements for Torus to support alternate approaches, particularly as they can be implemented for fresh investigations. Finally, the new platform is informed by a number of technical requirements, balancing lessons learned with the need to build a robust, cloud-native codebase that can be effectively developed and maintained by a relatively small development team. The Torus technical stack was carefully selected with these requirements in mind. These initial requirements have already drawn exceptional interest from industry and academy collaborators; Torus has quickly grown into an open-source, community-based project with the promise of an even larger community.

Early, enthusiastic participation from Arizona State University's Education Through Exploration Center (ETX)[2] has accelerated progress towards a broader, open-source community with a shared commitment to science-based courseware. ETX is a leading developer of highly immersive, adaptive courseware, with millions of dollars invested to develop materials on the Smart Sparrow platform, and commitments for additional development from a variety of grants (e.g., Horodyskyj et al., 2018; Mead et al., 2019). Pearson's 2020 acquisition of Smart Sparrow left ETX's prior work and future commitments at risk. The Inspark Teaching Network,[3] closely aligned with ETX, faced the same challenges. To address this risk, ETX became a key partner in the Torus community, recapitulating

[2] https://etx.asu.edu
[3] https://inspark.education

Smart Sparrow capabilities and migrating existing content to the new platform. At the close of 2022, ETX is central to the now-shared Torus effort, with all its content (and Inspark's) being delivered on an open Torus platform that provides feature parity with the now-defunct Smart Sparrow platform. Additional users and contributors include the State University of New York (SUNY) system, KTH Royal Technical Institute of Stockholm, Unicon, and WyeWorks. This growing community boasts an array of post-secondary and industry participants that have jointly invested $8.3 million in the platform.

In its current form, the Torus project demonstrates the viability of both the platform and the larger community-based approach. Over the past year, Torus has received open code contributions from more than 8 universities and companies; during that time, the platform has served 35,000 enrollments via 380 educators on 90 campuses. The platform has also seen expansive use in supporting independent enrollments and experiments. New and migrated content is being used by OLI for thousands of students, in courseware representing many domains. ETX is now developing new courseware on the platform. In the coming year, this use is projected to grow dramatically, with an estimated 110,000 enrollments as development and migration continue. The platform instantiates a three-tier client server architecture, combining a Postgres backend, an Elixir/Phoenix application layer, and an HTML/React/LiveView presentation tier. This technology stack was chosen to accelerate development, scale, leverage cloud-native capabilities, aggressively address accessibility, and drive engagement with a broader open-source community. Torus is built for scale in multiple dimensions and architected to support two-orders-of-magnitude growth from our legacy system. Our priorities in developing Torus have emphasized elements that will expand our user base by targeting historically limiting factors in adoption and contribution. Torus is under active development, with a focus on finalizing migration from the legacy system and a longer-term roadmap with a host of new features and affordances. Torus supports a growing array of analytics, supporting course designers, authors, and students. The authoring environment has already recreated key improvement views from the legacy system, allowing designers to more easily identify areas to be refined (Bier & Jerome, 2012). The system also provides direct connections to DataShop and LearnSphere, expanding analytic capabilities. New features include an audit framework to support detailed pre-release analysis of courseware; existing pedagogical, accessibility, and content audits are already in place and being actively refined, and new equity audit approaches are in development. The platform also addresses the legacy platform's analytics implementation, which offered a rigid approach that made expansion of and research into learning analytics a challenge (Bier, et al., 2014). Torus is designed to support a modular approach, separating prediction engine, domain and learner models, and visualization components to support a broader set of approaches and investigations.

3 Looking Ahead and Getting Involved

The Torus codebase,issue tracking, documentation, etc.. are available on GitHub[4]. OLI's production Torus instance is available to Interested developers, educators, and researchers are encouraged to explore the codebase and system[5]. A host of current projects ensure that the platform is rapidly developing and evolving. Support from Schmidt Futures and NSF are driving more instrumentation, A/B testing, and adaptive capabilities planned for 2023.[6] Investments from the Bill and Melinda Gates

[4] see https://github.com/Simon-Initiative/oli-torus

[5] can be accessed at http://proton.oli.cmu.edu

[6] See https://www.cmu.edu/news/stories/archives/2022/september/classroom-experimentation-stamper.html for details.

Foundation are also supporting new platform innovations, emphasizing equity-advancing features.[7] Support from NSF is also allowing research on learner sourcing content (Moore et al., 2021b) and integrating large language models (LLMs) like GPT3 (Moore et al., 2022). Monthly community meetings are held to announce new efforts, refine the Torus roadmap, and explore community needs. Interested educators, researchers, and developers are encouraged to participate.[8] This testbed for deploying and investigating learning analytics will be of ongoing interest to the LAK community.

REFERENCES

Bier, N. and Jerome, W. (2012). Learning Data Visualization. Open Education Annual Conference. Vancouver, Canada, October 17, 2012.

Bier, N., S. Moore, & M. Van Velsen (2019). Instrumenting courseware and leveraging data with the Open Learning Initiative (OLI). Companion Proceedings of 9th International LAK Conference.

Bier, N., Strader, R., & Zimmaro, D. (2014). An approach to Skill Mapping in Online Courses. Learning with MOOCs, Cambridge, MA.

Horodyskyj, L.B., Mead, C., Belinson, Z., Buxner, S., Semken, S., Anbar, A. (2018). Habitable Worlds: Delivering on the promises of online education. Astrobiology 18, no. 1. pp.86-99.

Kaufman, J., Ryan, R., Thille, C., & Bier, N. (2013). Open Learning Initiative Courses in Community Colleges: Evidence on Use and Effectiveness. Carnegie Mellon University, Pittsburgh, PA.

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

Lovett, M. (2012, July). Cognitively informed analytics to improve teaching and learning. EDUCAUSE Sprint, July 25, 2012: http:// tinyurl.com/r2mcnm59

Liu, R., Koedinger, K., Stamper, J., & Pavlik, P. (2017). Sharing and reusing data and analytic methods with LearnSphere. In Workshop and Tutorials Chairs (p. 475).

Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S. & Anbar, A.D. (2019). Immersive, interactive virtual field trips promote science learning. Journal of Geoscience Education, 67(2).

Moore, S., Nguyen, H., Bier, N., Domadia, T., & Stamper, J. (2022). Assessing the Quality of Student-Generated Short Answer Questions Using GPT-3. In Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. (pp. 243-257).

Moore, S., Stamper, J., Bier, N., & Blink, M. J. (2021a). A human-centered approach to data driven iterative course improvement. In *International Conference on Remote Engineering and Virtual Instrumentation* (pp. 742-761). Springer, Cham.

Moore, S., Nguyen, H. A., & Stamper, J. (2021b). Examining the effects of student participation and performance on the quality of learnersourcing multiple-choice questions. In Proceedings of the Eighth ACM Conference on Learning@ Scale. pp. 209-220.

Thille, C. & Smith, J. (2011) Cold Rolled Steel and Knowledge: What Can Higher Education Learn About Productivity?, Change: The Magazine of Higher Learning, 43:2, 21-27

What Works Clearinghouse (2020). "What Works Clearinghouse standards handbook, version 4.1." National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. https://ies. ed. gov/ncee/wwc/handbooks

[7] https://www.cmu.edu/news/stories/archives/2022/june/asu-cmu-chemistry.html

[8] See https://oli.cmu.edu/oli-communities/torus-community/ for schedule and additional information.

Using chatbots as a learning partner to promote student self-awareness and autonomy in an aspirational careers program.

John Kennedy, Deborah Devis, Kim Giannoni, Clare Reuther, Tamika Roach-Attard, Simon N. Leonard
University of South Australia
john.kennedy@unisa.edu.au

ABSTRACT: This presentation discusses the use of a chatbot as a learning activity embedded in an aspirational careers program. While the use of chatbots in higher education and tutoring contexts is commonplace, the use of a chatbot as an active partner in learning to encourage student skill development and self-awareness is novel. We will describe the implementation of a chatbot—*Monty*—aimed at students who are 10 to 14 years-old. The resulting learning analytics reveal patterns around students' abilities to communicate and develop empathy when limited exclusively to verbal communication strategies when solving a social problem by engaging with Monty. Implications for other uses of these types of tools are discussed.

Keywords: Actionable Learning Analytics, Self-Determination Theory, Careers Education, Educational Chatbots, Artificial Intelligence in Education.

1 INTRODUCTION

Careers education in some form is common to many education systems around the world. However, studies have shown that the evidence of the effectiveness and impact of these programs is fragmented and often weak (Hughes et al., 2016). Furthermore, while researchers have advocated for starting careers education with primary-age students for many years (e.g., Watson & McMahon, 2005), Hughes et al. (2016) identified that around 98% of studies assessing the impact of careers education programs focused explicitly on high-school age students (between the ages of 12 and 19). Almost all of the types of intervention identified by Hughes et al. (2016) also focused on student experiences in specific career areas in the form of mentoring, careers provision, or work-related learning or similar. Although the oft quoted *pseudo statistic* that "*N*% of jobs that will exist in 20*XX* haven't been invented yet" (e.g., Krueger, 2021; Tencer, 2017; White, 2017) is patently an exaggeration, there is clearly some wisdom in focusing on the development of transferable skills during careers education programs.

The Australian Blueprint for Careers Development (ABCD) identifies that careers education involves "the knowledge, skills, and attitudes that an individual needs to make sound choices and effectively *manage their own career* [emphasis added]" (National Careers Institute, 2022, p. 5). This concept of developing autonomy or taking control may be understood through the lens of Self-Determination Theory (Deci & Ryan, 2012). Increased autonomy and self-awareness of competence is posited to lead to improved intrinsic career motivation, enhanced workplace outcomes and career satisfaction. Inspection of ABCD Table 2 (National Careers Institute, 2022, pp. 12-14) reveals that effective dialogical communication skills underpin many of the ABCD career management competencies. Hence, any impactful careers education program has a need to progress students' skills in this area.

2 CAREERS COMPASS

Careers Compass (CC) is an *Aspirational* careers program developed by the University of South Australia that is targeted at students in Years 6, 8 and 10 (10 to 16 years old). The CC10 program centers around 30-minute duration, 1-on-1 conversations with specially trained mentors regarding personal career trajectories and intentions. However, this is not readily scalable to other year groups in a school environment, nor would it likely be effective given the age of the participants and their "distance" from the decision point (cf. Hughes et al., 2016). Rather, CC6 and CC8 are one day programs designed around a series of problem-based activities that are undertaken by students, sometimes in groups and at other times individually. These engaging activities assist students in developing transferable skills, such as effective collaboration, awareness of decision making, and successful communication. In order to complete the immersion experience and reduce the school-like *feel* of the program, the separate activities are linked by a single common thematic problem (Figure 1).

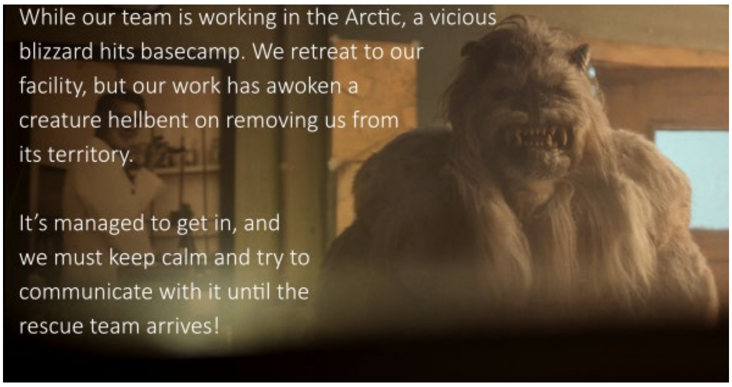


Figure 1: The problem statement and creature underpinning Careers Compass 6 and 8

2.1 Using a Chatbot as a Learning Partner

Chatbots are computer programs that combine Natural Language Processing (NLP) with a degree of artificial intelligence or machine learning to create the appearance of communication. Since the development of ELIZA (Weizenbaum, 1966), chatbot development has generally striven for increased humanlike interactions and enhanced flexibility with regards to dialogue structure and content. This has led to the widespread implementation of chatbots in society and in educational contexts (Pérez et al., 2020). However, within the school context, the implementation of these chatbots has tended to be confined to roles as educational agents—and thus reducing the workload of human educators—or as 24/7 tutors supporting the roles of teaching assistants.

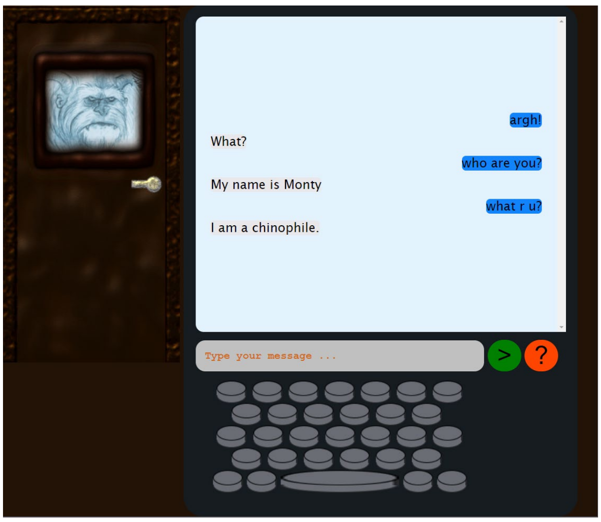


Figure 2: Screenshot of Monty's Interface

For one of the CC program activities, students are required to attempt to communicate with the blizzard creature, known as Monty. One of the human characters has engineered a translation device to communicate with Monty via text. However, as an Artic-dwelling, non-human—and very hungry—interlocutor, Monty has limited understanding of human language and can quickly get grumpy and threaten to eat the players. The students' aim is to communicate clearly and efficiently with Monty to keep him calm and to find out what he wants. A secondary aim is to discover more about Monty.

Monty has been developed as a hybrid scripted NLP chatbot using a combination of HTML and JavaScript for the user interface and PHP, Python and MariaDB for the backend. Students interact with Monty through a text-based interface (Figure 2) accessible through a web browser. The data used to create the initial scripts was sourced from roleplayed conversations. When a user's input does not reach a threshold similarity, the input is saved in an "unknown inputs log" and is used to refine the conversation scripts. The NLP semantic similarity process is implemented using sentence-transformers with a *distilled RoBERTa* model (Reimers et al., 2022). This model is limited to 128 words per sentence and transforms text to a 768-dimensional dense vector. A lightweight model such as this is ideal for this use case as it is fast and scalable even when running only on CPU. The data flow and logic model is shown in **Appendix A Online**.

3 UNDERSTANDING STUDENT CONVERSATIONS

Monty is currently under active development and the system has supported 110 student conversations to date. Conversations with Monty are logged by the tool and used to create graphs of conversations (**Appendix B Online** shows *Student A's* conversation with Monty). These graphs can be analyzed to generate a number of metrics that can describe a student's level of communication skill development in the context of the task. These metrics are designed to yield values in the range 0 to 1 with values closer to 1 theorized to indicate more well-developed skills.

- **Task Focus**: The number of dialogue categories found in the conversation log before the victory condition is met indicates the level of task orientation for the student. A focused conversation is less likely to wander between dialogue categories.
  $TF = \frac{1}{1+n_{categories}}$. Student A demonstrates a *TF* = 0.083.

- **Communication Predictability**: This is the fraction of inputs that generated known answers by the system. A more predictable conversation indicates that the participant is able to elicit meaningful responses from their partner more often.
  $CP = \frac{n_{known}}{n_{known}+n_{unkown}}$. Student A demonstrates a *CP* = 0.429.

- **Communication Clarity**: This is the mean of the cosine similarities returned by the semantic similarity process. Student A demonstrates a *CC* = 0.864.

4 IMPLICATIONS FOR PRACTICE AND RESEARCH

While the system described in this presentation uses a fundamentally simple approach, it is easy to see how a tool such as Monty can yield rich data that can inform learning. As part of the CC program, the communication metrics are fed into other parts of the reporting system and assist students in managing their own skills development in line with the principle of autonomy in self-determination theory. Once a significant number of conversations have been logged, it is intended to report students' skill development in comparison to all users and to age peers, alongside tailored feedback that students can act on to develop these skills further. This approach to providing actionable learning analytics directly to students will allow them to form an objective understanding of competence in this skill area and will empower students with agency to manage their own future learning.

Furthermore, student feedback about this activity has been very positive. Students find this approach to learning both fun and engaging as evidenced by these comments:

> "My favorite part of today was talking to the monster. It was so dumb but funny because he said he would eat kids" (Year 6 Student)

Students also demonstrated empathy in their feedback:

> "I said CALM DOWN and he didn't calm down. But then when my sister tells me to calm down, I get really angry so I stopped." (Year 6 Student)

It is clear from this research activity that gamified chatbots such as Monty, that operate as a learning partner rather than an expert tutor, have significant potential for use in the school classroom. The addition of multiple scenarios and win conditions, a teacher facing dashboard, and longitudinal tracking have applications for whole-school literacy development as well as general skills development. As the research in this area continues, we expect that the addition of other metrics to the system and the refinement of student feedback systems will enhance this system even further.

5 ACKNOWLEDGMENTS

The development of this project was supported by funding from the Commonwealth of Australia, Grant No. NCIPII000291. The authors would like to thank and acknowledge all the students and teachers who have contributed to the development of the project and the team at Anomaly.com.au for their expertise in video production.

REFERENCES

Deci, E. L., & Ryan, R. M. (2012). Handbook of Theories of Social Psychology: Volume 1. In. SAGE Publications Ltd. https://doi.org/10.4135/9781446249215

Hughes, D., Mann, A., Barnes, S.-A., Baldauf, B., & McKeown, R. (2016). Careers education: international literature review. http://www.educationandemployers.org/wp-content/uploads/2016/07/Careers-review.pdf

Krueger, N. (2021, 31 Aug 2021). Preparing Students for Jobs That Don't Exist. ISTE Blog. https://www.iste.org/explore/iste-blog/preparing-students-jobs-dont-exist

National Careers Institute. (2022). Australian Blueprint for Career Development. Canberra: Commonwealth of Australia Retrieved from https://www.yourcareer.gov.au/sites/default/files/documents/Australian%20Blueprint%20for%20Career%20Development.pdf

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. Computer Applications in Engineering Education, 28(6), 1549-1565. https://doi.org/https://doi.org/10.1002/cae.22326

Reimers, N., Becquin, G., & Espejel, O. (2022). all-distilroberta-v1. In Hugging Face. https://huggingface.co/sentence-transformers/all-distilroberta-v1

Tencer, D. (2017, 17 Jul 2017). 85% Of Jobs That Will Exist In 2030 Haven't Been Invented Yet: Dell. The Huffington Post. https://www.huffpost.com/archive/ca/entry/85-of-jobs-that-will-exist-in-2030-haven-t-been-invented-yet-d_ca_5cd4e7dae4b07bc72973112c

Watson, M., & McMahon, M. (2005). Children's career development: A research review from a learning perspective. Journal of vocational behavior, 67(2), 119-132. https://doi.org/10.1016/j.jvb.2004.08.011

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45.

White, D. (2017, 11 Sep 2017). We can't prepare for jobs that don't exist yet. Teach for Australia: Personal perspectives. https://teachforaustralia.org/we-cant-prepare-for-jobs-that-dont-exist-yet/

Designing a Student-Facing Learning Analytics Dashboard to Support Online STEM Practices

Fei Gao, Victoria Lower, Mohammed Abouheaf, Resmi Krishnankuttyrema, MD Sarder
Bowling Green State University
gaof@bgsu.edu, vlower@bgsu.edu, mabouhe@bgsu.edu, resmik@bgsu.edu, msarder@bgsu.edu

ABSTRACT: Despite the potential of learning analytics dashboards (LADs) to support learners' needs for autonomy, little research has been conducted on designing LADs to support student autonomy. In this paper, we reported the process of designing a student-facing LAD that offers students' autonomy support by providing necessary information for students to set their own goals and choose learning activities that are aligned with their goals. A leaderboard was also integrated into the LAD to promote student motivation. Reeves's (2006) design-based research model was adopted to develop the LAD. The final version of the LAD was presented, and the significance of the work was discussed.

Keywords: learning analytics dashboard, STEM, online learning, autonomy

1 INTRODUCTION

A learning analytics dashboard (LAD) or learning dashboard is "a single display that aggregate different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualization" (Schwendimann et al., 2017, p. 37). Research on student-facing LADs is particularly promising due to its potential to support students' autonomy and help learners become more self-regulated (Bodily & Verbert, 2017; Yoo, Lee, Jo, & Park, 2015). However, researchers suggested that there is a lack of actionability in current student-facing LAD design (Verbert et al., 2020). To address this concern, this study focuses on designing a LAD that prompts student to autonomously set their goals and choose appropriate learning activities.

2 LITERATURE REVIEW

2.1 Research on STEM Fluency

This study was conducted within an established online training program for STEM learning called STEM Fluency. This online program, built upon research-validated principles and methods including computer-based training with feedback, mastery-based training, distributed and interleaved practice, and multiple representations (Mikula & Heckler, 2017), is to improve undergraduate student mastery of essential STEM skills (i.e., basic procedural skills) via explicit practice. Research on STEM Fluency training has showed some positive results in enhancing both the accuracy and fluency of student essential skills (Heckler & Mikula, 2016; Mikula & Heckler, 2013; Mikula & Heckler, 2017). However, when students were given the freedom to choose what skills to practice in STEM Fluency after they completed the assigned tasks, most students chose to practice the skills that they were good at instead of those they needed to improve. To address this problem, this paper reports the attempt to design and develop a student-facing LAD that provides students process-oriented feedback and prompts that help students make choices that are more beneficial to their learning.

2.2 LADs and Student Autonomy

Research on LADs is still at its early stage (Schwendimann et al., 2017), and little research has been conducted on LADs and student autonomy support. Although LADs were designed help learners better monitor their learning activities, limited attention was given to support student autonomous decision-making during the learning phases such as planning and control (Valle, Antonenko, Dawson, & Huggins‐Manley, 2021). Our research directly addresses the research gap by designing a student-facing LAD that offers students' autonomy support and guides students to make well-informed learning choices on their own.

2.3 Research on Leaderboards

One concern about the implementation of student-facing LADs is the low usage of LADs from students (Bodily, Ikahihifo, Mackley, & Graham, 2018). To address this potential issue, we decided to incorporate a leaderboard, a popular gamification feature, into the LAD design. Research on the integration of gamification features into LADs has been limited. According to Sahin and Ifenthaler (2021), only 8 out of 76 studies on LAD include gamification features. In our study, a leaderboard was chosen because research on educational use of leaderboards has shown that, in general, leaderboards have a positive impact on participants' learning and motivation (Kalogiannakis, Papadakis, & Zourmpakis, 2021). It is expected that integrating a leaderboard into LADs may provide additional motivation needed for learners to use the information on LADs to improve their performance.

3 METHOD

The design and development of LAD follows the guidelines of design-based research (DBR) (Wang & Hannafin, 2005). We employed Reeves's (2006) DBR model which comprises four phases: analysis, solution, testing and refinement, and reflection. These phases overlap and proceed in a cyclic manner, with activity in previous phases often influencing activity in later phases. After several rounds of reiteration, we arrived at the version presented in Figure 1.

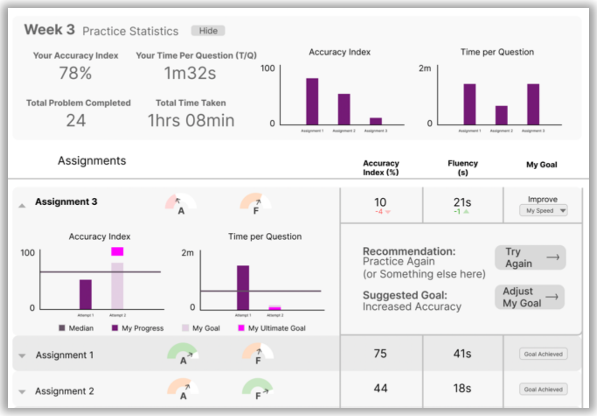


Figure 1. Final version of the LAD design

4 SIGNIFICANCE OF THE STUDY

This project is significant because it is one of the first few attempts to explore ways to design LADs to support learners' autonomy. The significance of the study also lies in the following aspects. First, the design of the LAD addresses the lack of actionability in current LADs design (Verbert et al., 2020) by incorporating actionable items into the LAD. Second, a gamification element, leaderboard, is integrated into the design of the LAD to create a strong support for learners' motivation. Finally, based on studies of the LAD, additional research will be conducted to examine learners' metacognitive strategies and decision-making, which are not well studied in LADs research (Sahin & Ifenthaler, 2021).

REFERENCES

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education, 30*(3), 572-598. doi:10.1007/s12528-018-9186-0

Heckler, A. F., & Mikula, B. D. (2016). Factors affecting learning of vector math from computer-based practice: Feedback complexity and prior knowledge. *Physical Review Physics Education Research, 12*(1), 010134. doi:10.1103/PhysRevPhysEducRes.12.010134

Kalogiannakis, M., Papadakis, S., & Zourmpakis, A.-I. (2021). Gamification in Science Education. A Systematic Review of the Literature. *Education sciences, 11*(22), 22. doi:10.3390/educsci11010022

Mikula, B. D., & Heckler, A. F. (2013). The effectiveness of brief, spaced practice on student difficulties with basic and essential engineering skills. In *2013 IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1059-1065). Piscataway, NJ: IEEE.

Mikula, B. D., & Heckler, A. F. (2017). Framework and implementation for improving physics essential skills via computer-based practice: Vector math. *Physical Review Physics Education Research, 13*(1), 010122. doi:10.1103/PhysRevPhysEducRes.13.010122

Reeves, T. C. (2006). Design research from a technology perspective. In J. v. d. Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 52-66). London: Routledge.

Sahin, M., & Ifenthaler, D. (2021). Visualizations and dashboards for learning analytics: A systematic literature review. In *Visualizations and dashboards for learning analytics* (pp. 3-22). Cham: Springer International Publishing.

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. doi:10.1109/TLT.2016.2599522

Valle, N., Antonenko, P., Dawson, K., & Huggins‐Manley, A. C. (2021). Staying on target: A systematic literature review on learner‐facing learning analytics dashboards. *British Journal of Educational Technology, 52*(4), 1724-1748. doi:10.1111/bjet.13089

Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., De laet, T., & Assoc Comp, M. (2020, 2020). *Learning analytics dashboards: The past, the present and the future.* Paper presented at the Proceedings of the Tenth International Conference On Learning Analytics & Knowledge, NEW YORK.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5-23. doi:10.1007/BF02504682

ACKNOWLEDGEMENTS:

This work was supported by the National Science Foundation under Award Number: 2142608.

The Middle-Man Between Models and Mentors: Using SHAP Values to Explain Dropout Prediction Models in Higher Education

Juan Andrés Talamás Carvajal
Tecnológico de Monterrey
juan.talamas@tec.mx

ABSTRACT: One of the challenges of prediction or classification models in education is that the best performing models usually come in a "black box", meaning that it is almost impossible for non-data scientists (and sometimes even experienced researchers) to understand the rationale behind a model prediction. In this poster we show how SHAP (SHapley Additive exPlanations) values can be used for model explainability as a baseline, and how this same tool might be used for further variable analysis and possibly even bias detection by obtaining SHAP values and figures for two dropout prediction models trained with student data from two different educational models implemented in the same University.

Keywords: CCS Concepts: • Applied computing → Education; Learning management systems. Additional Key Words and Phrases: Dropout, XAI, AI fairness

1 INTRODUCTION

This poster aims to demonstrate the usefulness of SHAP (SHapley Additive exPlanations) (Lundberg et al, 2018) as a tool for stakeholders which can help better understand a machine learning model and as a way to visualize possible bias in education. Our case study is based around the identification of students at risk of dropping out using a machine learning approach, with the variables used being general demographic information, extracurricular activities, and previous school level grades and information. We show these results for two distinct educational models from Tecnológico de Monterrey. The educational models are notoriously distinct from one another, something we will take advantage of to demonstrate SHAP as a general explainability tool.

2 MATERIALS AND METHODS

2.1 SHAP and Shapley values

SHAP is an explainability tool based on a game theory approach to fair distribution called Shapley values, where several players (model features) interact together to obtain a payout (prediction). The Shapley values refer to the marginal contribution of each player to the difference between the expected value (average) and the real value.

SHAP has the following properties that allow them to serve as reliable explanations to a particular model: Local accuracy: the explainer approaches the true model for a specific input as other values are removed; Missingness: a missing value or 0 has no effect on model impact; Consistency: if a model changes to a point where an input's contribution increases or stays the same, that input's explainer value should not decrease. (Lundberg et al, 2018)

It is important to note that SHAP can be applied both to linear and non-linear models. While the "Additive" part of the name might point towards linearity, this refers only to the process to arrive at individual predictions, and not to a necessity of a linear model.

2.2 Dataset

The dataset used in this paper was obtained from Tecnológico de Monterrey (Alvarado-Uribe, 2022). It features anonymized information of undergraduate students who have enrolled and attended at least one semester from 2014 to 2020. The data presented in this paper is available upon request in the Institute for the Future of Education's Educational Innovation collection of the Tecnológico de Monterrey's Data Hub at https://doi.org/10.57687/FK2/PWJRSJ (Alvarado-Uribe, 2022). The dataset contains data from Tec20 ("Classic" educational model) and Tec21 (Competence based educational model). The models differ greatly in their structure and processes, which inspired their comparison.

3 RESULTS

Ranking variable importance can be helpful as a first step, but it is with Swarm plots that SHAP values start to show their worth. Figure 1 shows both educational models side by side. These plots show a vertical ranking (y-axis) and a mapping of how each variable and its specific value impact the model output. An individual point's color indicates a high (red), low (blue), or purple-ish (intermediate) value on the variable its showing, while its position (x-axis) shows how that value impacted model output. Each point constitutes a single student's score in each feature.

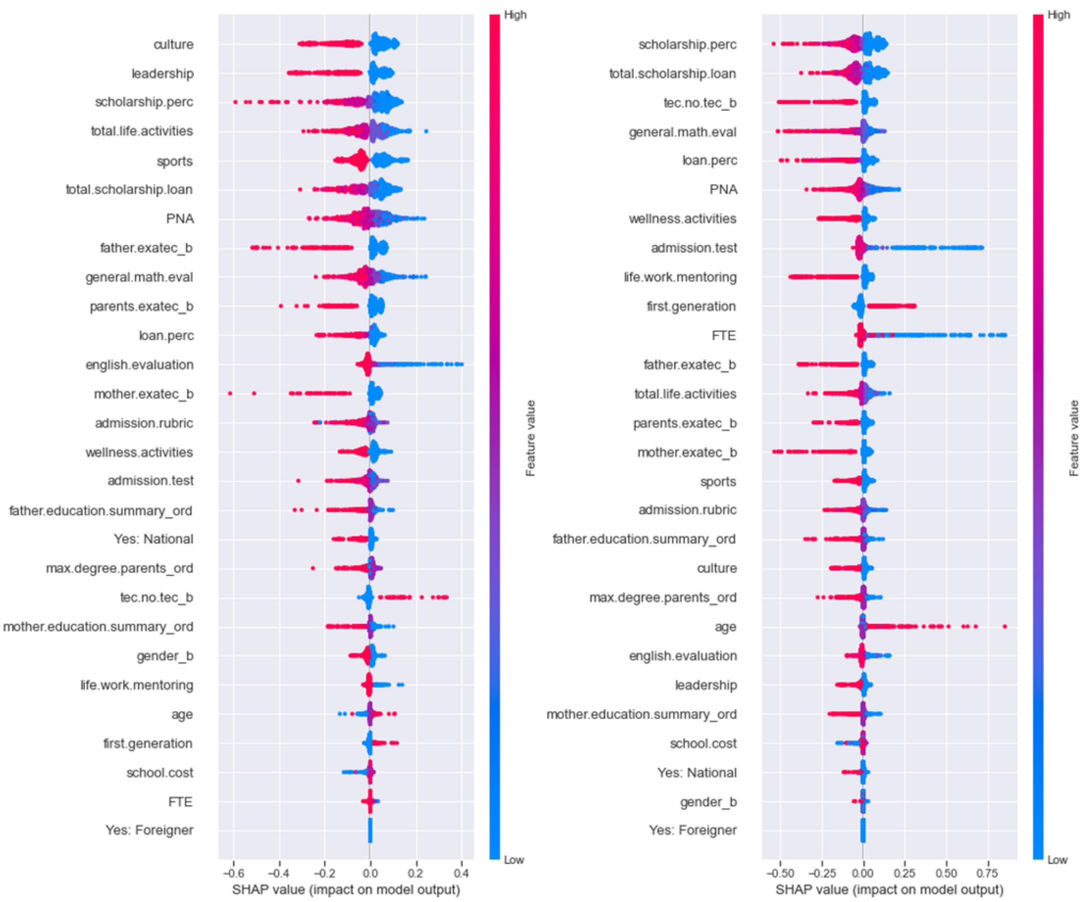


Figure 1: Side by side comparison of Swarm plots for Tec20 (left) and Tec21 (right) educational models

A quick example: A low value for "english.evaluation" on the Tec20 model (left) is indicated in blue and shows a positive SHAP value, pushing the model output towards our target variable (Dropout=True). That is, having a low "english.evaluation" increases the risk of student dropout. Swarm plots allow for quick discovery of variable effects that is both intuitive and informative. We could conclude from figure 1 that, in average, participation in leadership and culture activities (high values in "culture" and "leadership" variables) lead to student retention predictions in the Tec20 educational model, while low scores in the admission test and older students tend to dropout predictions in the Tec21 educational model.

4 DISCUSSION AND CONCLUSION

Using SHAP values to identify the most important overall variables, along with the general effect of those variables' values makes for an extremely powerful tool for average educational practitioners. Speaking to a student tutor or mentor about model precision and recall won't increase their trust on machine learning, but showing them a swarm plot like the ones above allows them to use their expertise to more easily understand model decisions and rationale. However, we believe that SHAP values can go even further.

Classic feature importance can be easily obtained from other tools, but the information available from SHAP values shown in the swarm and waterfall plots allows for any reasonably competent user to make their own analysis without the need of data science training, making the tool both transparent in its decisions, and giving stakeholders the necessary information to make data driven decisions. One possible use would be to identify variables that could introduce bias (gender is a classic example) and verify their overall effect on the model. If we find bias towards or against a specific group, it could be an indicator of a problem with the data collection, or even on a more systematic level.

Future research will focus on expanding the explainability and usefulness of the models, starting with the development of counterfactuals (how much a variable score needs to change to flip the prediction) to provide a viable path towards "breaking the prophecy" of the predictions of our machine learning models. In other words, finding what a student at risk of dropout needs and is able to change to reduce that risk.

REFERENCES

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K. W., Newman, S. F., Kim, J., & Lee, S. I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, *2*(10), 749–760. https://doi.org/10.1038/s41551-018-0304-0

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

Monotonic Variational AutoEncoder-based Individually Optimized Problem Recommender System

Takashi Hattori, Hiroshi Sawada, Sanae Fujita, Tessei Kobayashi, Koji Kamei, and Futoshi Naya
NTT Communication Science Laboratories
(takashi.hattori.cp|hiroshi.sawada.wn|sanae.fujita.zc|tessei.kobayashi.ga|
koji.kamei.zx|futoshi.naya.ke)@hco.ntt.co.jp

ABSTRACT: We propose a novel problem recommender system that can suggest moderately challenging problems to learners. By training a Variational AutoEncoder to reconstruct problem-answer data with a small number of latent variables, we can predict the likelihood of a learner's ability to correctly solve unanswered problems. Experimental results show that the system's predictions are accurate, and that it can recommend moderately challenging problems tailored to individual learners.

Keywords: Problem recommender system, correctness rate prediction of unsolved problems, Variational AutoEncoder

1 INTRODUCTION

People can achieve efficient learning by selecting appropriate material from many available options. In this paper, we propose a problem recommender system that chooses tasks tailored to individual learners. Our system combines ideas from multidimensional item response theory (Reckase, 2009) with a Variational AutoEncoder (Kingma et al., 2014) to make personalized recommendations.

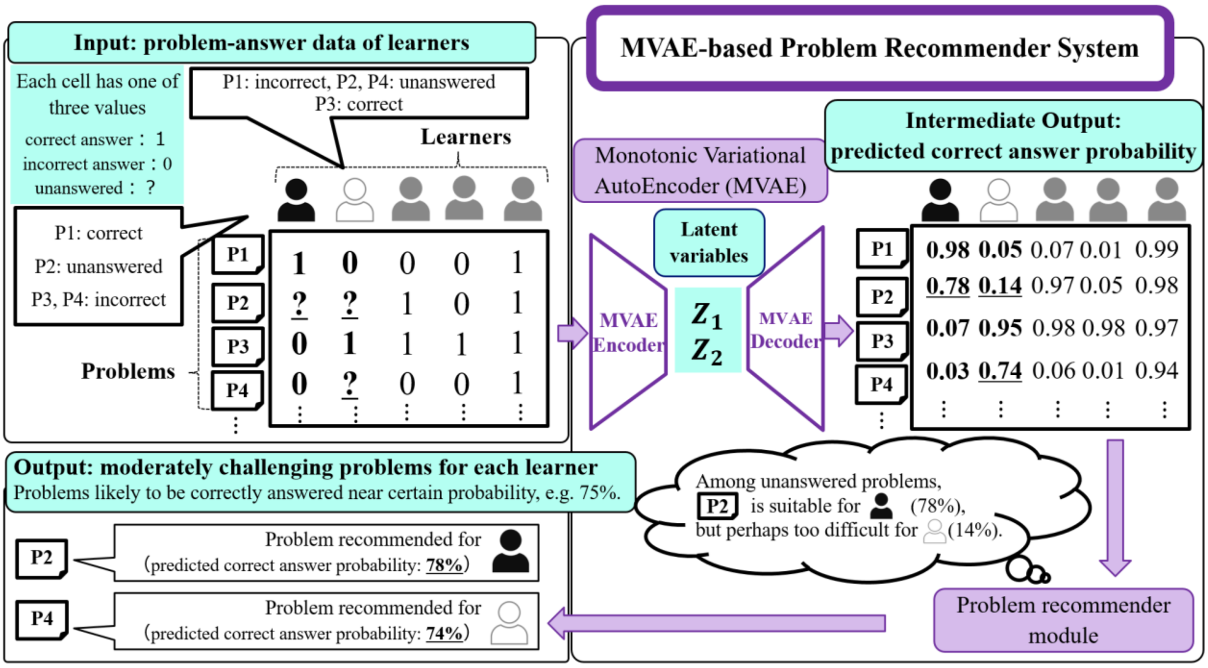


Figure 1: Overview of proposed problem recommender system

2 PROBLEM RECOMMENDER SYSTEM

Figure 1 shows an overview of our recommender system, which consists of a Monotonic Variational AutoEncoder (MVAE) and a problem recommender module. The MVAE predicts the probability that a learner can correctly solve unanswered problems. Based on the predicted correctness probability, the problem recommender module suggests moderately challenging, unanswered problems.

The system input is the problem-answer data of the learners. Each learner's response to a given problem has one of the following three values: *correct answer*, *incorrect answer*, or *unanswered*. The MVAE encoder compresses each learner's data into a set of latent variables with a smaller dimension. The MVAE decoder outputs the correctness probabilities based on these latent variables. The encoder and decoder are trained so that the output data match the answered portion of the input data as closely as possible. The MVAE decoder predicts the correctness probability of the unanswered problems using the latent variables obtained from training.

Learners can enhance their ability by reviewing the problems that they didn't answer correctly. However, answering overly difficult problems is exhausting and deflates learning. Thus, selecting moderately challenging problems for each learner is probably more effective for learning.

In addition to their usefulness for predicting correctness and recommending problems, MVAE latent variables offer another useful feature. By constraining the weights of the decoder's neural networks to be non-negative, the latent variables are trained to be monotonic with respect to the predicted, correct answer probabilities. This means that learners with larger latent variables are more likely to answer questions correctly. Furthermore, latent variables are trained to be independent from one another, and we expect them to represent different problem-solving capabilities. By analyzing the latent variables of each learner, the system can identify a learner's strengths and weaknesses.

3 EXPERIMENT, RESULTS AND DISCUSSIONS

We conducted a two-step experiment to evaluate the problems recommended by our system.

Eighteen Japanese adults in their 20s to their 50s participated in the experiment. In the data collection step, they were instructed to answer as many English language problems as possible in their spare time. We prepared three types of English problems: sort problems, filling-in problems, and choose-the-best-caption-for-a-image problems. These problems were accessible online from participants' devices. The average percentage of correct answers was 85%, and several participants in interviews after data collection stated that overall they felt the questions were easy. Half of the chosen problems were identical for each participant; the other half were selected separately for each participant. The shared answers are necessary for training the MVAE, and the separate questions provide unanswered (or new) questions for recommendations. The MVAE trained by the data collected in this step predicted the probability of the participants to correctly answer the unanswered problems in the second step.

In the second step, the participants answered English problems in the same manner as the first step. However, the problems given to them were selected differently. At a probability of 50%, the problems were either randomly chosen or recommended by the MVAE. In this experiment, we categorized problems with a predicted correctness rate close to 75% (10% lower than the average correctness rate in the data collection step) as moderately challenging and hence suitable for recommendations. More specifically, we recommended unanswered problems with a predicted correctness rate in a range of 70-80%. Since problems were posed without indication of the selection method, the participants could not distinguish between random and recommended problems.

Figure 2 compares the cumulative answer correctness rate over time of the random and recommended problems in the experiment's second step. The error bar indicates the standard errors.

The answer correctness rate of the random problems remained at approximately 85% throughout the experiment, just as in the first step. The results show that no significant change occurred in the learner ability or the difficulty of the problems during the experiment.

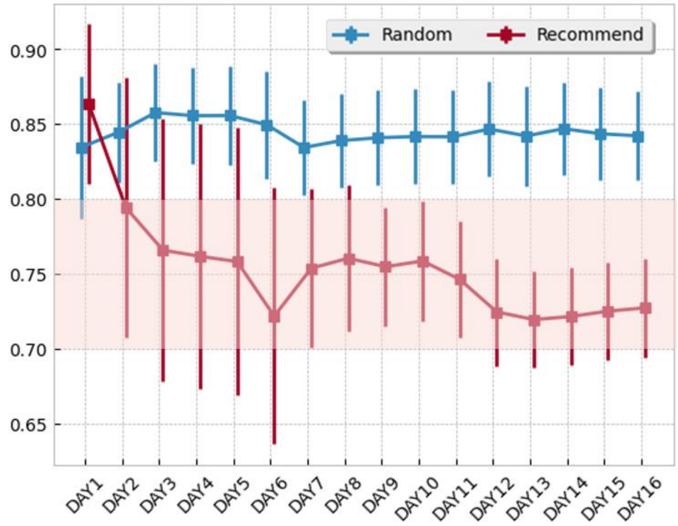


Figure 2: Comparison of cumulative answer correctness rate of random problems and recommended problems

On the other hand, the answer correctness rate of the recommended problems is lower than that of the random problems, hovering between 70% and 80%, which is the target range predicted by the MVAE. Although the standard error of the recommended problems is large at first, it gradually decreases and settles at a similar value to that of the random problems.

4 CONCLUSION AND FUTURE PERSPECTIVES

We proposed a Variational AutoEncoder-based problem recommender system and experimentally demonstrated that recommended problems were answered within the probability range predicted by the system. Future work will take advantage of learner characteristics analyzed by latent variables, which are difficult to discover using conventional collaborative filtering and clustering methods.

AKNOWLEDGEMENTS

This work was supported by Council for Science, Technology and Innovation, "Cross-ministerial Strategic Innovation Promotion Program (SIP), Big-data and AI-enabled Cyberspace Technologies". (Funding agency: NEDO (Japan))

REFERENCES

Reckase, M. D. (2009). Multidimensional item response theory models. In *Multidimensional item response theory* (pp. 79-112). Springer, New York, NY.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Building a Data-to-Dashboard Pipeline Using IPEDS Datasets through Google Cloud Platform for Automized Reporting System

Damji Heo Stratton
Solutions & Analytics Unit, Missouri Online, the University of Missouri System
dhsdfn@umsystem.edu
David Reid
Solutions & Analytics Unit, Missouri Online, the University of Missouri System
reidda@umsystem.edu
Lee Larrick
Solutions & Analytics Unit, Missouri Online, the University of Missouri System
larrickl@umsystem.edu
Scott Gossett
Solutions & Analytics Unit, Missouri Online, the University of Missouri System
smgrh6@umsystem.edu

ABSTRACT: This poster describes the ongoing project of building a data-to-dashboard pipeline by Missouri Online to better support the data-informed policy decision making process using data from the Integrated Postsecondary Education Data System (IPEDS) and Google Cloud Platform (GCP) environment. The diagram that illustrates the process to create the dashboard is included in this poster, along with the description of the dashboard. The challenge in implementation and notes for practice are also described. This case could offer insights to stakeholders who are interested in building such a system to better support the data-informed policy decision making process.

Keywords: Data visualization, cloud computing, data-informed policy decision making, learning analytics, the Integrated Postsecondary Education Data System (IPEDS)

1 BACKGROUND AND CONTEXT

Missouri Online has been supporting data-informed policy decisions regarding online programs within the University of Missouri System by providing stakeholders yearly conferral data for each online program and labor market data for the graduates of the programs. However, data entry and document formation were done manually, resulting in inefficiencies in producing market research reports. In addition, the reports depended on a proprietary database, which offered limited information on the characteristics of the competing institutions. Thus, Missouri Online adapted a more efficient data-to-dashboard pipeline process by using IPEDS datasets to have more options to compare data regarding competing institutions and conferral trends. IPEDS is a public database system managed by the National Center for Education Statistics, where the data are collected at the aggregated level from postsecondary institutions and do not have student-level information (Integrated Postsecondary Education Data System, n.d.). In our case, we used IPEDS data to understand the national trends of the conferrals from each online program and find the characteristics of competing institutions. Missouri Online also leveraged Google Cloud Platform (GCP) since it has multiple features to efficiently build the pipeline. For instance, we save the raw datasets into BigQuery and efficiently conduct Extract, Transforms, and Load (ETL) tasks using Jupyter Lab in the Vertex AI Workbench within the environment. Then, we store the curated datasets into designated storage buckets in Google Cloud. Finally, we connect the datasets to Tableau desktop to create dashboards. In addition, GCP enables enrichment of the datasets with the internal data from our internal database management system (DMS) more conveniently. The current project is in progress, which is part of the effort to design a more efficient data architecture for Missouri Online and to make insights from the data warehouse more accessible to a broader range of stakeholders.

The objective of this project is to build a data-to-dashboard pipeline for automating the reporting system to the stakeholders who make data-informed policy decisions regarding online programs in the University of Missouri System. The desired product is a dashboard that demonstrates the detailed characteristics of the competing institutions and trends in the conferral data using IPEDS datasets from 2016-2021.

2 DEVELOPMENT AND IMPLEMENTATION

2.1 The process

Figure 1 illustrates the detailed process to build the data-to-dashboard pipeline. The raw datasets from IPEDS are extracted and saved into a staging table within BigQuery. Then, data wrangling is conducted on Vertex AI to produce the requested final dataset. Lastly, the final curated dataset is exported to a designated bucket which is connected to Tableau. We employed the participatory design method to design and develop the dashboard (Verbert, Ochoa, De Croon, Dourado, & Laet, 2020). Thus, there are multiple prototype versions as the prototypes go through internal reviews within the unit and external reviews from the stakeholders, such as program coordinators who send reports to the college deans, department chairs, program chairs, and the president, followed by revising the prototypes and test-run the dashboards. The duration of the current project implementation was approximately 8 months.

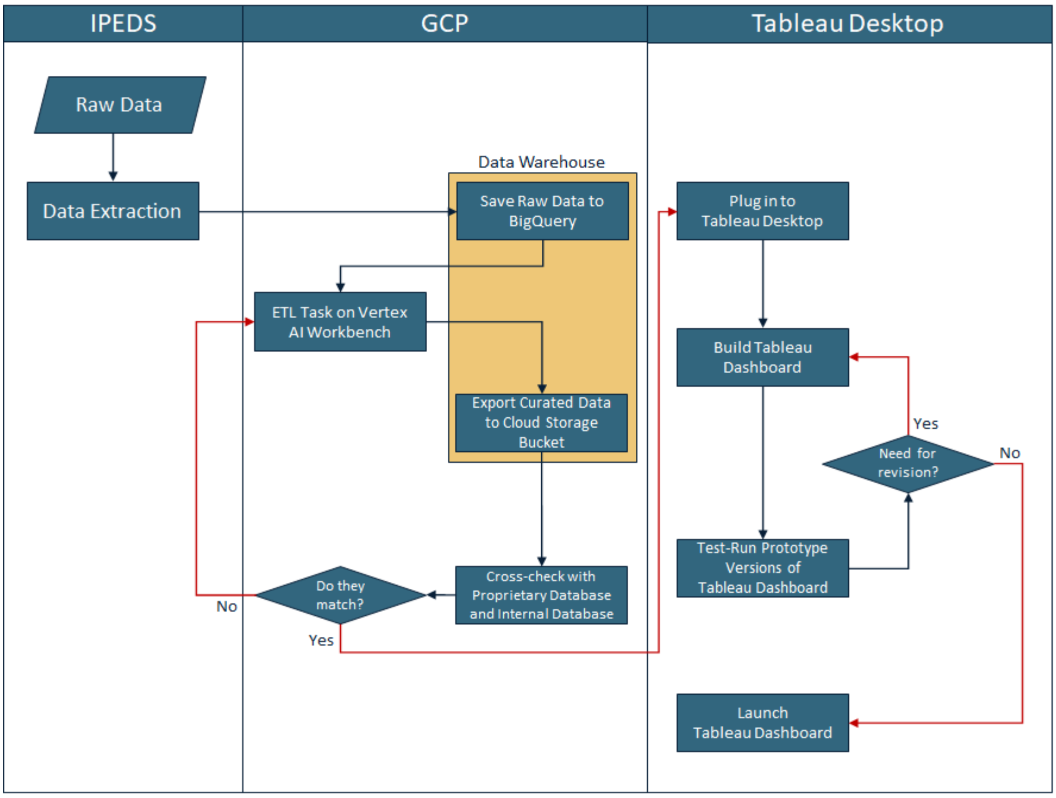


Figure 1: Data-to-dashboard pipeline diagram

2.2 Final product

The final product provides the stakeholder with self-service access to the conferral data in multiple unique views and filters for refinement. The final version of the dashboard consists of seven pages. The first page covers a brief overview of the entire dashboard. The second page includes the overview of the Carnegie classification 2021 update page. The third page is the map that shows geographic information of competing institutions as well as the campuses of the University of Missouri System using multiple filters where you filter the institutions based on CIP code, degree level, regional, non-regional, school type (private vs. public), the size of the institution, and Carnegie classification 2021 update categories. The fourth page includes multiple data visualization figures that show the trends of conferral data. The fifth page includes the box-n-whisker plots that represent the distribution of the conferral data from institutions based on each Carnegie classification 2021 update category. The sixth page includes the executive summary of the data visualization of the previous pages. The last page shows the comprehensive text tables that include detailed information on institutions and conferral data from 2016 to 2021. Figure 2 below demonstrates an example of the executive summary page of the dashboard.

3 FINDINGS OF EVALUATIONS

Dashboard prototypes were distributed to internal unit members for feedback leading to changes prior to showing the dashboards to stakeholders. Dashboards were then shown to stakeholders eliciting further feedback to ensure better usability. Features, filters, and designs were updated based on the stakeholder feedback. Multiple opportunities were given for evaluation of the project prototypes, ultimately providing a series of dashboards and visualizations the stakeholders could use for data-informed policy decision making.

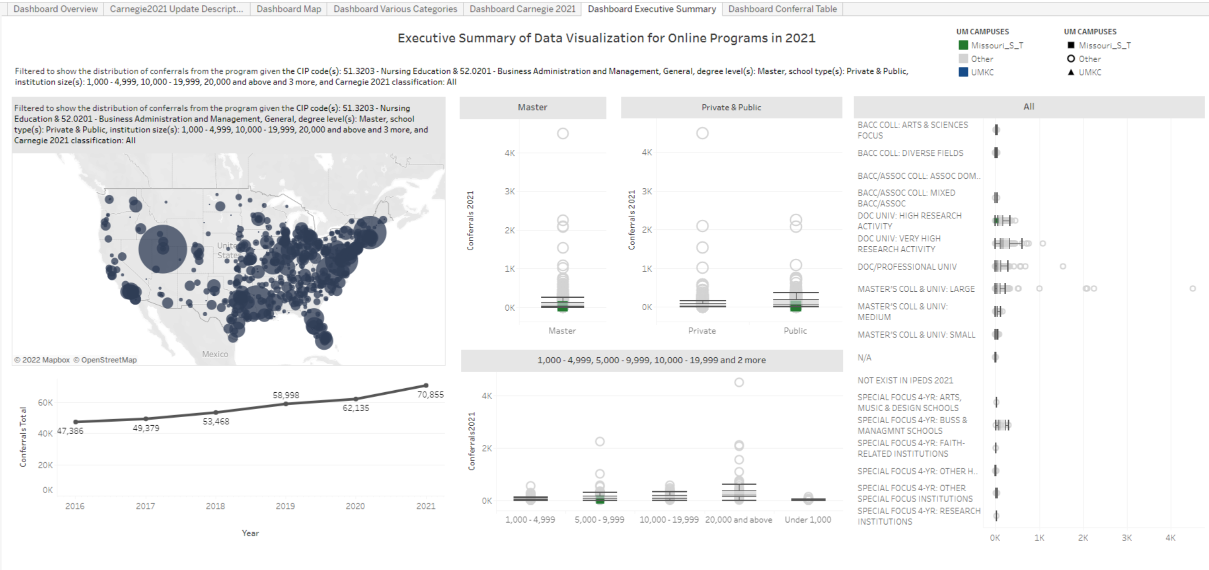


Figure 2: Executive summary page of dashboard

3.1 Challenges in implementation

To export the final dataset file to a designated bucket, a service account key was needed to access the bucket. The Jupyterlab notebook was shared, allowing any user who had the service account key to access the project. We found that we needed to re-organize and refine the project and roles of the users so that only those who were closely involved in the project could view the Jupyterlab notebook and the service account key file to the bucket. Currently, the process is being improved so that we can limit access to the notebook as well as the service account key file.

3.2 Notes for practice

To optimize the Google Cloud operating costs, it is recommended to test and refine the script in the local Jupyterlab environment before running it on Vertex AI. It is also worth considering converting the CSV file of the raw datasets to parquet formats as parquet files take much less disk space than CSV files. Another note is that because the end users are not necessarily used to the terminologies used to understand IPEDS dataset, it is helpful to add descriptions to help them to better understand the information from the dashboard.

4 CONCLUSION AND FUTURE WORK

The current project describes our in-progress work on building a data-to-dashboard pipeline using IPEDS datasets and GCP for automizing the report system to better support the data-informed policy decision making process. The final dashboard version includes the conferral data from programs that can be completed via distance education either completely or partially from 2016 to 2021 and detailed characteristics of institutions. The next step of the project includes matching the conferral data from each program from the internal DMS to the current IPEDS conferral data so that the dashboard can include the program title information that the four campuses of the University of Missouri System are currently using. In addition, imputed conferral data can be included as well as the conferral data from IPEDS datasets to treat missing data and potential issue of inaccurately entered data. Lastly, labor market data will be integrated into the dashboard.

REFERENCES

Integrated Postsecondary Education Data System (n.d.) IPEDS survey components. https://nces.ed.gov/ipeds/use-the-data/survey-components

Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020, March). Learning analytics dashboards: The past, the present and the future. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 35-40).

The Adoption of Learning Analytics in Higher Education: An Exploratory Study in the Dominican Republic

David Soto[1*], Shizuka Shirai[1], Mayumi Ueda[2], Manabu Higashida[1], Haruo Takemura[1]
[1]Osaka University, [2]University of Marketing and Distribution Sciences
[*]david.soto@lab.ime.cmc.osaka-u.ac.jp

ABSTRACT: The field of learning analytics (LA) is growing as a valuable solution for higher education institutions to understand students' learning processes and improve the quality of education. However, even with proven benefits, there has been a slow adoption of learning analytics, particularly in Latin American (LATAM) countries when compared to other regions. This paper presents an exploratory study conducted in the Dominican Republic regarding the adoption status, goals, and barriers to the success of learning analytics in higher education.

Keywords: learning analytics, higher education, adoption.

1 INTRODUCTION

The adoption of learning analytics (LA) in higher education institutions (HEIs) in LATAM countries is relatively low compared to other regions (Hilliger et al., 2020), where most surveys and studies around LA adoption are from universities in Europe, the United States, and Asia. One of the first initiatives addressing this gap was the foundation of the LALA project (Muñoz-Merino et al., 2020). This collaboration between European and LATAM universities has the objective of providing guidelines and best practices for the design, implementation, and adoption of LA tools to improve the quality of education in HEIs in LATAM. Using the Supporting Higher Education to Integrate Learning Analytics (SHEILA) model as a reference and adapting it to the LATAM educational context, the LALA framework proposes an integrated approach to embrace LA under four dimensions.

However, particularly in the Dominican Republic, to our knowledge the only study on the adoption and use of LA in the country was conducted by Chaljub et al. (2019) with the major goal of reviewing learning metrics from a theoretical and practical perspective in preparation for their integration into the Dominican Republic's educational system. In this way, this research further expands to investigate the factors impacting the adoption of LA in the Dominican Republic to better understand the strategies, outcomes, and barriers faced by HEIs in the country.

2 METHODOLOGY

The method for this exploratory study was an online survey. Looking for an integrated approach for the analysis of the adoption and challenges of LA, we adapted the model proposed by Tsai et al. (2020) in their exploratory study of HEIs in Europe. The questionnaire was adapted as is with only minor changes for translations to Spanish. The survey starts with demographic information, followed by a core question about the current institutional level of adoption of LA. From there, the survey branches through five different sections: (1) current development and adoption status of LA; (2) strategical plan for the implementation of LA; (3) resources for strategy development; (4) strategic, legal, and ethical considerations; and (5) self-evaluation of the maturity and institutional readiness. The survey was distributed to 46 institutions recognized by the Ministry of Higher Education (MESCyT) in the Dominican Republic for a period of 2 months.

3 RESULTS

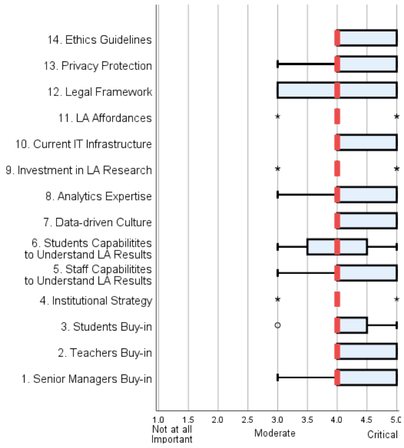
As for *demographics*, from the 46 formally recognized HEIs in the Dominican Republic, 15 responded to the survey, resulting in a 32% response rate. Lack of awareness about LA, conducting the survey exclusively online, and making direct contact to HEIs without a formal intermediary institution could be potential reasons for the low response rate. Institutional affiliation was balanced with responses from 7 public and 8 private institutions. Although the number of institutions is not particularly high, three of them have a strong influence on the general education of the country, enrolling nearly 50% of the country's entire student population. In terms of the *current adoption* of LA, only 5 institutions have implemented LA, where 1 is at a large-scale, 4 are at a small-scale, and most institutions have less than 5 years of experience in the field. The remaining 10 institutions have not implemented LA yet, and only 4 are in preparation to do so. The complete questionnaire (https://shorturl.at/AHMQW) and tables (https://shorturl.at/svDZ0) references are available online. In this paper, we focus on LA outcomes, key elements to success, and critical barriers currently faced by HEIs to adopt LA. Results are contrasted with previous research from European HEIs (Tsai et al., 2020).

First, we report the motivations, methods, strategies, and outcomes of LA. The first three items were collected from all institutions except for one that selected "Not Considering" to adopt LA. In resemblance to European institutions, the *main motivation to adopt LA* for Dominican institutions was to improve student learning performance, favored by 85.71% of the institutions. Also, the improvement of student satisfaction and teaching excellence were highly regarded. For the *means and usage of LA* to achieve motivational goals, institutions in the Dominican Republic focus on basic techniques such as measuring learning and teaching performance and trying to produce reports from collected data. This demonstrates the lead on experience of Europe when compared to LATAM countries, where previous research observed that European HEIs go beyond learning assessment to trying to understand how students learn or interact with learning resources in order to provide interventions that are appropriate for the learners.

In regards of *strategies and resources to adopt LA*, most institutions either developed or adapted a particular strategy for LA. Resources for the development of strategies included LA policies at the institutional level, teaching staff, and technological vendors. *Strategic, legal, and ethical considerations* insights reveal that ensuring that data is collected, stored, and shared anonymously was the main agreement for students' privacy protection. Finally, *outcomes* from the 5 institutions that have implemented LA show agreement on that LA provides effective solutions to their needs, they have redesigned curricula based on results, and personalized support from teachers to students has increased. In contrast, these outcomes were more inclined to a neutral position in European HEIs.

Additionally, regarding LA leaders, senior managers were the major stakeholders, followed by IT and teaching staff, and students were last. Whereas in Europe the learning and teaching department was first, followed by IT staff. In other aspects, ethics guidelines, IT infrastructure, a data-driven culture, and teachers' buy-in were considered the main *key elements for success* towards LA (Figure 1). This is homogenous with European HEIs where most elements were rated as very important as well.

Finally, LA affordances, IT infrastructure, investment in analytics, and senior managers buy-in were the most *critical barriers* to adopt LA (Figure 2), while in Europe results were more polarized and barriers were in terms of analytics expertise, senior managers buy-in, and legal frameworks. In this way, proper investment, structured strategies for LA, and the creation of a strong community of experts to spread awareness and best practices for an efficient adoption as proposed by Chaljub et al. (2019) could serve as a breakthrough for these barriers in the Dominican Republic.



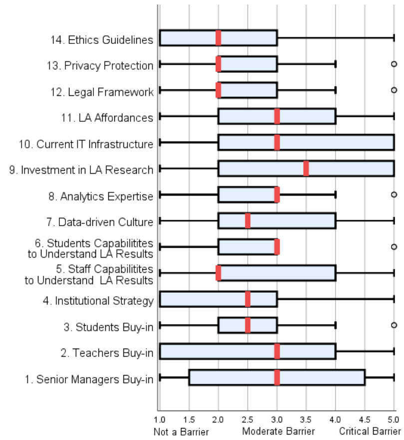


Figure 1: LA Key Elements for Success (N=15)

Figure 2: LA Barriers for Success (N=15)

4 CONCLUSION

In accordance with previous research, the adoption rate of LA in the Dominican Republic is still not high. This exploratory study provides an initial baseline for a learning analytics perspective in the country and upcoming case studies in LATAM. Future research should try to integrate personalized interviews about LA strategies and adoption frameworks, thus motivating HEIs to have a clearer view of their goals and how to approach them efficiently. Finally, involving in the research process important national educational institutions such as the Ministry of Higher Education (MESCyT) and the Dominican Association of Rectors of Universities (ADRU) could significantly increase the awareness of LA in the country and improve the response rates and results of subsequent case studies.

REFERENCES

Chaljub, J., Torres-Toukoumidis, A., Flores, A. & Bello, R. (2019). Analíticas de aprendizaje en la territorialidad virtual de la enseñanza: Caso República Dominicana. Hamut´ay, 6(3), 98-113. http://dx.doi.org/10.21503/

Hilliger, I., Ortiz ‐ Rojas, M., Pesántez ‐ Cabrera, P., Scheihing, E., Tsai, Y. S., Muñoz ‐ Merino, P. J., ... & Pérez ‐ Sanagustín, M. (2020). Towards learning analytics adoption: A mixed methods study of data ‐ related practices and policies in Latin American universities. British Journal of Educational Technology, 51(4), 915-937.

Muñoz-Merino, P. J., Kloos, C. D., Tsai, Y. S., Gasevic, D., Verbert, K., Pérez-Sanagustín, M., ... & Scheihing, E. (2020). An Overview of the LALA project. In Workshop on Adoption, Adaptation and Pilots of Learning Analytics in Under-Represented Regions 2020 (pp. 1-5). CEUR-WS.

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

The human-centered design of relevant, user-friendly and inclusive feedback dashboards for standardized tests in primary and secondary school education

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

Dr. Laura Herrewijn
AP University College Antwerp, Belgium
laura.herrewijn@ap.be

Dr. Gert Vanthournout
AP University College Antwerp, Belgium
gert.vanthournout@ap.be

ABSTRACT: From 2023-2024 onwards, standardized tests for Dutch and mathematics will be introduced in primary and secondary school education in Flanders (Belgium). The main purpose of these tests is to foster school development and improve educational quality. The authors' aim is to design high-quality, relevant, user-friendly and inclusive feedback dashboards to disclose the results of these tests to the different user groups (i.e. school principals, teachers, pupils and their parents), paying attention to the dashboards' content and user interface. In order to address this goal, a mixed-methods approach was set up in line with the Educational Design Research and Service Design frameworks. During the first project year, theoretical and practical insights, personas, design guidelines and a design framework were synthesized through desk research, expert interviews, and focus groups with secondary school principals and teachers. During the second project year, the design cycle continued to arrive at a first design prototype of the dashboard for secondary school principals. In this poster, the results of the research and design activities so far will be presented, and attention will be paid to how they informed the design of the content and the user interface of the prototype.

Keywords: Educational design research, School feedback system, Primary and secondary education, Dashboard design

1 EXTENDED SUMMARY

From 2023-2024 onwards, standardized tests for Dutch and mathematics will be introduced in primary and secondary school education in Flanders (Belgium). The main purpose of these tests is to foster school development and improve educational quality. To implement the tests, a Support Centre was created that is being led by an inter-university consortium. The Support Centre will design the tests, coordinate the implementation, administration and analysis of the results, and provide digital feedback to the school principals, teachers, pupils and parents. As part of the Support Centre, the authors of this poster are responsible for the design of the feedback dashboards that will disclose the test results to the user groups.

Hattie and Timperley (2007) define feedback as externally provided information about a person's performance or understanding. Feedback can be a powerful tool for learning, if used properly (Hattie, 2008). The quality improvement in primary and secondary school education that is envisaged with the introduction of the standardized tests will only work if schools deliberately and systematically work with the feedback from those tests (Van Gasse, 2021). In order to achieve this, the way in which schools receive the feedback (cf. the school feedback system) must meet a number of conditions (based on Van Gasse et al., 2015): (1) users should perceive the content of the feedback system as relevant; (2) the feedback system should provide its users not only with information but also with starting points for deliberate use of feedback; (3) the way in which the feedback is presented should facilitate accurate interpretation; and (4) the feedback system is clear and easy to use. Therefore, the author's aim is to develop high-quality, relevant, user-friendly and inclusive feedback dashboards for the different user groups, paying attention to the feedback dashboard's content (e.g. test results to be presented, data visualizations) and user interface (e.g. lay-out, structure, navigation).

In order to address this goal, a mixed-methods approach was set up in line with Educational Design Research ('EDR'; Phillips & Dolle, 2006) and Service Design (SD; Miller, 2015). During the first project year, the authors went through a first cycle of needs analyses and design activities. The needs analysis phase consisted of a desk research including a literature review and analyses of national and international good practices, as well as expert interviews (*N*=10) and focus groups with the user groups of secondary school principals and teachers (*N*=19). During the design phase, the insights of the needs analysis phase were accumulated and translated into a series of design guidelines, a stakeholder analysis and personas for secondary school principals and teachers, which in turn resulted in a design framework. Subsequently, these products were used to develop the first conceptual designs of the feedback dashboards' user interface (e.g. lay-out, structure, navigation) and content (e.g. data visualizations).

During the second project year, the research and design cycle continued with a consultation round of all researchers from the Support Centre, the pedagogical counselling services, the Flemish education inspectorate and the Flemish Department of Education. During this consultation round, the authors showcased the conceptual designs of the user interface and content of the feedback dashboards. The questions that arose during the first design cycle were also discussed. With all this input, the authors built a first interactive design prototype of the feedback dashboard for secondary school principals that included minimal functionalities (i.e. a minimal viable product) in the interface design tool 'Figma' (http://www.figma.com).

The research results show that there are large differences in digital and data literacy between and within our user groups. Therefore, the feedback dashboards need to be usable and understandable by users of all literacy levels. In the current prototype, this was designed for in a few different ways. First, a tutorial was added that explains users how to use and navigate through the dashboard. Second, two types of navigation were built into the prototype: users can search for information by means of the guiding questions (that assess what they are interested in and send them to the correct dashboard page and data visualization) and/or they can freely explore the dashboard by means of the menu. Third, the dashboard was designed to be layered: displaying the most essential and basic information first. Users with more (complex) information needs can subsequently get more details by adding filters, comparisons, confidence intervals,… directly to the data visualizations, or by clicking through to more in-depth analyses. Fourth, a reading guide is added alongside every data visualization, to support users in reading and interpreting all of the information. Finally, an export button gives user more freedom in using the dashboards' information, allowing them to print or download the data in different formats (e.g. a PDF report, an Excel file).

With this poster, the authors would like to present and discuss the theoretical and practical insights that they gathered through research during the first two project years, as well as the design products (e.g. design guidelines, design framework, conceptual designs) and the design prototypes that resulted from this. They will furthermore focus on a couple of challenges that they encountered throughout the design process, such as: how to develop feedback dashboards that are understandable for users with varying levels of digital and data literacy? And how to find a balance between the correct visualization of statistical information and user-friendliness?

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.

Miller, M.E. (2015, 14 december). How many service designers does it take to define service design. Practical Service Design. Geraadpleegd op 10 januari 2022, van https://blog.practicalservicedesign.com/how-many-service-designers-does-it-take-to-define-service-design-6f87af060ce9

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.

Stickdorn, M., Hormess, M., Lawrence, A., & Schneider, J. (2018). This is service design doing. Applying service design thinking in the real world: A practitioners' handbook. Sebastopol: O'Reilly Media

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

Van Gasse, R. (2021, 23 augustus). Aan de slag met schoolfeedback uit centrale toetsen. Op weg doorheen terminologie en aanpak. Edubron Blogt. Geraadpleegd op 10 januari 2022, van http://www.edubronblogt.be/onderzoek/aan-de-slag-met-schoolfeedback/

Evaluating Automatic Speech Recognition for Non-Native Speakers

Leonard Bongard
Department of Computer Science Technical University of Darmstadt Germany
Leonard.bongard@stud.tu-darmstadt.de

Anna Filighera
Multimedia Communications Lab Technical University of Darmstadt Germany
ORCID: 0000-0001-5519-9959

Tim Steuer
Multimedia Communications Lab Technical University of Darmstadt Germany
ORCID: 0000-0002-3141-712X

Thomas Tregel
Multimedia Communications Lab Technical University of Darmstadt Germany
ORCID: 0000-0003-0715-3889

ABSTRACT: It is claimed that sufficient speaking data exists to create powerful Automatic Speech Recognition (ASR) systems. Despite the success of ASR systems for native speakers, the performance of non-native speakers is yet to be evaluated. Therefore, we assembled an English as a Second Language speaker (ESL) corpus with German natives as speakers to build an optimized ASR system using state-of-the-art deep learning approaches. We found evidence that ASR remains unsolved for ESL speakers.

Keywords: Automatic Speech Recognition, Deep Learning, Speech Corpora Creation

1 INTRODUCTION

Automatic Speech Recognition (ASR) techniques are becoming more popular as the transcription performance of ASR systems approaches reliable results, making them suitable to use for a variety of tasks in Computer-Assisted Language Learning (CALL). However, this holds only for speakers with English as First Language (EFL), as the performance of state-of-the-art ASR techniques drops significantly when speakers with English as Second Language (ESL) try to use them. The performance diminishes as ESL speakers diverge from EFL speakers in terms of accent, pronunciation errors, and grammatical mistakes (Wang, 2020). Because of this loss in performance, CALL systems are limited in potential functionality and have to work around this constraint. However, tasks like giving Vocalization feedback for language learners, cannot rely upon such constraints as word-by-word transcriptions are needed in order to work properly (Filighera, 2022).

Therefore, this work proposes a selection of data-driven experiments to tackle the loss in performance. With a small set of manually created free-speech data and a subset of the Common Voice corpus, we evaluate if more training-data would enhance the performance for ESL speakers.

2 DATASET CONSTRUCTION

State-of-the-art ASR systems are trained on huge amounts of EFL data, like the Librispeech corpus which contains transcripts of telephone calls (Baevski, 2020). However, only a minimal quantity of ESL corpora exists, with the crowdsourced Mozilla Common Voice (CV) project (Ardila, 2020) being an exception. The CV corpus is a quality-assured collection of voluntary contributions, where every participant can read out given sentences. The project is maintained to this day and the English subset consists of over 2300 hours of speaking data. The corpus provides useful meta-information per entry about each participant, such as accent and locale. While this information is non-obligatory, we used this information to create a subset of the CV corpus containing only recordings of participants that specified either “german” as the accent or “deutsch” as the locale. We manually searched for German-specific keywords to create this filter. The final subset contains around 73 hours of ESL-speaking data. We split 80% of the dataset off for fine-tuning (Train), 10% for validation (Dev), and another 10% for the final evaluation (Test).

As we couldn’t ensure that the CV subset represents non-native speakers well enough, we decided to create a second corpus with randomly chosen ESL speakers from Germany. 15 ESL speakers answered open-end questions (e.g. “What is your favorite place to go?”) for approximately 15 minutes each. The number of questions answered varies between seven and thirty for each speaker and should encourage them to speak freely. The speech recordings were first transcribed automatically and thereafter corrected manually.

3 EXPERIMENTS

The Wav2Vec2-Large model is currently one of the best-performing Deep Learning approaches for ASR. Another benefit, as the authors of the model claim, is that even small amounts of training data, as little as 10 minutes, can result in a working ASR approach regardless of language (Baevski, 2020). Thus, we decided to fine-tune the given “Wav2vec2-large-960h"[1] checkpoint with our CV subset to create our own model called “Wav2vec2-large-960h-CV". To optimize the performance, we did random search for hyperparameter tuning on the parameter learning rate and weight decay as suggested by the authors of the model. For evaluation, we chose the Word Error Rate (WER) metric (lower is better). To avoid overfitting, we implemented early stopping.

Out of eight runs, the best-performing model achieved a WER score of 0.076 on the CV Subset validation Set and a WER score of 0.081 on the test set. However, the evaluation with our second corpus revealed that the performance, contrary to the expectation, worsened. The “Wav2vec2-large-960h" checkpoint has a WER score of 0.249. Our fine-tuned model reaches only a WER score of 0.314, meaning that roughly every third word is transcribed incorrectly. Even though we couldn’t ensure the correctness of our CV corpus, we assume that the lack of performance is due to the differences in speaking between spoken-out sentences from the CV subset and free speech from our corpus as this is the major difference between both corpora. Free or spontaneous speech contains hesitations, disfluencies, and grammatical errors, all of which are not included in the CV corpus.

[1] https://huggingface.co/facebook/wav2vec2-large-960h

Furthermore, we fine-tuned the models on our free speech corpus. Given different amounts of training data, we wanted to observe if the fine-tuned model starts adjusting for free speech. Therefore, we created 5 different splits of our free speech corpus, with an increasing number of speakers in the Train and Dev set per split. The Test set only contains unseen speakers. We again used random search for hyperparameter-tuning on the same parameter with 5 runs per split. Table 2 summarizes the average results on the Dev and Test set.

Table 2: Fine-tuned Wav2vec2-large-960h and Wav2vec2-large-960h-CV models to evaluate if an increasing number speakers in the Train set leads to an improved WER for unseen speakers.

		Wav2vec2-large-960h		Wav2vec2-large-960h-CV	
N of Speakers in Train and Dev set	N of Speakers in Test Set	WER Dev Set	WER Test Set	WER Dev Set	WER Test Set
10	5	0,176	0,172	0,165	0,172
11	4	0,189	0,179	0,158	0,171
12	3	0,171	0,171	0,167	0,169
13	2	0,171	0,172	0,163	0,176
14	1	0,177	0,172	0,158	0,163

Our fine-tuned CV model performs slightly better on the Dev set, however, equally in comparison to the fine-tuned Wav2vec2 checkpoint on the Test set. These results indicate that fine-tuning with the CV subset improves the overall performance for ESL ASR, despite the verbal differences in speaking. Whether the CV model adjusted itself to the free speech with an increasing amount of training data couldn't be examined yet as the amount of data is not enough both for training and for evaluation.

4 CONCLUSION

An excellent ASR performance is essential for many tasks. Although this requirement is attainable for native speakers, it remains a limitation for non-native speakers. We optimized the performance of our non-native ASR wav2vec2 model, however, our results give evidence that current ASR approaches keep struggling with ESL speakers. Without more high-quality data, we assume that the problem for ESL speakers will remain and therefore, this data limitation should be tackled by future work.

REFERENCES

Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., ... & Weber, G. (2020, May). Common Voice: A Massively-Multilingual Speech Corpus. In Proceedings of the 12th Language Resources and Evaluation Conference (pp. 4218-4222).

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems, 33, 12449-12460.

Filighera, A., Bongard, L., Steuer, T., & Tregel, T. (2022, July). Towards A Vocalization Feedback Pipeline for Language Learners. In 2022 International Conference on Advanced Learning Technologies (ICALT) (pp. 248-252). IEEE.

Wang, Y., Luan, H., Yuan, J., Wang, B., & Lin, H. (2020, October). LAIX Corpus of Chinese Learner English: Towards a Benchmark for L2 English ASR. In INTERSPEECH (pp. 414-418).

How do professional learners engage with learning analytics dashboards?

Sarah Alcock, Maria Aristeidou, Bart Rienties
Institute of Educational Technology, The Open University
sarah.alcock@open.ac.uk, maria.aristeidou@open.ac.uk, bart.rienties@open.ac.uk

Soraya Kouadri Mostéfaoui
Computing & Communication, The Open University
soraya.kouadri@open.ac.uk

ABSTRACT: Learning Analytics Dashboards (LAD) research is mainly carried out in schools and higher education, whilst professional learning has often been overlooked. To address this gap, this fine-grained longitudinal study of 12 professionals over a period of five months has taken exploratory steps into the context of professional accountancy learning. It investigates perceptions and use of a static assessment LAD which incorporates data visualization and personalized written feedback, aiming to purposefully affect learning. Overall, findings demonstrate learners took a positive view of and subsequently actively used the LAD, choosing elements directly related to their needs to inform next revision steps. Learning insights which both provide an understanding of past performance and also specifically recommend how the professional learners might improve performance had the highest frequency of use. The implication for the LAD design community is LAD should include context and learner specific elements, providing personalized next step guidance.

Keywords: Learning Analytics Dashboard, professional learning, assessment, feedback, personalization, accountancy.

1 INTRODUCTION

To date, most Learning Analytics Dashboards (LAD) research and implementation studies have been carried out in schools and higher education institutions (Schwendimann et al., 2017; Sclater et al., 2016). Professional learning – ongoing learning in a professional workplace – has been largely overlooked by the Learning Analytics (LA) research community (Buckingham Shum et al., 2022). This research begins to address the gap by investigating learners' perceptions and use of a LAD in a professional accountancy learning environment. This knowledge can inform future LAD design and implementation across professional learning contexts.

2 CONTEXT

The study context was a commercial tuition provider preparing professional accountants for the final admitting exam of the Institute of Chartered Accountants in England and Wales (ICAEW). This longitudinal research followed 12 UK based professional learners from the beginning of the course to the final exam in an authentic learning context, answering calls for practice-based evidence for effects of a LA tool on learning (Wise et al., 2021). Participants had already taken up to 14 high-stakes ICAEW examinations, so were experienced post-graduate level learners. The assessment LAD visualizes complex competency-based exam marking data to support learners in improving exam performance. LAD elements include standard bar/line charts showing grade by Requirement and by Skill, along with (as per Figure 1) a unique 'map' of the marking key (1) with personalized written feedback (2) and how to improve (3).

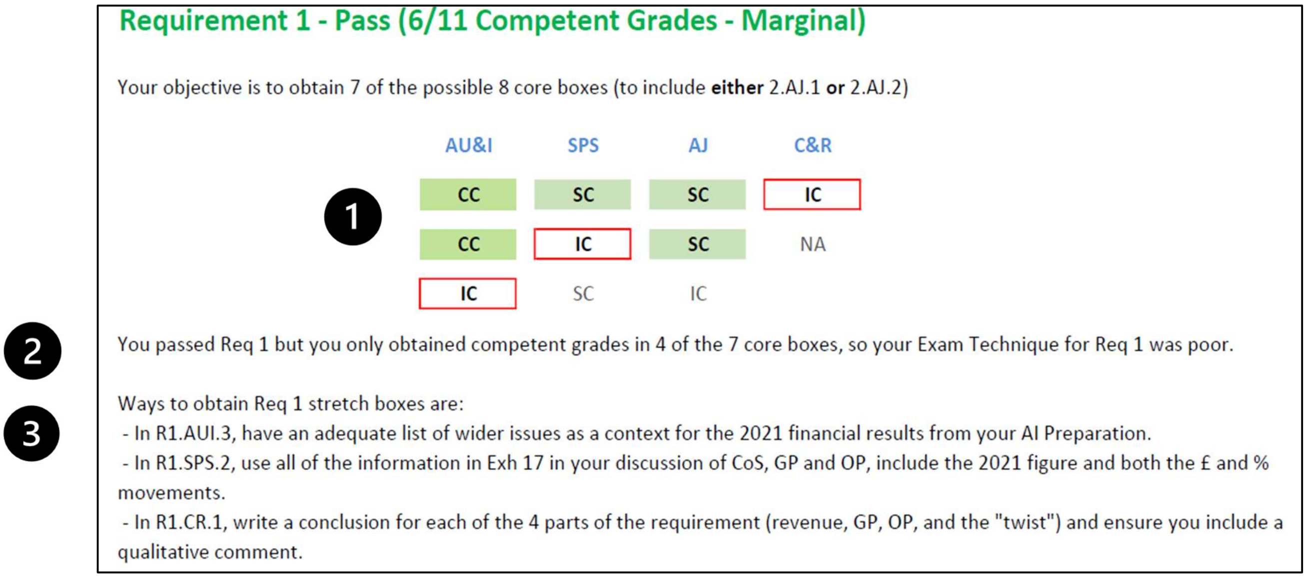


Figure 1: 'map' of the marking key (1), including personalized feedback (2) and how to improve (3)

3 METHODOLOGY

The research took a mixed methods approach, using classroom observations, questionnaires (with 5-point Likert scales and free text options) and post-final exam semi-structured interviews (n=3). Full data was collected for nine learners. Figure 2 shows the data collection timetable. The questionnaires were analyzed and responses checked for normality and skewness. While Factor Analyses were conducted, given the small sample size we opted not to report these. Inductive and deductive processes were used to thematically code each interview transcript (n=3) for references to learning activities as well as context-specific behaviors and pressures.

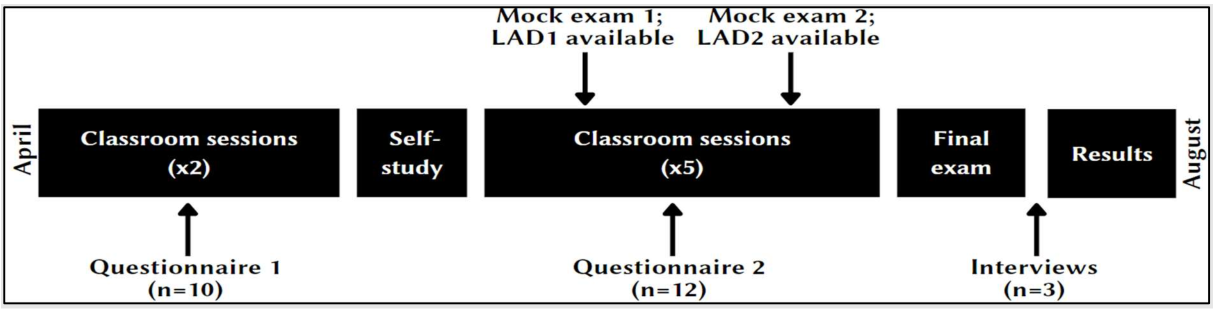


Figure 2: Data collection timetable

4 RESULTS AND DISCUSSION

Participants at the beginning were relatively positive about their intended use of LAD. Using a cut-off at 3.4, 70% of participants indicated to want to use the LAD. At the second measurement, on average, half of the offered LAD elements were used by participants, while five out of 12 respondents used less than a third of the LAD elements. The most used elements were the marking key 'map' and associated text (n=10, 83%), followed by the personalized feedback / how to improve text (n=9, 75%). This indicates professional learners are looking for both specific insights into and a personalized explanation of the learning data available, supporting previous findings in higher education studies (Rets et al., 2021, Sedrakyan et al., 2020). The post-exam interviewees (n=3) all had an overall positive view of the LAD and took a similar approach in using it to “trigger” (participant 9) their home study revision activities. However, the LAD could not solve the learning challenges of every participant. Participant 13, a low-reported user (only one element, the ‘map’), gave insight into their concerns: “I think my motivation and energy levels limit my amount of revision more than the results on the dashboard.” The data from this study, therefore, supports previous findings (Rets et al., 2021) that ‘good’ learners find the LAD useful, but those who are struggling are less likely to use it.

5 CONCLUSION

The LAD studied presented individual exam marking analysis, focused on each learner’s performance compared to assessment standards. It does not aim to make predictions of future performance based on the past but offers an automated, personalized explanation of a mock exam result using published standards for accountability. This study has shown the context specific LAD to be an effective tool for professional learning improvement for those learners that engage with it, with learners showing a preference for context specific data visualizations and associated ‘what next?’ personalized feedback. Although the LAD could not solve all learning difficulties, in this small-scale study it was widely positively received. Further research across multiple classes and cohorts of professional learners is both warranted and planned for 2023.

REFERENCES

Buckingham Shum, S., Littlejohn, A, Kitto, K and Crick, R. (2022). Framing Professional Learning Analytics as Reframing Oneself, *IEEE Transactions on Learning Technologies (Early Access)*, 1–17. Https://doi.org/10.1109/TLT.2022.3190055.

Rets, I., Herodotou, C., Bayer, V., Hlosta, M and Rienties, B. (2021). Exploring critical factors of the perceived usefulness of a learning analytics dashboard for distance university students, *International Journal of Educational Technology in Higher Education, 18*(*1)* https://doi.org/10.1186/s41239-021-00284-9

Schwendimann, B., Rodriguez-Triana, M., Vozniuk, A., Prieto, LP., Boroujeni, M., Holzer, A., Gillet, D. and Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research, *IEEE Transactions on Learning Technologies*, *10(1),* 30–41. https://doi.org/10.1109/TLT.2016.2599522

Sclater, N., Peasgood, A and Mullan, J. (2016). *Learning analytics in higher education: A review of UK and international practice Full report*. Jisc. Retrieved from https://www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v2_0.pdf.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S. and Kirschner, P. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation, *Computers in Human Behavior, 107* https://doi.org/10.1016/j.chb.2018.05.004

Wise, AF., Knight, S and Ochoa, X. (2021). What Makes Learning Analytics Research Matter. *Journal of Learning Analytics, 8(3)*, 1–9. https://doi.org/0.18608/jla.2021.7647

Evaluating the Efficacy of Tutoring Services: A Data Mining, Machine Learning, and Propensity Score Matching Approach

Author(s): Aaron Thomas

ABSTRACT: The following poster is an attempt to evaluate the impact of tutoring services upon student academic success and an exploration of how the effectiveness of tutoring varies by course using data mining, machine learning, and propensity score matching. One of the problems of post-hoc evaluations of student services such as tutoring is that these services are open to the entire population of the student body, which confounds those likely to pass without tutoring versus those likely to be non-successful without some type of intervention. In order to unconfound these populations, course level predictive models were developed for the traditional gateway courses to identify students in need of a tutoring intervention. A second step included using a propensity score matching analysis of those in need and who used the tutoring services in comparison to those who were in need but never engaged with tutoring. Further it was possible to identify students who began tutoring before being detected by the predictive models as compared to those who initiated tutoring after. Results indicated that students who engaged in tutoring before identification were more likely to pass the course than those who initiated tutoring after a predictive signal.

.**Keywords**: Program evaluation, machine learning, predictive analytics, propensity score matching

1 RECENT FINDINGS CONCERNING TUTORING

Tutoring across grade levels has been found to positively and significantly impact academic performance at all levels of education. Alegre et al. (2019) in a meta-analysis of 42 studies found that peer-tutoring at the K-12 had a moderate effect upon student grades (*d*=0.38). Steenbergen-Hu and Cooper (2014) in a meta-analysis of 39 studies on intelligent tutoring systems and human tutors found a moderate effect (*g*=.32-.37) with ITSs significantly less effective than human tutoring. The impact of tutoring upon academic performance is a well-studied and documented educational intervention.

While tutoring is a well-established educational intervention, it is important for institutions to evaluate the efficacy of their programs, especially in light of financial limitations for student support services. The following poster will describe initial efforts to leverage data mining, predictive modeling, and propensity score matching to evaluate tutoring services.

1.1 Population

Three terms (Fall 2020, Spring 2021, Fall 2021) of high-enrollment core-tracking courses for many degree programs were selected because 1) these courses have higher than normal non-success rates 2) a predictive system was employed to predict as early as week 4 of a term success and non-success. It is to be remembered that for the purposes of this analysis the population reported here are those *students predicted at any point after week 4 of the term to be non-successful*. As this part of the evaluation concerns the causal relationship between tutoring effects upon student grades, it was necessary to focus upon the subset of students predicted to be unsuccessful.

2 THE CASE FOR MACHING LEARNING AND PROPENSITY SCORE MATCHING

One of the primary obstacles to evaluating the impact of voluntary services or interventions is the lack of a control group that predominates in experimental studies where variance for key variables among the groups can be equally distributed via random assignment. In the case of tutoring, however, there are significant confounds between the degree of need of tutoring service among students. Tutoring services tend to be used by two types of students, those who are likely to pass the course and those who are in danger of academic non-success. While one student may seek tutoring to improve a B+ to an A-, another may need tutoring to pass the course with the minimum, typically a C grade in a traditional academic setting. As a result of these two types of students, the empirically observable impact of tutoring services varies between these two populations of students varies greatly.

To differentiate between at-risk students and non-risk students, machine learning methods, specifically random forest algorithms, were used to predict whether a student was at-risk for a particular course. The overall accuracy of the model ranged between 90%-98% across core-tracking courses based upon previous validation efforts.

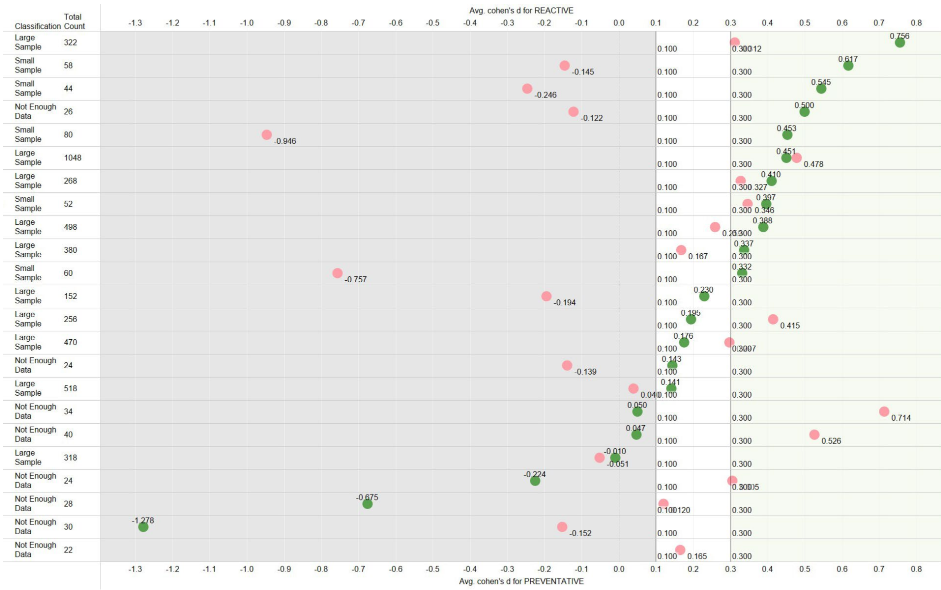
To establish a causal connection between tutoring and academic success, propensity score matching, a statistical technique which matches a treatment case (tutoring) to a control case (no tutoring) based upon a propensity score which is the probability of the non-intervention group being assigned to the quasi-experimental control group was employed. PSM supports the even distribution of covariates between groups (Leite 2016). This method supports the assumption that the groups to be compared are comparable in practical and theoretical terms.

The following potential confounding variables were used in a PSM: schedule pass rate, total cumulative credits, total transfer credits, total test credits, GPA , attempt ,and first time in college vs transfer.

3 RESULTS

The impact of tutoring vs no tutoring upon end of semester course grades (0-100%) can be classified as positive medium effect (d=.24) across all courses and is comparable to effect sizes reported in the literature (d > .30). Viewed from an institutional perspective where a passing grade is C or above, the tutoring group had a significantly higher proportion of success students (52.02%, N=2,376) vs the no tutoring group (42.59%, N=2,376) ($h_{effect\ size}$=.20). The proportion of passing grades for PREVENTATIVE tutoring (61.56%, N=947) was significantly higher than both the NO TUTORING group (42.38%, N=2,376; $Z_{1.96}$=10.21) and the REACTIVE tutoring group (45.70%, N=1,259, $Z_{1.96}$=9.35). There were significant differences between the reactive tutoring group (45.70%, N=1,429) and NO TUTORING control (42.38%, N=2,376; $Z_{1.96}$=2.01). The results support the importance of initiating tutoring as soon as possible although there are still noticeable positive effects for reactive tutoring. See Figure 1 for the difference in effect sizes for preventative vs reactive tutoring.

Figure 1: Preventative vs reactive tutoring. (pink=reactive, green preventative)



4 LIMITATIONS

One of the major limitations of the following evaluation is the reliability of Canvas LMS data for grades. Some courses rarely used the Canvas gradebook or some courses have misconfigured gradebooks that do not store grade data so *these courses are likely to be underrepresented in this evaluation*. Also the quantity and quality of the tutoring session may be an important factor in the effectiveness of tutoring as compared to the non-granular approach taken in this evaluation where a student is classified as a tutoring case if they complete one or more tutoring sessions.

Another issue to consider is that the population considered here are only those students identified as needing tutoring services. There may be positive effects for tutoring students who were not in jeopardy of course non-success. The strength of claims of causality reported here is limited by the nature of a quasi-experimental study.

REFERENCES

Alegre, F., Moliner, L., Maroto, A., and Lorenzo-Valentin, G. (2019b). Peer tutoring and mathematics in secondary education: literature review, effect sizes, moderators, and implications for practice. *Heliyon* 5:e02491. http://10.1016/j.heliyon.2019.e02491

Leite, W. (2016). *Practical propensity score methods using R*. Sage Publications.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of educational psychology*, *106*(2), 331-347. https://doi.org/10.1037/a0034752

Technology and Collective Learning in Improvement Networks

Ouajdi, Manai

The Carnegie Foundation for the Advancement of Teaching manai@carnegiefoundation.org

Susan, Haynes

The Carnegie Foundation for the Advancement of Teaching haynes@carnegiefoundation.org

ABSTRACT: The Networked Improvement Learning and Support (NILS) platform is an online tool designed to accelerate the initiation and development of Networked Improvement Communities in a disciplined manner. Its main goal is to promote social, organizational learning through curation and synthesis and tacit to explicit knowledge conversion to facilitate knowledge construction and ownership by the communities regarding improvement practice in education. In this proposal we will discuss the NILS platform, a use case, and a plan of analytics development that advances knowledge dissemination and monitors the health status of networks.

KEYWORDS: Networked Improvement Community, Improvement Science, Behavior Modeling, Statistical Analysis, Social Network Analysis, Recommender Engine, Machine Learning, Network Health.

1 NETWORKED IMPROVEMENT COMMUNITIES

In the past decade the Carnegie Foundation for the Advancement of Teaching has pioneered a fundamentally new vision for the research and development enterprise in education. In particular, the Carnegie Foundation seeks to join the discipline of improvement science with the powerful capacities of networks to foster innovation and social learning for education reform. This approach is embodied in what we call Networked Improvement Communities (NICs) [1]. NICs are scientific learning communities distinguished by four essential characteristics: (a) focused on a well specified common aim; (b) guided by a deep understanding of the problem, the system that produces it, and a shared theory of practice improvement; (c) disciplined by the rigor of improvement science; and (d) coordinated to accelerate the development, testing, and refinement of interventions, their more rapid diffusion out into the field, and their effective integration into varied educational contexts. Examples of NICs are Carnegie Math Pathways, which aims to improve outcomes of students who placed into remedial math, Building a Teaching Effectiveness Network, which aims to improve the retention of effective new teachers [1], and Mountaineer Mathematics Master Teachers, which aims to increase the opportunity for 20,000 middle and high school students to "do mathematics" in class.

2 NETWORKED IMPROVEMENT LEARNING AND SUPPORT (NILS) PLATFORM

To accelerate improvement work by NICs, we have developed an online platform for social learning: the Networked Improvement Learning and Support (NILS) platform. NILS is designed to promote social, organizational learning and to disseminate tacit and explicit knowledge [2] for improvement in education by moving much of what we currently do face-to-face into a virtual learning environment.

As we support NIC leaders seeking to initiate, grow, and sustain their networks, we recognize the need to ensure that the networks' improvement efforts are grounded analytically, tested empirically, and accelerated through network-wide social learning. This requires front-line practitioners to shift from consumers to producers of the knowledge that advances improvement practice. How can we assist them in building improvement science habits and mindsets? Accordingly, we need a support infrastructure for documenting, capturing, and organizing improvement knowledge for access and use throughout widely distributed networks. The design and creation of such a platform is critical for the Foundation to advance widespread use of the NIC as a social arrangement to accelerate improvement work in the field of education.

The NILS platform is designed to accelerate the initiation and development of NICs in a disciplined manner. The landing page shows the driver diagrams that NIC members are acting on and their in-progress tests of change on their network's formulated theory of practice improvement. The driver diagram provides a common, controlled language that facilitates NIC members' understanding of their measurable common aim, along with a set of hypotheses and actual ideas for change to achieve the aim [1]. It also enables NIC members to track Plan-Do-Study-Act (PDSA) cycles with relevant data. The PDSA cycle is a basic method of inquiry in improvement science and can be used to turn ideas into action and connect action to learning [1]. NILS shows pictures of NIC members as they work in the platform together and collaborate, so as to promote group solidarity and affinity, and cultivate a sense of belonging to the networked community. NILS allows NIC members to participate in and comment on group work, network activities, and PDSA cycles. The platform provides space for dialogue and discussion, and emphasizes the curation and synthesis of network information and learning as a means for NIC members to jointly construct and own a body of knowledge for improvement practice. NILS renders to the affordances of various technological devices and is designed to be integrated into practitioners' current workflow for ease of use (see Figure 1).

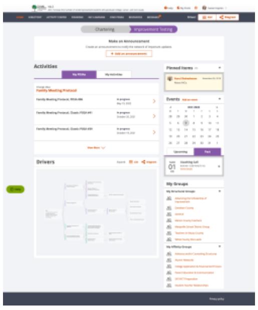


Figure 1

3 PLAN OF ANALYTICS

As users utilize the NILS platform, the system automatically collects user behavioral data as a form of click-stream data. Production data is available through PDSA cycles and discussion boards. Qualitative data on networks' system usage is collected through monthly interviews and social graphs of network members' interactions on NILS are generated for analysis. By leveraging machine learning techniques [3], we plan to analyze these data in order to examine how change ideas evolve over time as a consequence of interactions among NIC members, and more particularly, to identify what types of interactions involving what levels or groups of NIC members contribute most to this evolution. This analysis of interactive patterns serving to evolve change ideas can contribute to informing users of the best practices in knowledge and learning dissemination. We plan to develop a recommendation engine to notify users about groups in the NIC that are currently working on similar issues and suggest that the users join those groups to learn from their practice. This may be particularly useful for those who struggle with implementing change ideas and seemingly need support from the outside. From social network standpoints [4], we will identify what we call informal leaders who are, for example, active in executing change ideas and leverage them for further knowledge dissemination (see Figure 2). At the same time, we are interested in any social, cultural norms that may develop over time in NICs, which may facilitate or inhibit knowledge dissemination. Related to this, we want to examine any shift in NIC members' mindsets and habits regarding improvement science. Our analysis also focuses on the status of health or sustainability of the NIC and the development of a relevant analytics dashboard.

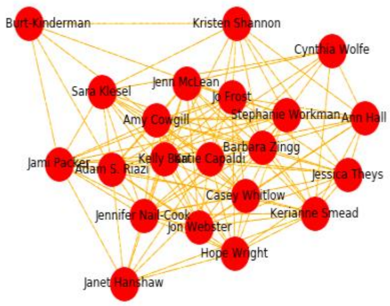


Figure 2

4 USE CASE

Multiple NICs working on the NILS platform have generated thirty use cases that illustrate the system's value in supporting the initiation, development, and maturation of networks. An example of a network using NILS is West Virginia's Mountaineer Mathematics Master Teachers network, which has engaged in a technical partnership with the Carnegie Foundation to enable the identification, support, and leveraging of experienced and exemplary secondary mathematics teachers as master teachers who, though the use of the tools of improvement science and networked improvement communities, lead efforts to improve math teaching and learning across West Virginia, with a focus on the extent to which students have meaningful and robust opportunities to engage with mathematics.

5 CONTRIBUTION TO THE FIELD

We will present our developmental work on the NILS platform. From our presentation, the field will be able to learn patterns of knowledge dissemination and evolution, any shifts in users' mindsets and habits, analytic techniques to identify those patterns, as well as data visualizations informative to NIC members from different levels in the community.

6 BEYOND NILS

The Carnegie Foundation aims to integrate NILS with two other systems to expand the platform's feature set and capabilities and to further support the work of improvement networks: a network health assessment tool and the Community Learning Improvement Platform (CLIP). The network health assessment tool is designed to gauge the health of networks and features a health survey and reporting dashboard. The CLIP system exports network learning from NILS to the wider education improvement community.

REFERENCES

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.

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. New York, NY: Oxford University Press

Murphy, K. P. (2012). Machine learning: A probabilistic perspective. Cambridge, MA: The MIT Press.

Laat, M. De, Lally, V., Lipponen, L., & Simons, R. J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. International Journal of Computer-Supported Collaborative Learning, 2, 87-103

Learning analytics perceptions and expectations in a Mexican university: a qualitative study

Tomás Bautista-Godínez[1]; José Gerardo Moreno-Salinas[1]; Verónica Luna de la Luz[1]; Irma Rocío Zavala Sierra[1], Roberto Santos Solórzano[1], Melchor Sánchez Mendiola[1]
[1]National Autonomous University of Mexico (UNAM)
tomas_bautista@cuaieed.unam.mx; gerardo_moreno@cuaieed.unam.mx; veronica_luna@cuaieed.unam.mx; irma_zavala@cuaieed.unam.mx; roberto_santos@cuaieed.unam.mx; melchor_sanchez@cuaieed.unam.mx

ABSTRACT: The adoption of Learning Analytics (LA) in Higher Education Institutions (HEIs) in Latin America and the Caribbean is still in an exploration stage. A large public Mexican university started the process of integrating LA in its educational spaces, exploring the perceptions and expectations about the management and use of educational data for LA adoption. A qualitative study was carried out using semistructured interviews with high-level institutional authorities, and focus groups with students and teachers, from different disciplinary areas. Results indicate that perceptions and expectations are oriented toward improving school performance through data-based feedback, with ethical responsibility. The more specific constructs of LA still need to be disseminated and internalized in Mexican universities' educational stakeholders, in order to increase the likelihood of its successful adoption.

Keywords: Learning analytics adoption, Learning analytics perceptions, Qualitative study, Latin America, Higher education.

1 INTRODUCTION AND CONTEXT

In the last five years, a movement of exploration, research, and implementation of Learning Analytics (LA) has been launched with intensity in some universities in Latin American and Caribbean (LAC) regions. However, compared to the Europe, Oceania, and North America regions, LA adoption remains relatively low. There is a need for an LA research critical mass in the LAC region (Cechinel et al., 2020), that requires identification of the LA field and capacity recognition within educational institutions. According to the Horizon Report 2022, LA has enormous potential to become a source of educational innovations to address various complex educational processes in the Mexican context (Sánchez-Mendiola, 2022). The main framework of this research is the Learning Analytics Latin America (LALA) project originated from SHEILA. This study reports the first phase of an ongoing investigation, in order to recognize the perceptions and expectations of the stakeholders involved in the educational processes, regarding the current situation of LA in a large Mexican university.

2 METHODS

The original LALA strategy was adapted to the context of a public HEI in Mexico, to recognize the perceptions and expectations of teachers, students and university authorities regarding the management and use of educational data towards the integration of LA (Hilliger et al., 2020). This document presents a qualitative approach that included semistructured interviews with deans of participating schools, and focus groups with a diversity of university stakeholders.

2.1 Participants and Instruments

The interviews were conducted with four school deans, and the focus groups included different types of participants: Administrative staff (35), students (32), and teachers (32). The schools included four disciplinary areas: physics, mathematics and engineering sciences; biology, chemistry and health sciences; social sciences; humanities and arts. The interview guides in Spanish were modified using words more commonly used in Mexico, to ensure that the participants understood all the questions (bit.ly/3kOm7MO).

2.2 Data analysis

Content analysis was used to process the information registered in the interviews and focus groups transcripts. Excel spreadsheets were used to organize the data in tabs, independently for each type of participant. The process was carried out in two stages: in the first, one-dimensional ideas from each participant were identified, and in a second stage the key concepts, common themes or ideas were identified based on what was expressed by all the participants in the group. Analysis of the perceptions and expectations of authorities' interview results were coded using a similar process by the qualitative researcher and the study group.

3 RESULTS AND DISCUSSION

General use of data. The perceptions of the different groups of participants are at the level of institutional and curricular analytics, and still distant from a systemic LA. The relative absence of the understanding of the LA construct in this large public Mexican university, constitutes a phenomenon similar to other Latin American universities, as Cechinel et al. (2020) and Hilliger et al. (2020) described, which has motivated the use of terms similar to the LA construct: "use of educational data".

Transparency, ethics and privacy. Regarding formal explicit consent about the use of , authorities and students stated they had little information on the mechanisms implemented by the university. In relation to the policies about access to data by teachers and students, students recognize that they are not certain about its existence. From the students' perspective, providing their data to teachers is risky, since it is personal data that can be misused if it is archived outside the formal protection of the academic institution. The foregoing also exposes the problem of data ownership, as discussed by Pardo & Siemens (2014).

Academic use of data. For the study populations, the use of educational data should be targeted to benefit academic training, that is, the application of LA should be considered for various activities within educational processes, for example: planning and creating courses, workshops and specific educational materials that respond to training needs. Another aspect highlighted by the participants about the use of educational data, was related to addressing low performing students, lagging and/or school dropouts, as has been reported by Pontual-Falcao et al. (2019).

Data-based feedback. Feedback is one of the central themes mentioned by participants in this study. Students prefer feedback to be done in person by their teacher or tutor. Teachers emphasize the qualitative aspects, beyond the quantitative ones, where the students' experiences throughout their learning process are taken into account, so that it is the student him/herself who identifies strengths and weaknesses.

Data-related actions. All participants considered that data obtained from teachers and students are helpful to generate actions to improve their performance. They pointed out the need to identify the reasons for dropout and academic lag in order to address these problems. Authorities mentioned the importance of collecting and including student input and balancing it with the interests of faculty and administrative staff. Teachers were concerned that the university "knows everything" about them, and even those who knew little about this issue realized the importance of data availability and use, and how data can become useful or dangerous when it is transformed to information and knowledge.

4 CONCLUSIONS

The initial analysis of the perspectives and expectations of authorities, teachers and students in this large public university about the use of educational data, allows recognition and identification of the context, and helps define a starting point in the systematization of LA processes. University authorities and administrative staff collect data for generating reports and making administrative decisions. Faculty and students recognize the importance of educational data and its potential use, but they are not clear on its operational use and have concerns regarding its ethical use and privacy issues. The university needs to address challenges to adopt LA: 1) it will be essential to integrate the data that is currently stored and managed separately in different instances and moments; 2) institutional policies must be implemented in agreement with current regulations; 3) it is crucial to disseminate information about LA in the community, with credible local examples about its benefits and drawbacks.

REFERENCES

Cechinel, C., Ochoa, X., Lemos dos Santos, H., Carvalho Nunes, J. B., Rodés, V., & Marques Queiroga, E. (2020). Mapping Learning Analytics initiatives in Latin America. *British Journal of Educational Technology*, *51*(4), 892–914. https://doi.org/10.1111/bjet.12941

Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y., Muñoz-Merino, P. J., Broos, T., Whitelock-Wainwright, A., Gašević, D., & Pérez-Sanagustín, M. (2020). Towards learning analytics adoption: A mixed methods study of data-related practices and policies in Latin American universities. *British Journal of Educational Technology*, *51*(4), 915–937.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics: Ethical and privacy principles. *British Journal of Educational Technology*, *45*(3), 438–450.

Pontual-Falcão, T., Ferreira, R., Lins Rodrigues, R., Diniz, J., & Gasevic, D. (2019). Students' Perceptions about Learning Analytics in a Brazilian Higher Education Institution. *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, 204–206.

Pontual-Falcão, T., Mello, R. F., Rodrigues, R. L., Diniz, J. R. B., Tsai, Y.-S., & Gašević, D. (2020). Perceptions and expectations about learning analytics from a brazilian higher education institution. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 240–249. https://doi.org/10.1145/3375462.3375478

Sánchez-Mendiola, M. (2022). Higher Education in Mexico. In K. Pelletier, M. McCormack, J. Reeves, J. Robert, & N. Arbino, *2022 EDUCAUSE Horizon Report, Teaching and Learning Edition* (pp. 45–46).

Using Multi-Modal Network Models to Visualize and Understand How Players Learn a Mechanic in a Problem-Solving Game

Zack Carpenter
University of Minnesota
carpe787@umn.edu

Yeyu Wang
University of Wisconsin
ywang2466@wisc.edu

David DeLiema
University of Minnesota
ddeliema@umn.edu

Panayiota Kendeou
University of Minnesota
kend0040@umn.edu

David Williamson Shaffer
University of Wisconsin
dws@education.wisc.edu

ABSTRACT: The incipient work in this poster aims to extend work on multi-modal learning analytics by exploring how blending think-aloud, eye gaze, and log data in network models informs how players learn a game mechanic. Preliminary models demonstrate differing patterns between players who have learned and have not yet learned the mechanic.

Keywords: Multi-modal learning analytics, network models, game-based learning, eye gaze.

1. INTRODUCTION

Video games are playful contexts that fuel learning through problem-solving (Gee, 2005) and provide traces of action for multi-modal learning analytics research (MMLA) (Emerson et al., 2020). By incorporating evidence from multiple data streams, MMLA offers an opportunity to understand deeper how problem-solving actions shape learning. Here, we use eye-gaze, log, and think-aloud data in multi-modal network models to understand how players learn a central game mechanic through problem-solving moves, such as noticing deviations from preference and searching for causal explanations. In doing so, we intend to contribute to the work on MMLA by answering the following question: *How do eye gaze and game actions provide markers of learning through problem-solving in a puzzle-based video game*?

2. METHOD AND DATA

We studied learning in the game *Baba is You* (Teikari, 2019). The physics of Baba is You are altered by moving text blocks. Figure 1a shows that players start as the white character (Baba) enclosed in the wall with the text WALL-IS-STOP. To win the level, the player must move Baba to push either the WALL, IS, or STOP text to “break” the rule. The player then could move through the wall and combine text to form the rule FLAG-IS-WIN. The player wins when they move Baba over the flag object. This work focuses on how players learn the STOP mechanic—that is, how players realize that impassable objects are caused by the WALL-IS-STOP (level 1) and FLAG-IS-STOP (level 2) rules.

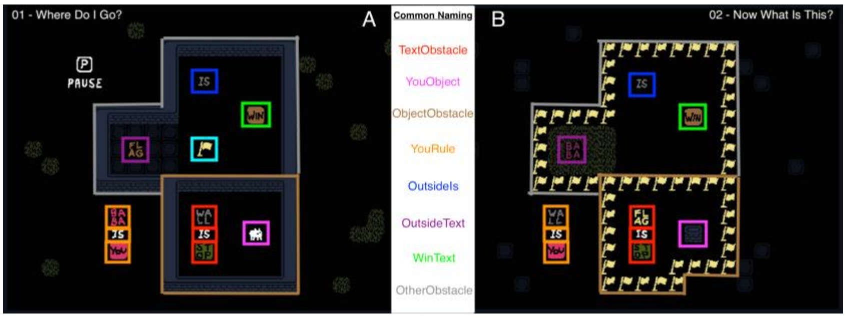


Figure 1: The Levels used to investigate learning the STOP mechanic. Objects with similar purposes across levels are marked with the same color and labeled to facilitate cross-level comparisons.

Data from 10 of the 18 recruited undergraduates are considered here due to attrition and equipment failure. In each of the two one-hour sessions, an Eye Link II eye tracker (SR Research) was calibrated, and students played while thinking aloud. We segmented the data streams to include moments players were trapped inside the initial wall/flag enclosure. This resulted in 27 cases across ten players on the first level and 32 cases on the second level. Trans-Modal Analysis (TMA)[1] was used to analyze the data. TMA is a novel extension of Epistemic Network Analysis (Shaffer et al., 2016) and Ordered Network Analysis (Tan et al., 2022). Descriptions of the three data streams incorporated into the models are as follows. (1) The think-aloud data identified 6 of the 10 players as learning the STOP mechanic. For example, a player categorized as "learned" hit the wall and said, "Oh, I can't get through. Oh, it's because WALL-IS-STOP (*breaks the rule)*". When players learned the STOP mechanic on the first level, subsequent trapped instances were coded as learned. (2) The codes briefly outlined in Table 1 were extracted from the log data. (3) Eye gaze was recorded at 250 HZ. Fixations on the colored areas of interest in Figure 1 represent the codes in the models.

Table 1: Description of Log Data Codes

Code	Description	Example in Log File
Start or restart	Player enters the level or restarts it from the beginning.	event_start; input_restart_
Deviation	Player can't get YouObject through ObjectObstacle.	input_up_ input_up_
Rule Break	Player moves a text block away from TextObstacle.	event_rule_remove_12:13:wall is stop
Passed Boundary	Player moves YouObject passed the ObjectObstacle.	change_update_baba:10:10:1 input_up_

3. PRELIMINARY FINDINGS: TMA MODELS OF TWO LEVELS

Integrating the eye gaze codes in Figure 1 and the log data codes in Table 1 resulted in the TMA model in Figure 2. We limit our discussion to salient patterns (the thicker lines in the models) related to learning the STOP mechanic. Figure 2A compares learned versus not-yet-learned on level 1.

[1] "ECR: Trans-Modal Analysis (TMA): A Mathematical and Computational Framework for Equitable Assessment of Multimodal STEM Learning Processes," National Science Foundation Grant DRL-2201723. TMA is an approach to constructing network models of complex problem-solving that incorporate connections of learning behaviors across different modalities. TMA uses a Temporal Influence Function (TIF) for each modality of data to account for different functionalities of modes.

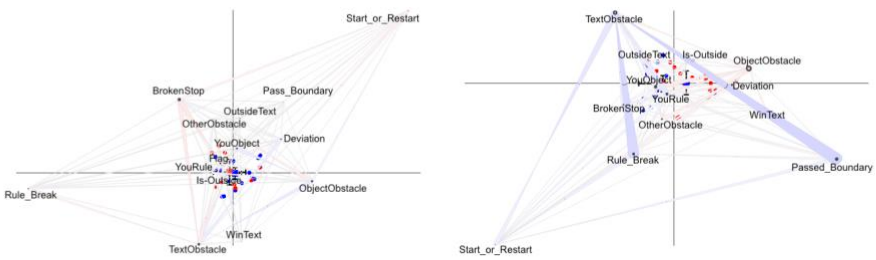


Figure 2: TMA models comparing groups on level 1 (left) and level 2 (right). Blue edges indicate patterns for the learned group and red edges indicate patterns for the not-yet-learned group.

It shows that looking from the TextObstacle to the ObjectObstacle and looking from the TextObstacle to experiencing the deviation is more common for the learned group. This indicates that the learned group interacts more with objects and text related to the STOP mechanic, as suggested by this group experiencing the deviation (i.e., hitting the wall) and searching for a cause (i.e., glancing at WALL-IS-STOP). Figure 2B compares learned versus not-yet-learned players on level 2. It shows that looking at TextObstacle and breaking the TextObstacle rule or passing the boundary is more common for the learned group. For the not-yet-learned group, looking at the ObjectObstacle and then experiencing the deviation is more common. Overall, players who learned the STOP mechanic on level 1 seemed to transfer this knowledge to a similar situation in level 2, and those who did not learn the mechanic on level 1 were looking at and engaging in similar actions as the learned group did on level 1.

4. DISCUSSION, LIMITATIONS, AND FUTURE DIRECTIONS

Two TMA models were used to compare players who learned and did not learn a game mechanic. On level 1, players who learned the mechanic made more connections between WALL-IS-STOP, the wall enclosure, and the deviation and transferred these experiences to FLAG-IS-STOP, the flag enclosure, and passing the boundary on level 2. These initial results make methodological contributions by using novel network models and could inform game designers to incorporate timely deviations. Apart from the small sample and limited level selection, some key limitations are that the TMA window size needs a stronger justification, and there were possibly too many variables to easily gauge connections. Next steps include devising data-driven methods for determining window size and extending the analyses to different levels.

REFERENCES

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-learning and Digital Media*, 2(1), 5-16.

Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, 51(5), 1505-1526.

Arvi Teikari. 2019. Baba Is You. Game [PC]. (March 2019). Hempuli Oy, Finland.

Tan, Y., Ruis, A., Marquart, C., Cai, Z., Knowles, M., Shaffer, D.W. (In Press): Ordered network analysis. In International Conference on Quantitative Ethnography. Springer.

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.

Distraction Analytics: Understanding Students' Distraction Patterns during Digital Learning

Jacob Choi[1] Lin Lin-Lipsmeyer[2] Chris Bigenho[3] Lujie Karen Chen[4]
[1]Marriotts Ridge High Schoo, Maryland USAl [2]Southern Methodist University, Texas USA
[3]Lewisville Independent School District, Texas USA
[4]University of Maryland Baltimore County, Maryland USA
[1]jacob.jh.choi@gmail.com [2]llipsmeyer@mail.smu.edu [3]bigenhoc@lisd.net [4]lujiec@umbc.edu

ABSTRACT: It is well-known that digital distraction and the resulting media multi-tasking may lead to impairment of learning. However, there needs to be a better understanding of how the distraction process unfolds over time and what aspects of the distraction are mostly detrimental to learning. This project analyzes the detailed distraction process with a dataset collected from a group of high school students who were presented with a "distractor" - an AI-based chatbot that was available to chat - while watching TED talk educational videos. We performed a series of temporal and semantic analyses of students' chat history with the bot to shed light on the patterns of distractions and their relationship with learning outcomes.

Keywords: chat bot, multi-tasking, learning analytics, temporal analytics, digital learning, cognitive load theory

1 MOTIVATION AND THEORETICAL BACKGROUND

Digital distractions are increasingly prevalent in learning environments with the young generation of students (McCoy, 2020). Numerous studies have found that students engaged in media multitasking while working on a primary learning task are more likely to have impaired learning outcomes (Dietz and Henrich, 2014). Those findings are supported by the cognitive load theory, which states that competing attention from multiple sources and multiple tasks (e.g., chatting) may add cognitive overload and negatively affect learning due to the limited capacity of working memory. This is especially relevant when learning new materials (Sweller, 2011). However, there is a lack of research that provides details on how different types of distraction patterns may lead to changes in cognitive load and thus impact learning. Instead of treating distraction as a binary condition, i.e., the student is either distracted or not distracted, we need new data and analytical tools to model the nuanced distraction patterns that unfold over time and assess its differential implication for learning outcomes.

2 DATASET, PREPROCESSING, AND FEATURE EXTRACTION

We use a dataset collected from 34 high school students (19 male and 15 female) in the Southwest part of the United States. Students were broken into two groups, each of whom being distracted by a chatbot while watching one of the educational TED talks. One of the videos was on dark matter, and the other was about happiness. The two video lectures were of comparable length (about 20 mins) and with similar lexical difficulty. The students were asked to take a 15-item multiple-choice test following each of the videos. Students were also asked to take notes as they would usually do. For our analysis, we used the timestamped chat logs and transcriptions and the post-video test scores. On average, students sent 3.4 messages per minute, waited 24.2 seconds between messages and sent 4.1 words per message.

To gain deep insights into the natural conversational dynamics and semantics, we segmented messages into episodes of conversational acts, which were clusters of chat messages that occurred within time intervals of 40 seconds. We then extract episode-level features that characterize those patterns with respect to the ***intensity*** and ***volume*** of the chat. Intensity features described how often the chat episodes occurred. We used the between-episode intervals, episode counts, and duration of the episodes as proxy measures. Volume Features described the amount of the chat, approximated by the number of words. Those episode-level features were then aggregated into video-level features for association analysis with learning outcomes as described below.

3 Relationship Between Distraction Patterns and Learning Outcomes

Figure 1 illustrates the relationship between the two chat features and their relationships with learning outcomes. As shown from the left figure, larger cumulative gaps between episodes (i.e., lower chat intensity) are associated with better learning. On the other hand, students who were engaged in longer chat episodes on average (as reflected in Episode Duration Mean features, right plot) tended to have worse learning outcomes. **Figure 2** (left) illustrates the relationship between the number of chat episodes and learning outcomes. At first glance, it seems counter-intuitive that more chat episodes were associated with better learning. Figure 2 (right) provides additional insights by clarifying the relationship between episode count (chat frequency/intensity) and episode word count (chat volume). As shown, there is an inverse relationship between how often they chatted and how long they chatted when they did chat. There were two types of students in terms of chat patterns (1) those engaged in many short chats (marked in red) and (2) those invested in a few long chat episodes (marked in blue). Putting those two plots together, we can understand that those engaged in a few lengthy chats seem to be worse off than those involved in many short chats.

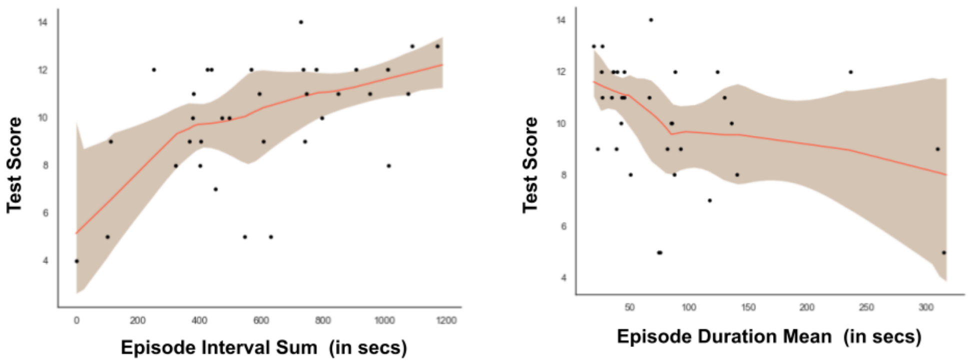


Figure 1: The relationship between chat intensity features and learning outcomes.

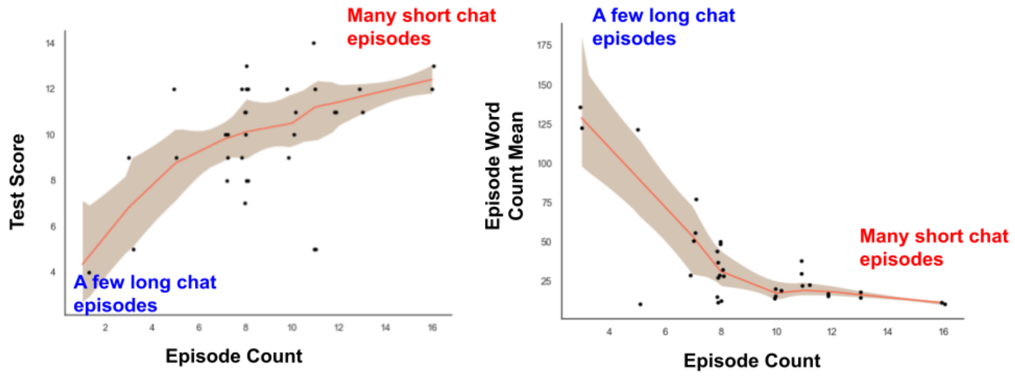


Figure 2: The relationship between chat volume features and learning outcomes.

4 Discussion and Future Work

This study presents a fine-grained analysis of distraction patterns and explores their relationships with learning outcomes. We note that students with distraction patterns characterizing frequent and/or long chats suffer the most in learning outcomes, which seems to resonate with cognitive load theory as those chat patterns likely induced higher demand on working memory, thus leading to more severe impairment of learning. By investigating the distraction process at a fine-grained level, this study provides nuanced evidence of the impact of digital distraction on learning outcomes. In addition, it opens up a new avenue of research to study the temporally and causally linked instructional events and observed distraction episodes and their interplay with learning outcomes. As part of future work, we plan to study the relationship between temporal patterns of pedagogical events systematically (e.g., para-linguistic features of the speaker, usage of visual aids, or employment of active learning techniques), students' note-taking behaviors, the temporal distribution of the knowledge units across lecture and their relationship with learning outcomes. We believe those insights may be used to inform design instructions that are adaptive to the young generation's dynamic and heterogeneous attention and distraction patterns.

REFERENCES

Dietz, S., & Henrich, C. (2014). Texting as a distraction to learning in college students. *Computers in Human behavior*, *36*, 163-167.

McCoy, B. (2020). Gen Z and Digital Distractions in the Classroom: Student Classroom Use of Digital Devices for Non-Class Related Purposes. Journal of Media Education, 11(2), 5-23.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), Cognitive load theory (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Interpretable neural networks vs. expert-defined models for learner behavior detection

Juan D. Pinto, Luc Paquette, Nigel Bosch
University of Illinois Urbana-Champaign
jdpinto2@illinois.edu, lpaq@illinois.edu, pnb@illinois.edu

ABSTRACT: Despite their powerful predictive abilities, neural networks' lack of transparency makes it difficult to interpret how they make decisions. Recent work has proposed approaches for increasing the interpretability of neural networks, including convolutional neural networks (CNNs). However, analysis of the meaning of discovered convolutional filters is yet to be done. Consequently, the present study explores interpretable CNNs to measure a well-studied learner behavior—that of gaming the system. Phase 1 of our research, presented in this paper, aims to answer: (1) How does the accuracy of a CNN model compare to an expert-created model for gaming-the-system classification? (2) What is the impact on performance when making CNN filters more interpretable via regularization? Our findings suggest that there is good reason to pursue interpretable machine learning models for detecting student behaviors. Future work will consist of an in-depth analysis of the specific patterns the CNN has identified in student actions.

Keywords: Interpretability, pattern mining, convolutional neural networks, gaming the system, learner behavior detection.

1 INTRODUCTION

Neural networks are increasingly used in educational contexts for tasks such as knowledge tracing (Sarsa et al., 2022) and learner behavior detection (Botelho et al., 2019). However, despite their powerful predictive abilities, neural networks' lack of transparency makes it difficult to interpret how they make decisions. In an effort to address this issue, Jiang & Bosch (2021) proposed an approach for increasing the interpretability of a specific type of neural network—convolutional neural networks (CNN)—by implementing regularization in the loss function that constrains the weights of the convolutional filters to be binary rather than continuous. Since each learned filter serves as a different pattern that the network is seeking in the data, making filter weights binary makes it far easier to understand which combinations of student actions the model considers important. This approach follows a philosophy of interpretability, which is fundamentally different than explainability methods such as SHapley Additive exPlanations (SHAP; Lundberg & Lee, 2017). Rather than conducting *a posteriori* analyses of inputs and outputs, Jiang & Bosch (2021) alter the inner workings of the model itself to make it easier for humans to understand.

While Jiang & Bosch (2021) demonstrated the effectiveness of their technique for discovering useful predictive patterns, they did not analyze the meaning of the discovered filters. To fill this gap, the present study explores interpretable CNNs to measure a well-studied learner behavior—that of gaming the system, or "attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material" (Baker & de Carvalho, 2008). Models of gaming the system have been created using different approaches, including machine learning, knowledge engineering, and a hybrid of the two (Paquette & Baker, 2019). Paquette et al. (2014) elicited details about the specific learner action patterns that experts consider when labeling gaming behavior. This provides an interesting baseline against which to compare an interpretable CNN, both in terms of accuracy and the specific patterns identified.

Phase 1 of our research, presented in this paper, aims to answer the following research questions: (1) How does the accuracy of a CNN model compare to an expert-created model for gaming-the-system classification? (2) What is the impact on performance when making CNN filters more interpretable via regularization? This project is about more than gaming the system or marginal performance improvements. Rather, we explore whether interpretable neural networks can provide valuable predictive information, while allowing us to understand the patterns they recognize in learner behavior. Future work will more specifically compare the CNN's discovered patterns to patterns identified through expert knowledge elicitation.

2 METHODS

To more accurately compare our results with prior studies, we used data previously reported in Paquette et al. (2014). It consists of sequences of actions from 59 students using the Cognitive Tutor Algebra system during an entire school year. Cognitive Tutor tasks students with solving multi-step mathematical problems and can provide on-demand hints. A total of 10,397 clips (i.e., student action subsequences) were previously labeled by an expert (Baker & de Carvalho, 2008) and contained 708 examples of gaming the system behavior (6.8%). Each clip contained a minimum of 5 actions and needed to last at least 20 seconds. If a clip was shorter than 20 seconds, additional blocks of 5 actions were added until the duration was at least 20 seconds. Overall, 84% of all clips contained 5 actions, while 12% contained 10 and the rest were various lengths. We used the same held-out test set as Paquette et al. (2014), consisting of 25% of the total data. The same study identified *constituents*, which are designed to capture elements of the students' problem-solving behavior that experts pay attention to when looking for gaming the system behavior: for example, whether a student reuses the same answer in a different part of the problem interface or the student enters consecutive similar answers. Each constituent is a binary feature (i.e., the constituent is either observed or not for a given action) and is designed to have a meaningful interpretation with regards to how students solve problems. We used these same constituents as our inputs when training the CNN.

We trained a neural network with a single convolutional layer. To make use of clips that were not exclusively 5 actions long, we included an adaptive max pooling layer that condensed the output from the convolutional filters into the same length, making it possible to feed them all into the same fully connected layer for prediction. We further divided our training data into randomly selected training/validation sets of 80%/20%. We separately trained three models to select an early stopping point before overfitting to the training data began. We then used the average number of epochs among these to train a final model on the combined training + validation set.

To answer RQ2, we conducted the process described above twice: once with a model without regularization, and once with a model using the regularization term described in Jiang & Bosch (2021). This regularization is minimized when the parameters of convolutional filters are either 0 or 1, which enables interpretation of a filter as a sequential pattern that matches action presence and absence.

3 FINDINGS AND FUTURE WORK

Based on the validation stage, we trained the model with regularization (the more interpretable one) for 300 epochs on the entire training set. It obtained a Cohen's kappa of .322 (accuracy=93.1%, precision=.490, recall=.278) on the held-out testing set. These results are comparable to the performance of the expert model described in Paquette et al. (2014) (kappa=.331, accuracy=88.7%, precision=.307, and recall=.528 on the held-out test set). It outperformed the machine learning model created by Baker & de Carvalho (2008), which achieved a kappa of .24 on held-out data (as reported in Paquette et al., 2014).

We also trained the model without regularization for 90 epochs, which obtained a kappa of .333 (accuracy=93.7%, precision=.582, recall=.261) on the held-out testing set. The accuracy improvement over the model with regularization was therefore very minimal. Given our goal of generating interpretable models, and the potential of our model with regularization to be more interpretable, these results are promising.

Our findings so far suggest that there is good reason to pursue interpretable machine learning models for detecting student behaviors such as gaming the system. CNNs in particular provide a powerful approach for analyzing sequential data that can be designed with both accuracy and interpretability in mind. To evaluate the claim of interpretability, the next phase of our project will consist of an in-depth analysis of the specific patterns the convolutional layer has identified in the student actions. As part of this analysis, we plan to compare these patterns with those defined as relevant to gaming-the-system behavior by a domain expert.

REFERENCES

Baker, R. S., & de Carvalho, A. M. J. A. (2008). Labeling student behavior faster and more precisely with text replays. *Proceedings of the 1st International Conference on Educational Data Mining (EDM)*, 38–47.

Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, *12*(2), 158–170. https://doi.org/10.1109/TLT.2019.2912162

Jiang, L., & Bosch, N. (2021). Predictive sequential pattern mining via interpretable convolutional neural networks. *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)*, 761–766.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Paquette, L., & Baker, R. S. (2019). Comparing machine learning to knowledge engineering for student behavior modeling: A case study in gaming the system. *Interactive Learning Environments*, *27*(5–6), 585–597. https://doi.org/10/gg2h95

Paquette, L., de Carvalho, A. M. J. A., & Baker, R. S. (2014). Towards understanding expert coding of student disengagement in online learning. *Proceedings of the 36th Annual Cognitive Science Conference*, 1126–1131.

Sarsa, S., Leinonen, J., & Hellas, A. (2022). Empirical evaluation of deep learning models for knowledge tracing: Of hyperparameters and metrics on performance and replicability. *Journal of Educational Data Mining*, *14*(2), Article 2. https://doi.org/10.5281/zenodo.7086179

A Trial Study on Understanding Operational Tendency of Device Use Based on Operation Logs during Home Study Using Japanese Authorized Digital Textbooks

Yukako Miyanishi, Toru Nagahama, Taku Kawada, Tatsuya Horita
Graduate School of Information Sciences, Tohoku University
miyanishi.yukako.q8@dc.tohoku.ac.jp, toru.nagahama.a7@tohoku.ac.jp,
taku.kawada.d4@tohoku.ac.jp, horita@horilab.info

ABSTRACT: In this trial, in order to understand the operational tendencies of each individual in home study using Japanese digital textbooks, we attempted to extract features from operation logs and classify individual tendencies based on the features. Principal component analysis was conducted using four metrics, and three features were extracted: frequent page movement, the length of time spent on clicks, and the number of device usages. Clustering based on the three features resulted in the following categories: little or no device use, short operation time and very high click frequency, operational tendencies that may have high engagement, and suspected of leaving devices unattended. Finding optimal values for the features extracted by PCA could help infer engagement from the operation logs.

Keywords: clustering; digital textbook; engagement; home study; learning log; learning styles

1 INTRODUCTION

Textbooks in Japan are certified by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to ensure their quality[1]. They were previously supplied in paper form, but are now legally permitted to be provided in digital form starting in 2019. This allows for analysis of learning logs, making them more learner-friendly. However, although there have been a few LA studies on digital materials (e.g. Owatari et al., 2020; Sun et al., 2018), there have been no studies on how learners use digital textbooks that have passed the MEXT's certification process. By analyzing data from digital textbooks used uniformly in Japan, it will be possible to obtain an overall picture of how textbooks are used in school education. This analysis could be used to improve targeted learning and instruction in the future. This study aimed to extract features of operation logs and classify individual operation tendencies based on the features, especially focusing on how they are used in home study.

2 METHODS

In this trial, in order to focus on the tendency of self-study operation without teacher intervention, we targeted home study. The target data consisted of operation logs during English home study (excluding weekdays from 8:00 am to 4:00 pm during class time) for approximately 10 months for first grade students at a public junior high school in Ibaraki Prefecture (71 students, around 12 years old, in three classes with one teacher in charge). We focused on four types of operation logs: page

[1] https://www.mext.go.jp/en/content/20210325-mxt_kouhou02-200000029_1.pdf

Table 1: Metrics used in PCA.

Metrics	Description	Mean (SD)	Median
Number of device usages	Number of device usages during the period	3.3 (3.6)	2.0
Device use length	Total seconds per device use	1,382.6 (1,111.5)	1,152.5
Click count	Number of operations per device use	28.7 (27.9)	24.7
Interval duration	Average time elapsed after each operation per device use	76.9 (81.5)	50.0
Pages accessed	Number of pages accessed per device use	7.7 (5.3)	7.0

browsing, display of pop-up contents (e.g., videos), extracted display of screen elements (e.g., figures), and enlargement of the entire page.

We excluded operations unrelated to essential ones, based on Miyanishi et al. (2022), specifically device use less than 1 second and operations considered as leave (more than 4,464 seconds). This resulted in 63 users (6,982 cases) included in the analysis.

To extract the features of the operations, a principal component analysis (PCA) was performed on the standardized metrics in Table 1. Next, based on the extracted components, individual operational tendencies were classified using k-means clustering (number of clusters: 4). Finally, statistical significance was tested by Kruskal-Wallis test for each cluster and the features were analyzed.

3 RESULTS AND DISCUSSION

Table 2 shows the principal component scores and the results of their interpretation. The results of the Kruskal-Wallis test showed statistically significant differences in all metrics ($p<.01$). Figure 1 shows the results of the multiple comparisons by Steel-Dwass test for each cluster.

Cluster 1 (n=30): The users in cluster 1 are considered to have not used the device very much because each metrics was small. **Cluster 2 (n=4):** The users in cluster 2 are considered to have performed many clicks that involved switching pages frequently because of the large number of clicks and pages accessed. While some of the clicks were intended to be like operations with a lot of enlarging the screen, others were considered to be page jumping in a short period that are not relevant to learning (zapping). **Cluster 3 (n=23):** The users in cluster 3 are considered to be less likely to zap or leave the device because they do not have as many clicks and access pages as users in cluster 2, and they do not spend as much time on operations as users in cluster 4. Assuming that the number of device usages reflects their proactivity, we can assume that they were relatively proactive in using the device and performing clicks that were essentially related to learning. **Cluster 4 (n=6):** The users in cluster 4 are considered to have been reluctant to operate the device due to the low number of device usages. In addition, they may have left the screen open for long periods of time because they often operated the device for long periods of time. However, it is also possible that they may have been engaged in other learning activities during the prolonged operation. This point was not examined in this trial.

This trial is only limited to understanding device operational tendencies. When inferring learning engagement, defined as “active use of devices” in this trial, from device operation logs in future studies, it is expected that users as represented by cluster 3 will have the highest engagement. In this study, we focused only on device operation logs, but in the future, it will be necessary to examine metrics more closely to estimate engagement, and to conduct analysis based not only on operation logs but also on the context of education and the theoretical frame.

Table 2: PCA results.

Variables	Loading		
	PC1	PC2	PC3
	Frequent operation on a wide range of pages	Long time device usage	Number of device use during the period
Number of device usages	0.18	0.30	**0.93**
Device use length	0.12	**0.69**	-0.21
Click count	**0.67**	0.02	-0.12
Interval duration	-0.22	**0.66**	-0.21
Pages accessed	**0.67**	-0.02	-0.17
Accumulated percent	0.39	0.71	0.89

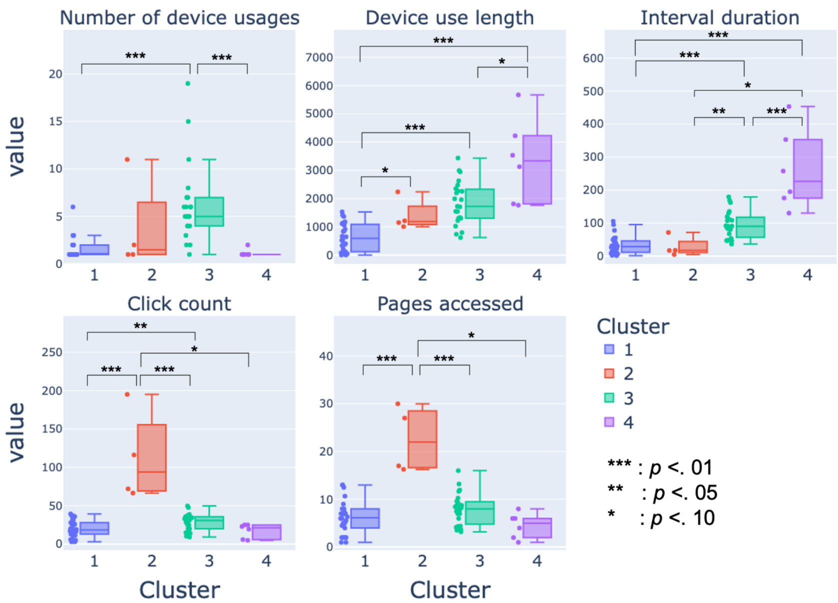


Figure 1: the statistics of the metrics for each cluster.

ACKNOWLEDGEMENT

This work was supported by Grant-in-Aid for Challenging Exploratory Research Number JP22K18611. This study was conducted in cooperation with the Tsukuba City Board of Education, Tokyo Shoseki Co.,Ltd., Lentrance Inc. for preparing data.

REFERENCES

Miyanishi, Y., Nagahama, T., Kawada, T., Kiyoto, K., Tonooka, T., Matsuda, R. & Horita, T. (2022). A Trial Study of Understanding the Status of Device Usage during Home Study Based on Learning Logs of Junior High School English Digital Textbooks. *2022 Autumn Conference of Japan Society of Educational Technology, 2022* (4), 98-105. https://doi.org/10.15077/jsetstudy.2022.4_98

Owatari, T., Shimada, A., Minematsu, T., Hori, M. & Taniguchi, R. (2020). Real-Time Learning Analytics Dashboard for Students in Online Classes. *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE),* 523-529. https://doi.org/10.1109/TALE48869.2020.9368340

Sun, J.-Y., Lin, C.-T. & Chou, C. (2018). Applying Learning Analytics to Explore the Effects of Motivation on Online Students' Reading Behavioral Patterns. *International Review of Research in Open and Distributed Learning, 19*(2), 210-227. https://doi.org/10.19173/irrodl.v19i2.2853

Investigation to Deploy Speech Rhythm Converters for English-Language Learners in Japanese High Schools

Sadao Hiroya, Koji Kamei, Takashi Hattori and Futoshi Naya
NTT Communication Science Laboratories
Nippon Telegraph and Telephone Corporation
sadao.hiroya.kv@hco.ntt.co.jp

ABSTRACT: Many native Japanese speakers tend to speak English with a mora-timed rhythm. This poster presents the implementation of an English speech rhythm training method for native Japanese speakers using speech rhythm conversion technology. Preliminary experiments at a Japanese high school confirmed the effectiveness of the speech rhythm conversion. The system was also extended to accommodate training on the internet.

Keywords: speech rhythm conversion, English-language learners, web-based training

1 INTRODUCTION

Many native Japanese speakers have difficulty communicating successfully in English with native English speakers. The difference between Japanese and English is not only in pronunciation, accent, and intonation but also in speech rhythm. English has a stress-timed rhythm and Japanese has a mora-timed rhythm (Grabe and Low, 2002), but native speakers of Japanese tend to speak English with a mora-timed rhythm. In speech rhythm training, it is common to imitate a native speaker's English speech. However, there are many differences other than speech rhythm, and it is difficult to know what to imitate, making effective training difficult.

We have developed a speech rhythm conversion technique that converts only the speech rhythm of English speech uttered by native Japanese speakers into a stress-timed rhythm. Since only the speech rhythm is converted, it is expected that trainees will be able to acquire the correct speech rhythm efficiently by imitating the converted speech. In this paper, we report on the development of an application for English speech rhythm training using this technology and the verification of its effectiveness through a demonstration experiment with Japanese high school students.

2 SPEECH RHYTHM CONVERSION TECHNIQUE

The speech rhythm conversion technique (Hiroya and Taguchi, 2020) consists of non-negative temporal decomposition (NTD) (Hiroya, 2013) and deep learning (DL). Speech contains both frequency and temporal information. NTD decomposes the vocal tract spectral sequence estimated from speech into a frequency spectrum and a temporal function for each phoneme (Figure 1). The temporal function is smooth with a value of [0,1]. DL devised a rule to convert the temporal function of native Japanese speakers into that of native English speakers using a corpus of English speech by 350 native Japanese and 700 native English speakers. The speech rhythm conversion is performed by replacing the temporal function of the English speech of the native Japanese speaker with that of the stress-timed rhythm obtained by DL. With this speech rhythm conversion, only the speech rhythm can be

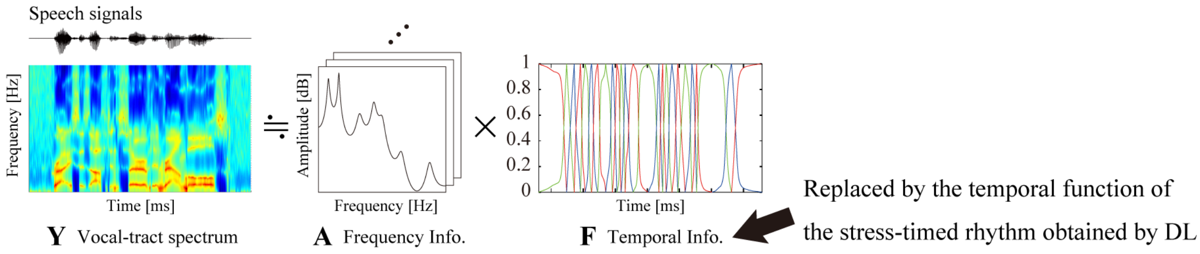


Figure 1: Non-negative temporal decomposition

changed to a stress-timed rhythm without changing the frequency information of the native Japanese speaker. It can be converted for any English speech. The difference in phoneme duration with native speakers was reduced from 42.7 ms to 27.7 ms after the conversion (Hiroya and Taguchi, 2020).

It is desirable to be able to display the speech rhythm score of the English speech during speech rhythm training. Therefore, in this study, the score was obtained by modeling the relationship between the change in the temporal function before and after the conversion and the speech rhythm rating by the native English speakers (out of 100 points) using DL.

3 IMPLEMENTATION OF TRAINING TOOLS

Speech rhythm conversion was demonstrated in a prototype implementation. The first client application was dedicated to Android. It connected directly from the client to a server on a Wi-Fi network. In the first experiment, conducted in 2021, five servers were deployed in a high school, where twenty-two students were divided into five groups, and the clients and servers were connected so that conflicts are unlikely to occur. However, when converting, it was necessary to occupy all the computation units of the GPU on the server, so sometimes conflicts occurred and there was a wait for the conversion.

The latest client application is available in an HTML browser (Figure 2). To allow for more participants without conflicts, we made some changes to the conversion server and introduced a load balancer to distribute conversion requests across a group of servers (Figure 3). As mentioned earlier, servers need to be occupied, but as long as all servers are not congested, requests can be assigned to servers in a round-robin fashion so that conflicts do not occur. The load balancer also automatically resends requests if all servers are busy. Therefore, it can handle multiple conversion requests from many students without changing the communication protocol between clients and servers. The entire system was then deployed over the Internet after a network security assessment to allow students to access it from home on their own devices.

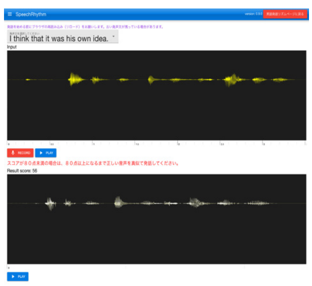


Figure 2: Speech Rhythm Conversion Client Application

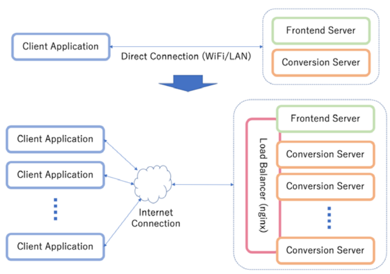


Figure 3: Extended Implementation of the Speech Rhythm Conversion Servers

4 PRELIMINARY EXPERIMENTS

Twenty-two second-year students (16M, 6F, 16 y.o.) from Miura Gakuen High School in Yokosuka, Japan, participated in the first experiment. The experiment took place in a classroom, and students were given Android smartphones with the application installed. In the pre- and post-training tests, students spoke the same 20 English sentences, and their pre- and post-test scores were compared to assess the training effect. For training, students uttered the English sentences differently from those on the test, and if their score was less than 80, they repeated the imitation of the speech with the speech rhythm converted until their score was 80 or higher. The training session lasted 30 minutes. In addition, five of the students who participated in the experiment were tested three months later to verify the sustainability of the training. During this period, the students were not given the application or encouraged to practice speech rhythm. Results of the experiment showed that training significantly improved the speech rhythm score ($t(21) = 2.08$, $p < 0.001$). The improvement rate was 12% (average score 59.2 to 66.3). The training effect was sustained after three months.

To verify long-term training effectiveness, the second experiment began in November 2022 and will last until February 2023, with eleven participants attending the same high school. Participants used the web application at home via the Internet on their personal devices. The English sentences for training are 20 sentences each week, and pre- and post-training tests would be performed.

5 CONCLUSION AND FUTURE PERSPECTIVES

This paper proposed a method for training English speech rhythms for native Japanese speakers using a speech rhythm conversion and confirmed its effectiveness. We also proposed its expansion to the Internet. By changing the model, speech rhythm training in other languages is also possible. In addition, synergistic effects can be expected by combining the training with other training such as pronunciation. Since the availability of the system has increased because of this deployment, we plan to develop training and analysis methods for educational settings, including high school teachers.

ACKNOWLEDGMENT

We would like to thank the participants of the experiments from Miura Gakuen High School. This work was supported by Council for Science, Technology and Innovation, “Cross-ministerial Strategic Innovation Promotion Program (SIP), Big-data and AI-enabled Cyberspace Technologies”. (funding agency: NEDO)

REFERENCES

Grabe, E., Low, EL. (2002). Durational variability in speech and the rhythm class hypothesis. *Proceedings of the 7th Conference on Laboratory Phonology*, 515-546.

Hiroya, S. (2013). Non-negative temporal decomposition of speech parameters by multiplicative update rules. *IEEE Transaction on Audio, Speech, and Language Processing, 21*(10), 2108-2117.

Hiroya, S., Taguchi, F. (2020). An English speech rhythm conversion method using non-negative temporal decomposition and deep learning, *Proceedings of the Spring Meeting of Acoustical Society of Japan*, 1079-1080.

Fusion of Explainable Recommender System and Open Learner Model

Yiling Dai, Brendan Flanagan, Kyosuke Takami, Hiroaki, Ogata
Kyoto University
[dai.yiling.4t, flanagan.brendanjohn.4n, takami.kyosuke.2z]@kyoto-u.ac.jp,
hiroaki.ogata@gmail.com

ABSTRACT: Explainable recommender systems enhance personalized learning by helping learners understand why a recommendation was made. As an important part of personalization, the learner model is being opened to the learners to promote meta-cognitive skills such as planning and monitoring learning. Educational explainable recommender systems and open learner models are discussed independently in the literature as they do not necessarily overlap. In this study, we propose a fusion system which provides explanations of recommended learning activities by visualizing the concept hierarchy to the learner. We describe the design of the system and the learning effects we can expect.

Keywords: educational recommender system, learner model, concept hierarchy, explainable recommender system, meta-cognitive skills

1 INTRODUCTION

Recommender systems provide personalized guidance for the next learning activity based on learners' understanding of the knowledge. Providing explanations for the recommendation can increase the transparency of the system, improve learners' trust towards the system, and promote learners' acceptance of the recommendation (Khosravi et al., 2022). From another perspective, personalization in technology-enhanced learning relies on the construction of a learner model which holds the information about the learner's current knowledge states, learning style, and so on (Bull, 2020). Beyond the primary role of enabling the system to generate personalization, opening the model to learners can promote meta-cognitive skills such as monitoring, planning, and reflection in learning (Bull, 2020).

Both explainable recommender systems and open learning models share the purpose of enabling learners to understand their learning progress and how the system supports their learning. However, these two concepts were insufficiently discussed in the same research context as: a) recommender systems do not necessarily provide explanations based on the learner model; and b) open learner models do not necessarily make recommendations on learning. In this study, we are interested in knowing whether we can achieve both benefits if a recommender system explains its recommendations based on the learner model it builds. Several works have attempted to incorporate the element of learner models into the explanations of recommended learning activities (Abdi et al., 2020; Barria-Pineda et al., 2021; Dai et al., 2022). However, these studies did not adopt graphical presentation of the learner model and did not explore how the topic-based explanation improves the learning from a meta-cognitive perspective. To address this, we propose a learning activity recommender system which explains its recommendations via visualizing the concept hierarchy of its learner model.

2 PROPOSED SYSTEM

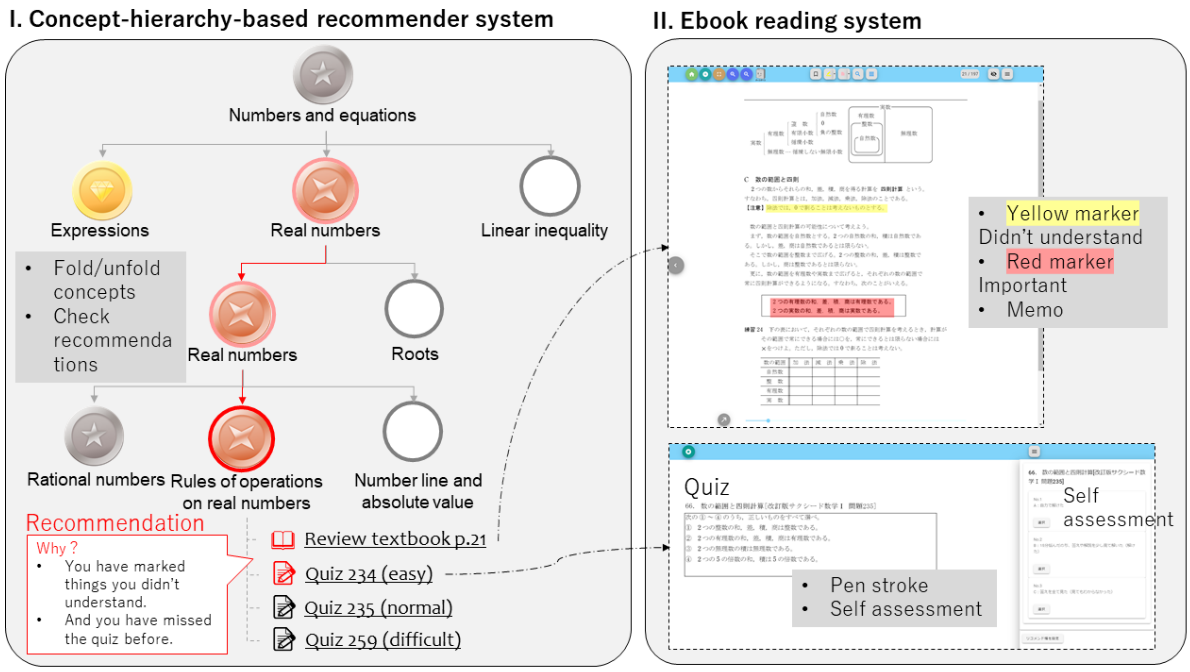


Figure 1: System design

We construct the learner model on a concept hierarchy where the nodes represent important concepts in this subject and the edges represent "part-of" relationships between concepts. As illustrated in the left part of Figure 1, the concept of `Numbers and equations` contains sub-concepts such as `Expressions`, `Real numbers`, and so on. At the lowest level, the concepts are associated with learning materials including specific pages in the textbook and the quizzes with different levels of difficulties. We use badges to indicate leaners' completion rates of the concepts, where gold indicates high, silver indicates medium, bronze indicates low, and white indicates zero completion rates, respectively. We assume that the concept hierarchy help learners to manage their knowledge--- navigating among available learning contents, monitoring knowledge acquisition, and fostering an overall perspective of the domain knowledge.

The primary purpose of the system is to support the learning of the concepts such as reviewing the textbook and solving quizzes (Flanagan & Ogata, 2018). We extend Flanagan et al.'s (2019) learning behavior modeling framework to capture various learning activities with a more fine-grained level adapted from Bloom's taxonomy (Swart, 2010). As illustrated in the right part of Figure 1, the links under a concept redirect the learners to an ebook reading system, where the learners can mark the content they find difficult or important, take memos, write down their answers, and report their self-assessment. We devise a simple incrementing method to estimate learners' completion rate of each leaf concept in the hierarchy based on the learning activities. For example, adding a marker is an action of identifying knowledge, which indicates the recognition of knowledge and gets a score of 1. Besides, deleting a marker is an action of reviewing the knowledge, which indicates the comprehension of knowledge and get a higher score of 2. After computing the completion rates of the leaf concepts, we propagate the completion rates to the ancestor concepts.

For the recommendation, we devise a simple incrementing method to compute the recommendation score of the concepts. If learning activities (e.g., undeleted marker, unattempted quiz) requiring remedial actions happen, we increment the recommendation score of the concept. We then adopt the greedy algorithm to rank the concepts based on the necessity of remedy. As illustrated in Figure 1, we display and explain the recommendation by highlighting the concept and adding textual information of the contributing learning activities.

3 CONCLUSION AND OUTLOOK

In this study, we proposed a learning activity recommender system which provides explanations by opening the concept hierarchy to the learners. As a fusion of explainable recommender system and open learner model, we expect the learning effects of both sides. In the future work, we plan to evaluate the system's effects on improving the acceptance of the recommendations and the meta-cognitive skills through real-life experiments.

REFERENCES

Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020). Complementing Educational Recommender Systems with Open Learner Models. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 360–365. https://doi.org/10.1145/3375462.3375520

Barria-Pineda, J., Akhuseyinoglu, K., Želem-Ćelap, S., Brusilovsky, P., Milicevic, A. K., & Ivanovic, M. (2021). Explainable Recommendations in a Personalized Programming Practice System. *Proceedings of International Conference on Artificial Intelligence in Education*, 64–76.

Bull, S. (2020). There are Open Learner Models About! *IEEE Transactions on Learning Technologies*, *13*(2), 425–448. https://doi.org/10.1109/TLT.2020.2978473

Dai, Y., Takami, K., Flanagan, Brendan, & Ogata, Hiroaki. (2022). Investigation on Practical Effects of the Explanation in a K-12 Math Recommender System. *Proceedings of the 30th International Conference on Computers in Education*.

Flanagan, B., & Ogata, H. (2018). Learning Analytics Platform in Higher Education in Japan. *Knowledge Management & E-Learning*, *10*(4), 469–484.

Flanagan, B., Rwitajit, M., Gökhan, A., Wang, J., & Ogata, H. (2019, March). Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments. *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)*. http://hdl.handle.net/2433/243240

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Swart, A. J. (2010). Evaluation of Final Examination Papers in Engineering: A Case Study Using Bloom's Taxonomy. *IEEE Transactions on Education*, *53*(2), 257–264. https://doi.org/10.1109/TE.2009.2014221

Integrating Learning Analytics-driven Practices in Special Needs Education

Yuko Toyokawa, Rwitajit Majumdar, Izumi Horikoshi, and Hiroaki Ogata
Kyoto University
toyokawa.yuko.59t@st.kyoto-u.ac.jp, dr.rwito@gmail.com, horikoshi.izumi.7f@kyoto-.ac.jp, Hiroaki.ogata@gmail.com

ABSTRACT: Information and Communication Technology (ICT) and digital learning tools have been widely used as assistive tools to support unique needs in special needs education. Such technology-enhanced learning environments enable us to apply learning analytics (LA) to process log data to support learning and teaching for different needs. However, the utilization of logs in learning practices in special needs education is still under-researched. This paper proposes how Learning and Evidence Analysis Framework (LEAF) can be applied in special needs education and discuss opportunities and challenges in conducting evidence-based learning in the distinctive learning environment using the concept of Universal Design for Learning (UDL).

Keywords: Learning Analytics, special needs education, LEAF system, learning logs, Universal Design for Learning

1 BACKGROUND: UNIVERSAL DESIGN FOR LEARNING AND SPECIAL NEEDS EDUCATION

The promotion of individually optimized learning support is attracting attention. However, each learner is different, and in a learning environment that involves students with special needs education, learning becomes even more complex. Research in special needs education, therefore, required systematic guidelines for identifying the types and the levels of evidence for scientific and effective evidence-based practices (Odom, et al. 2005). Universal Design for Learning (UDL) is widely used as a guideline that proactively accommodates the diverse and variable needs of all learners. It systematically designs learning plans with the intention of addressing individual differences of learners and promoting the inclusion of all learners (Rose, 2000). The concept of UDL is applied to the proposed approach of Learning Analytics-driven practices in special needs education using the Learning and Evidence Analysis Framework (LEAF) to support learners with different needs. LEAF is a framework to support an evidence-based education to accumulate, analyze, and predict the resulting learning behavior and outcomes to support various stakeholders in their own context (Ogata, et al. 2018). In regular classes, most of the curricula do not accommodate the diversity of each individual, while in special needs education, a curriculum is created that suits each individual needs. LEAF can be adapted to diverse learning environments of special needs education as a scaffolding for learning support that accumulates “Yes, I did it!” of each learner.

2 LEAF FOR SPECIAL NEEDS EDUCATION

2.1 LEAF and its potentialities

Teachers choose an optimal method from various ones and use the one that is considered to be optimal for each individual. These days, there are many assistive tools and applications available that can be used to support individual “Yes, did it!", but what is missing is visualizing and sharing logs. Teachers cannot observe learning initiatives at home, and parents cannot observe initiatives at school on a daily basis. It is also important to provide opportunities for learners to be able to learn by themselves anytime, anywhere. LEAF provides a learning environment and a place for reflecting and confirming that builds up “Yes, I did it!”. There are four components of LEAF: a Learning Management System (LMS) like Moodle, an e-book reader called BookRoll, a Learning Record Store (LRS) to accumulate and encode logs from BookRoll, and a set of LA dashboards called LOG PALETTE that visualizes the encoded logs. Teachers can upload visual and auditory materials integratedly in BookRoll, and learners can zoom in and out on the screen. Learning facilitating functions such as markers can be used for highlighting critical information to emphasize the information and memo can be used for annotation purposes. Users can also write directly on the BookRoll memo interface using a touch pen or fingers. Learning logs are visualized in LOG PALETTE, providing important support for evidence-based approaches in the context of special needs education. Learning support with LEAF is intended for students with special educational needs (e.g., learning disabilities) and teachers, as well as other stakeholders such as homeroom teachers, parents, and school administrators.

2.2 Pilot study

BookRoll in LEAF has been used for daily learning activities in a resource room in an elementary school (Toyokawa, et al., 2022). The resource room is a place where students with milder disabilities or difficulties attend to receive additional support while registering in a regular class. LOG PALETTE has been used to share their learning process and outcomes with teachers and parents (not all parents) and analyze data. For example, a log of a child's direct handwriting behavior (ADD, UNDO, REDO) for one semester revealed two patterns of handwriting behavior. One is no modification and the other is a repeated undo/redo pattern. Giving one example of the later pattern, it shows that undo and redo were mainly seen at the character level. There were six UNDO and REDO kanji (kanji), and katakana twice. Hiragana was undone twice, one without REDO and one rewritten to Kanji. At the sentence level, pen stroke analysis revealed that the names of friends and the words “everyone”, “this member” and “friend” appeared in the sentences, and they were erased or rewritten several times. These are examples of what pen stroke analysis possibly presents, such as learners' writing difficulties, feelings, and what they value in their daily lives. These logs can be shared with teachers, homeroom teachers, and parents, and can be used as material for discussing the child's growth and subsequent support. Figure 1 shows the framework of the LEAF and sample logs from LOG PALETTE pen stroke analysis.

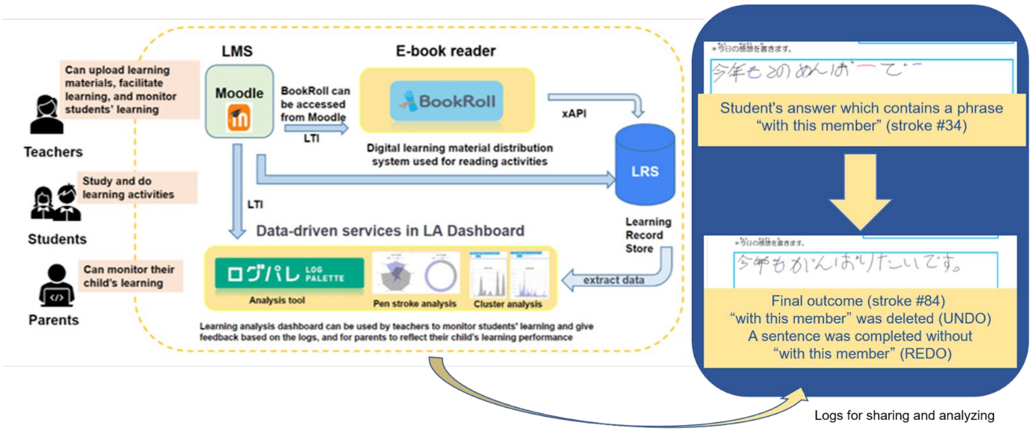


Figure 1: The framework of LEAF and logs for sharing and analyzing

3 CONCLUSIONS AND FUTURE WORK

Our aim is to provide a data-driven special educational environment in which we can find learners' stumbling points and special abilities from their learning logs and use them as materials for judging the contents of support, which can provide confidence to those students and support. Log accumulation and visualization provide learners with special needs and their families with along with other stakeholders various possibilities to maintain continuous and seamless support. For future challenges, first, in terms of system affordance and dashboard design, conciseness and attractiveness, such as variation of colors and font should be considered to increase accessibility so that anyone can understand at a glance. Moreover, additional auditory options shall be considered as alternatives for visual information. Second, it would be an issue to provide continuous support to teachers with the cooperation of researchers and system developers. The teacher's guidance greatly affects the growth and development of learners. In addition to improving the expertise of teachers regarding special needs, support using technology requires knowledge of operating the system and data literacy. Lifelog maintenance and transfer are the last issues needed to be mentioned. Special needs support requires constant review and updates of data. And the presentation of the log is asked throughout their lives. Securely managing learner logs across schools and institutions and seamlessly transferring data into the future is essential. While collaborating with stakeholders including the learner's families, our ultimate goal is to establish a learning environment that can provide optimal individualized support over the long term, and pursue more flexible and practical educational support.

REFERENCES

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional children*, *71*(2), 137-148.

Ogata, H., Majumdar, R., Akapinar, G., Hasnine, M. N., & Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning. In *26th ICCE 2018* pp. 493-496

Rose, D. (2000). Universal design for learning. *Journal of Special Education Technology*, *15*(3), 45-49.

Toyokawa, Y., Majumdar, R., & Ogata, H. (2022). Learning Analytics Enhanced E-book Reader in a Japanese Special Needs Class. In *IEEE ICALT* (pp. 274-278).

Toward Trustworthy Explainable Recommendation: Personality Based Tailored Explanation for Improving E-learning Engagements and Motivation to Learn

Kyosuke Takami, Yiling Dai, Brendan Flanagan, Hiroaki Ogata

Kyoto University

[takami.kyosuke.2z, dai.yiling.4t, flanagan.brendanjohn.4n]@kyoto-u.ac.jp,

hiroaki.ogata@gmail.com

ABSTRACT: Explainable recommendation, which provides an explanation about why a quiz is recommended, help to improve the transparency, persuasiveness and trustworthiness. Recently, some explainable recommenders were proposed in the educational field. However little research focused on the tailored intervention. We proposed personality based tailored explanation for improving engagements and motivation to learn. Students are clustered into different segments based on Big Five personality traits and tailored explanation for recommended quizzes are provided based on their personality trait. As a preliminary result we found significant click rate on recommended quizzes with tailored explanation compared with control group. This result suggests that tailored explanation will be effective for improving engagements, motivation to learn and trustworthiness.

Keywords: Tailored intervention, Big five personality, explainable recommendation, K-12

1 INTRODUCTION

Explainable recommendation, which provide explanations about why an item is recommended, can improve transparency, persuasiveness and trustworthiness (Zhang and Chen 2020). Recently, some explainable recommenders were proposed in the educational field. However, the way of generating explanation are limited such as explanations about difficulty, learning history or relevance knowledge for recommended quizzes and not consider students' cognitive perception or personality. In the public health research, tailored interventions, which are designed to reach one specific person based their unique characteristics, have been shown to be effective for behavioral change (Sohl and Moyer 2007). In the perception survey of explainable recommender, some students were convinced by the explanations than others (Takami et al. 2022), suggesting that the explanations should be more tailored to the individual student. This study draws inspiration from tailored intervention research in public health research, we propose personality based tailored explanation as one of explanation for recommended quizzes. We show the preliminary results indicating the effectiveness of our personality based tailored explanation approach for e-learning engagements. This approach expected to improve not only engagements but also motivation to learn and trustworthiness.

2 OUR APPROACH

First, we perform clustering on the Big Five personality traits and good at math scales by k-Means. 217 high school students in mathematics course were divided into optimal three groups labeled Diligent (high consciousness), Fearful (high neuroticism) and Agreeable (high agreeableness) as shown Figure 1. We prepared three different explanation for each three group according to previous personality based persuasiveness study in psychology (Wall et al. 2019), e.g. Diligent tend to be persuaded by authority. Thus, for Diligent cluster, Authority based explanation "Three of your top achievers have solved" are generated from top achievers learning log as shown Figure 2. For Fearful and Agreeable cluster, Commitment and Peer explanation are also generated from learning log for each cluster. We implemented this explanation generation function into existing explainable recommender system.

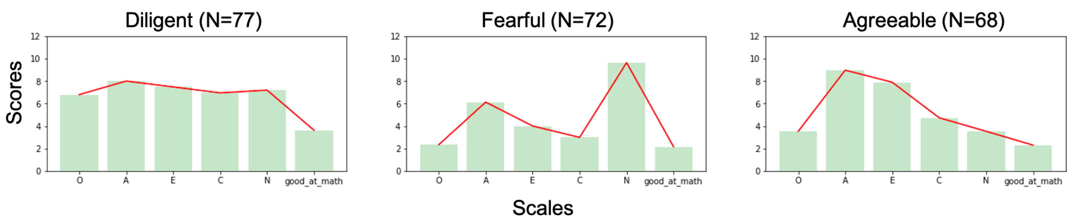


Figure 1: Clustered personality traits

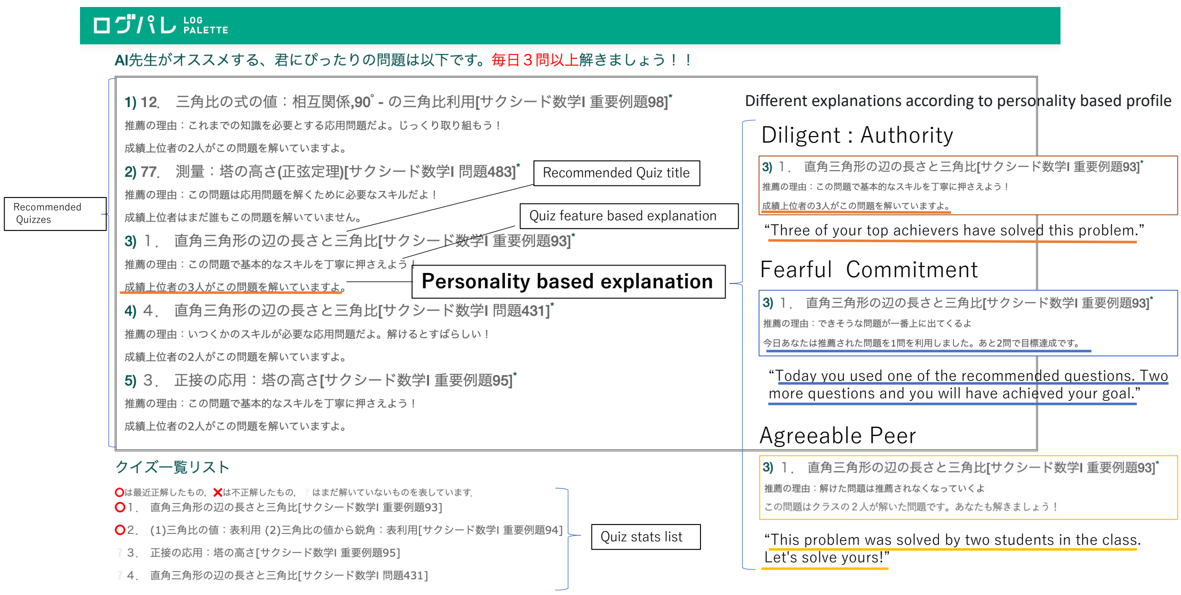


Figure 2: Screenshot of recommender UI

3 PRELIMINAL RESULTS

We conducted an 18 day A/B test to compare usage between intervention group (n= 106) with personality based explanation and control group (n= 111) without personality based explanation. As the blue and red colors in the Figure 3 show, the number of access and the number of clicks on recommended quizzes are higher in the in the intervention group than in the control group. Table 1 shows the number of view (accessed on recommendation page) and the number of clicked on the recommended quizzes. Acceptance rate in tailored intervention group (with personality based explanation) is 3.9 times higher than control group (without personality based explanation). These results suggest the positive effectiveness of our personality based explanation to engagements and motivation to learn. A Chi-square test and the results revealed significant difference among View count and Click count (χ^2(2)=40.5057, P<0.001). This analysis revealed that tailored intervention group significantly tend to click the recommended quizzes more than control group.

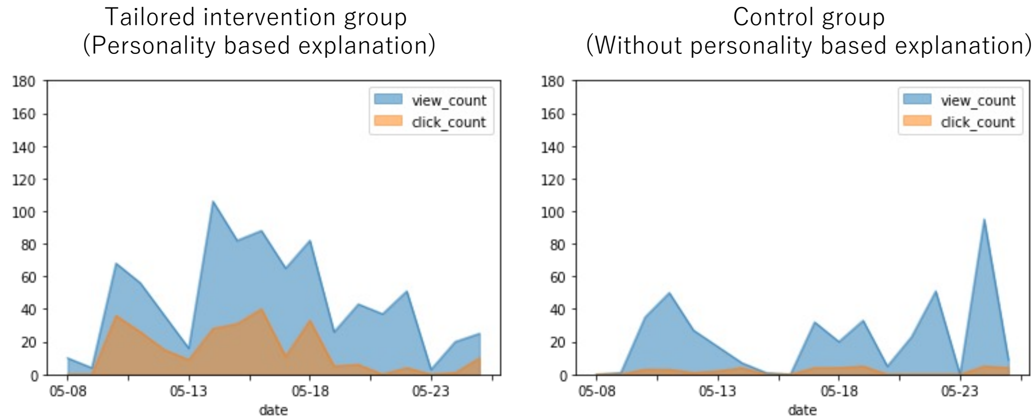


Figure 3: Overview of Recommender Usage

Table 1: Recommender Usage: Summary of system Accessed and Clicked counts

Condition	View count (accessed)	Click count (on recommended quizzes)	Acceptance Rate
Intervention	818	255	31.2%
Control	406	35	8.6%

Acceptance Rate = click count / view count

4 CONCLUSION

Explanations about why an item is recommended is important in explainable recommendation in educational setting for improve persuasiveness and trustworthiness. However, the various ways of explanation are possible. We propose personality based tailored explanation for improving engagements, motivation to learn and trustworthiness

REFERENCES

Sohl, Stephanie J., and Anne Moyer. 2007. "Tailored Interventions to Promote Mammography Screening: A Meta-Analytic Review." *Preventive Medicine* 45 (4): 252–61.

Takami, Kyosuke, Yiling Dai, Brendan Flanagan, and Hiroaki Ogata. 2022. "Educational Explainable Recommender Usage and Its Effectiveness in High School Summer Vacation Assignment." In *LAK22: 12th International Learning Analytics and Knowledge Conference*, 458–64. LAK22. New York, NY, USA: Association for Computing Machinery.

Wall, Helen J., Claire C. Campbell, Linda K. Kaye, Andy Levy, and Navjot Bhullar. 2019. "Personality Profiles and Persuasion: An Exploratory Study Investigating the Role of the Big-5, Type D Personality and the Dark Triad on Susceptibility to Persuasion." *Personality and Individual Differences* 139 (March): 69–76.

Zhang, Yongfeng, and Xu Chen. 2020. "Explainable Recommendation: A Survey and New Perspectives." *Foundations and Trends® in Information Retrieval* 14 (1): 1–101.

Perspectives of Multimodal Data Sharing and Privacy in VR Learning Rooms

Qinyi Liu
University of Bergen
qinyi.liu@uib.no

André Rabello-Mestre
University of Bergen
andre.mestre@uib.no

Mohammad Khalil
University of Bergen
mohammad.khalil@uib.no

ABSTRACT: With the emergence of learning analytics, data-sharing practices and privacy have become a major concern for different stakeholders in educational settings. Recently, this issue has been accentuated by the rapid development of virtual reality (VR) tools which, because of their multimodal nature, are likely to raise further concerns regarding student privacy. In this context—of quick technical development and evolving data-privacy frameworks—it is critical to have a clearer grasp of the privacy perceptions of different stakeholders regarding the use of VR in education. To address this issue, this study looks at how PhD candidates in education perceive VR learning rooms from the positionality of students, instructors, and researchers. The purpose of this poster is to investigate the privacy perceptions of 30 PhDs, exploring their different identities and roles regarding data collection and privacy

Keywords: Multimodal learning analytics (MMLA); virtual reality (VR); immersive virtual environments; student privacy

1 INTRODUCTION

Privacy issues are a growing concern in the field of learning analytics (LA) and scrutiny over data sharing practices and student privacy continues to grow. An important reason behind this is that poor handling of privacy issues has hindered the development of the field (Prinsloo et al., 2022). It is also evident that future progress of LA will have to address the technical and legal challenges associated with the ethical use of student data. It is in this context that we are also seeing the rapid adoption of virtual reality in educational settings. Platforms such as Second Life, Active Worlds, and Mozilla Hubs have attracted considerable attention from educators (Reisoğlu et al., 2017) and have been employed to promote "alternative learning ecologies" that, among other things, potentialize student reflexivity and promote experiential learning (Gamage et al., 2011; Kuznetcova & Glassman, 2020). However, while immersive virtual environments offer a plethora of possibilities in designing new spaces for learning, their multimodal nature presents further challenges for those wishing to engage in data sharing and privacy inquiry using these tools.

2 CONTRIBUTION

Unfortunately, it remains the case that privacy issues are seldomly addressed in the educational context where LA is advancing (Prinsloo et al., 2022). When it comes to multimodal data-collection this problem is compounded, and systematic reviews have shown the challenge of privacy is consistently overlooked (Alwahaby et al., 2021). For these reasons, this research seeks to contribute to our understanding of privacy in MMLA by investigating the data-sharing perceptions of PhD candidates in VR environments. Because PhDs often perform multiple academic roles – e.g., coursework, teaching and research duties – they have a unique perspective on both the value and the privacy cost of educational data that is valuable for this pilot study.

The contribution of our research can be summarized as follows: (1) an investigation of the privacy expectations of PhDs candidates in relation to multimodal VR learning spaces; and (2) an investigation of how academic roles inform diverging perspectives of data sharing and privacy.

3 METHODOLOGY

3.1 Participants and Ethics

We planned to recruit a sample of 30 PhD students from two universities who had undertaken teaching assignments. As the purpose of this pilot study is to understand whether perceptions of data sharing and privacy in VR learning classrooms may differ across academic roles, PhDs who have undertaken teaching duties have multiple experiences and perspectives from students, instructors, and researchers. Therefore, the group of doctoral candidates is well-positioned to take on the variety of academic perspectives needed in this preliminary study. Prior to the start of the experiment, each participant will be given written informed consent to participate in the study. As the experiment does not require the collection of any personal information, the study does not require special ethical approval under Norwegian law.

3.2 Experiment Design

The experiment will take place in a small VR lab at the University where participants will be briefed on the various steps of the procedure prior to the experiment. Once participants are ready to start, they will put on a commercial VR device and be asked to enter an immersive VR application designed in Mozilla Hubs for the purpose of the study. The virtual environment consists of a multimodal data-literacy learning room. In this open-ended space, they will learn about the various ways in which MMLA collects data from participants, with different sections of the space offering materials in the form of texts, graphics, videos, audio, and GIFs. In the following example, we depict a section of the experimental space with information about biometric data collection (Figure 1). Participants will learn through this VR experiment that immersive virtual environment, when combined with multimodal devices, can be extremely rich for the purpose of LA: many previously uncollectable data such as facial expressions, muscle tension, and ambient light, can be captured. At the end of the VR task, participants will fill out an online questionnaire using a tablet. The whole process is expected to take about 30 minutes.

Figure 1: Virtual space designed for the experiment describing eye-tracking data collection.

3.3 Survey

As previously mentioned, participants will be asked to fill out a questionnaire at the end of the experiment. The questionnaire will be divided into three parts: their perceptions of data sharing and privacy as students regarding VR learning classrooms, their perceptions of data sharing and privacy as teachers regarding VR learning classrooms, as well as their perceptions as researchers. The purpose of the questionnaire is to understand their expectation of data sharing and privacy in the VR learning room from different roles. The questionnaire using psychometric scales is excerpted below (Table 1).

Table 1: Excerpt from the survey on students' perspective of MM data sharing in VR.

Data type	Description	Willingness to share
Biometric Data	includes the user's eye movements, heart rate, facial expressions, and muscle tension	1-5
Environmental Data	Such data pertains to the user's surroundings, such as temperature, lighting, and sound	1-5
Behavioral Data	behavioral patterns, such as track how one studies and interacts with the virtual environment	1-5
User IDs and Passwords	these are used for logging in to the various virtual worlds that users will be interacting with	1-5
Time and Location Data	where do you use the VR equipment, such as stores or restaurants	1-5

4 EXPECTED RESULTS AND FUTURE WORK

We expect to see a significant discrepancy between participants' perceptions of data privacy when different academic roles are taken into consideration. While doctoral candidates are well positioned to understand the educational advantage of multimodal data collection as they relate to teacher inquiry and learning analytics, we expect that participants will also be hesitant to consent to their data in a variety of use cases. In the future, we will include a broader range of educational data stakeholders (e.g., undergraduate students) to conduct the experiment.

REFERENCES

Alwahaby, H., Cukurova, M., Papamitsiou, Z., and Giannakos, M. (2021). The evidence of impact and ethical considerations of Multimodal Learning Analytics: A Systematic Literature Review.

Kuznetcova, I., & Glassman, M. (2020). Rethinking the use of Multi-User Virtual Environments in Education. *Technology, Pedagogy and Education*, 29(4), 389–405.

Gamage, V., Tretiakov, A. & Crump, B. (2011). Teacher Perceptions of Learning Affordances of Multi-User Virtual Environments. *Computers & Education*, 57(4), 2406–2413.

Prinsloo, P., Slade, S., and Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*, 53, 876–893.

The Role of Cognition and Metacognition in Data Science Problem Solving: Insights from a Field Study

Maryam Alomair[1, 2], Shimei Pan[1], Lujie Karen Chen[1]
University of Maryland Baltimore County[1], King Faisal University[2]

ABSTRACT: Data Science (DS) is an interdisciplinary field with applications in many domains. Data Science Problem Solving (DSPS) requires learners to hone in higher-order reasoning and decision-making skills to apply them to solve real-world problems. Mastering DSPS requires a careful design of educational interventions to provide learners with ample practice opportunities. This study is the first attempt to explore a set of Caselets - bite-size self-paced scalable case studies to support the deliberate practice of DSPS with graduate-level students in a university setting. Through various predictive models, we note that prior knowledge and self-reported metacognition play a major role in predicting learning gain. Moreover, a graph-based analysis reveals interesting structures of self-reported metacognition, prior knowledge, and Caselet's performances at varying granular levels. We discuss the implication of those analytics to future iterations of pedagogical design and data collection.

Keywords: Data science problem solving, Data science education, Cognition, Metacognition, Knowledge Components

1 INTRODUCTION

Data Science (DS) education is often biased toward teaching component skills, including a conceptual understanding of methods and the programming skills involved in data wrangling and modeling, which prepares learners mostly to know WHAT and HOW. However, this is quite different from what learners must do in the real world, where they solve DS problems by answering WHAT, WHEN, and WHY. The difference between teaching component skills and problem-solving skills is a giant and challenging gap (Rittle-Johnson & Koedinger, 2009). To address this challenge, we utilized Caselet (Chen & Dubrawski, 2018) - a scalable case-based learning intervention - to provide learners with opportunities to practice DSPS skills using authentic, real-world problems.

While cognition involves acquiring knowledge as a mental action, metacognition is the process of thinking about one's thinking which is a higher order of cognition. Metacognition requires learners to know how to learn and reflect on their learning while solving problems (Winne & Azevedo, 2014). Caselets, as a special case of problem-solving, require learners to monitor and regulate their problem-solving process. Therefore, quantifying metacognition is needed to evaluate the effectiveness of metacognitive awareness in solving problems and learning gain in the DSPS context. Researchers have used the Metacognition Awareness Inventory (MAI) as a self-reported measurement (Schraw & Dennison, 1994) to assess metacognitive processes.

This study primarily focuses on analyzing the cognition and metacognition aspects of DSPS. It aims to (1) elucidate the relationship between cognition, metacognition, and learning gain; (2) inspect the similarity among DS knowledge components uncovered from the Caselet practice performance.

2 LEARNING CONTEXT

We collected the data from a graduate-level data science course in a university setting, with 24 graduate students. We piloted Caselets (i.e., DSPS practices) that were introduced by (Chen, & Dubrawski, 2018). By design, the skills required to solve the Caselets might not strictly align with the course contents. As such, learners were required to do research independently. This extra learning requires learners to use metacognition skills to discover their knowledge gap and find ways to fill it.

3 DATASET DESCRIPTION

The dataset includes aspects of cognitive and metacognitive processes. Metacognition is measured by MAI that has 52 items divided into two main components: 1) Knowledge about Cognition (KC) which measures declarative knowledge, procedural knowledge, and conditional knowledge, and 2) Regulation of Cognition (RC) which measures planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation. The cognitive aspect consists of three assessments: pre-assessment, formative assessments (i.e., Caselets and course assignments), and post-assessment. All the cognitive assessments are Multiple-Choice Questions (MCQs) mapped to DS knowledge components. The frequency and occurrence of those components in the three assessments vary; some occur in all.

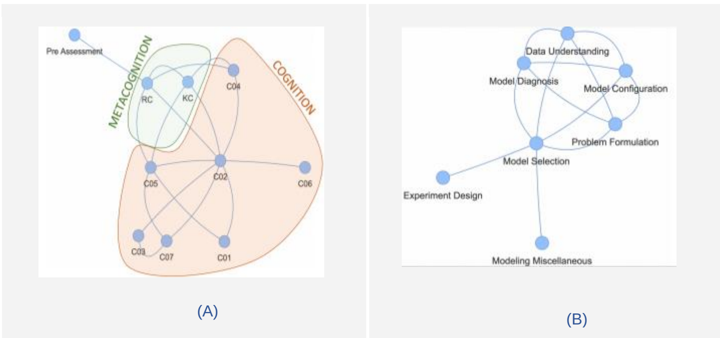


Figure 1: (A) Graph showing pairwise correlation among cognition and metacognition components (nodes) with each edge representing a significant correlation; (B) Graph showing pairwise cosine similarity among Caselets' knowledge components; only edges with significant correlations are retained.

4 EXPLORATORY DATA ANALYSIS

We use pairwise Pearson correlation to explore the correlations between and within learners' performance for MAI, pre-assessment, and Caselets. The result is presented as a sparse graph in Figure 1A, which shows that cognitive processes are highly correlated, and MAI components (KC and RC) are also highly correlated. Though no direct relationship (edge) is observed between the pre-assessment and the Caselets, several edges link MAI and Caselets, suggesting a plausible role of metacognition in explaining the Caselet practice performance. Similarly, we quantify cosine similarity based on the Caselets' performance vector in each knowledge component to capture the structure among them. As shown in Figure 1B, Model Selection node centers on the graph with relationships to all the other components, i.e., learners who answered model selection related questions correctly tend to answer the other questions correctly, and vice versa. This correlation reflects the importance of mastering this component. In contrast, Experiment Design node seems uncorrelated to the other components.

5 PREDICTIVE MODEL

We conduct ANOVA goodness-of-fit tests on a set of nested linear regression models to evaluate the contribution of given factors in predicting learning gains. As shown in Table 1, compared with the baseline model using only the cognitive factors in the pre-assessment, adding metacognition proxied by MAI significantly improves the model performance. However, adding Caselet performance does not show significant improvement, which is worth more investigation. Finally, when assignments are added, the model shows significant improvement.

Table 1: ANOVA test result for model comparison (Model 2 vs. Model 1)

Model 1	Model 2	P- value
Pre-Assessment	Pre-Assessment + MAI	**4.00e-10***
Pre-Assessment + MAI	Pre-Assessment + MAI + Caselets	1.96e-01
Pre-Assessment + MAI + Caselets	Pre-Assessment + MAI + Caselets + Assignment	**3.82e-03***

6 DISCUSSION AND FUTURE WORK

Using data collected from a field study where we piloted Caselets with data science students, we explored cognition and metacognition aspects of the learning and their contribution to predicting learning gain in DSPS. We note that MAI seems to be able to provide predictive value in addition to prior cognitive knowledge. These findings resonate with a prior study (Winne & Azevedo, 2014) on the critical role of metacognition in learning. The graph-based analysis of knowledge components further elucidates the relationship among various knowledge components involved in DSPS, which may be used to guide instructors in allocating limited support resources to focus on critical knowledge components.

As part of ongoing work, we model the learners' knowledge growth in DSPS using a larger dataset. We will also consider metacognition while modeling the learners' knowledge growth. In addition, we have designed a system to capture the learner's digital traces while practicing Caselets. Those traces will help to capture fine-grained cognitive and metacognitive processes. Those lines of research will allow us to gain deep insights into learners' cognition and metacognition toward mastery of DSPS. They may also provide data-driven evidence for designing effective pedagogical interventions to support learners to become competent problem solvers for real-world DS problems.

REFERENCES

Schraw, G., & Dennison, R. S. (1994). *Assessing metacognitive awareness*. Contemporary educational psychology, 19(4), 460-475. https://doi.org/10.1006/ceps.1994.1033

Chen, L., & Dubrawski, A. (2018). Accelerated apprenticeship: teaching data science problem solving skills at scale. In Proceedings of the Fifth Annual ACM Conference on Learning at Scale, 1-4. https://doi.org/10.1145/3231644.3231697

Winne, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 63–87). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.006

Rittle-Johnson, B., & Koedinger, K. (2009). Iterating between lessons on concepts and procedures can improve mathematics knowledge. British Journal of Educational Psychology, 79(3), 483-500. https://doi.org/10.1348/000709908X398106

Computer-Generated formative Feedback using pre-selected Text Snippets

Sylvio Rüdian*[1,2], Clara Schumacher[1], Niels Pinkwart[2]
[1]Humboldt-Universität zu Berlin, [2]German Research Center for Artificial Intelligence
*ruedians@informatik.hu-berlin.de

ABSTRACT: In this paper, an approach is introduced to generate formative textual feedback with the idea to use already existing prepared text snippets that are pre-selected by a supervised machine learning model. The approach is based on existing tools that are extended to be suitable for teachers with low technical skill levels. It uses the trusted learning analytics approach. As state-of-the-art technology still does not provide high-quality results, the teacher is always held in the loop as the domain expert who is supported by a tool, and not replaced.

Keywords: Formative Feedback, Online Learning, Teacher Support, Higher Education

1 INTRODUCTION

Feedback is an essential factor for supporting successful learning processes and outcomes (Hattie & Timperley, 2007). In higher education, summative assessment (and feedback) are predominant, used for grading and comparing learners at the end of a course or unit. Formative assessments focus on providing learners with feedback on their performance and next steps. Hence, feedback can be defined as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007). Elaborated feedback offering information on task-, process- and self-regulation level has been found to be most effective for learning success (Wisniewski, Zierer, & Hattie, 2020). However, providing learners with informative feedback at scale is challenging (Lim, et al., 2020) and a time-consuming task. Thus, it is not surprising that formative feedback is often limited to sample solutions or short paragraphs. This proposed approach aims at supporting teachers in creating feedback with text blocks that are pre-selected automatically, combined, and arranged to create feedback for new learner submissions.

2 RELATED WORK

Despite the examination of computer-generated feedback for decades; the creation of highly informative feedback is very complex, where machines can be supportive, but do not replace teachers (Chen & Cheng, 2008). The variety of texts created by learners is manifold, which is hard to evaluate automatically (Landauer, 2003). Due to the low quality of computer-generated feedback, its application can lead to high frustration levels (Ware, 2011). Even available state-of-the-art automatic writing evaluation tools, like proofreading tools to detect mistakes in submissions of language learners, do not meet teachers' expectations (Rüdian, Dittmeyer, & Pinkwart, 2022). Finally, the teacher remains essential in the learning process to provide feedback. Thus, instead of providing computer-generated feedback to learners directly, a teacher-in-the-loop approach is of high importance. Therefore, the process to create feedback must be intuitive without the need for complex adjustments. In the domain of education, decisions coming from computer-generated feedback tools must be explainable. This is a key component of the trusted learning analytics approach (TLA) (Hansen, Rensing, Herrmann, & Drachsler, 2020). One solution is the tool OnTask (Pardo, et al., 2018), which can principally be used to generate texts based on pre-defined text snippets and rules that use trace data. Based on such rules, decisions can be justified and explained. If for example, the learner submits a text and the tool recognizes that the learner skipped watching a related learning video, which is implemented as a rule, then feedback is given using the snippet with the advice to have a deeper look at the learning material. However, it is essential to educate teachers so that they get an understanding of the versatility of such software. Teachers must have scenarios in mind, which must be implemented in rules. From the practical perspective, this is a pitfall as teachers want to focus on their domain to create learning material and not on scenarios that possibly can exist (Ware, 2011).

3 FRAMEWORK

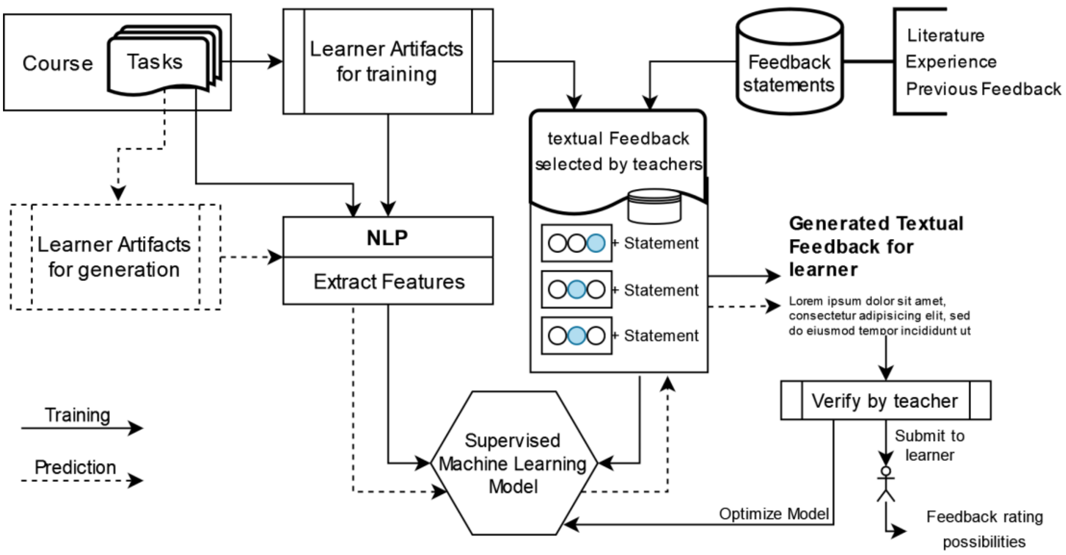


Figure 1: Approach to train a supervised machine learning model and process to generate feedback for new learner submissions.

The approach (Fig. 1) proposes a teacher-in-the-loop approach that is based on pre-defined text snippets to provide feedback on task-level for writing tasks. Such snippets can be extracted from already given feedback texts or from best practices in the literature. Text snippets must meet the condition to be related to a scale (e. g., Likert scale, binary (yes/no) scale). In the training process, teachers create feedback by selecting pre-defined text snippets. The idea of using such snippets is not new, but a helpful step for teachers to reduce the required time to create feedback. Then, snippets are stored including the rating on the scale, e. g. whether a learner correctly applied a concept, or not. NLP features are extracted from user artifacts (e. g. textual submissions). Features can be based on sentiment analysis, word-sense disambiguation, argument mining, or others. Such features are then used to train a supervised machine learning approach, with ratings as labels. Explainable methods such as random forest trees are favored. For all labels that can be predicted with acceptable accuracy, a model is stored. Then, for new learner artifacts of the same task, ratings can be predicted. The teacher gets those predictions so that related text snippets are automatically pre-selected when the teacher aims to create feedback. Based on those selections, a final feedback text is generated. Besides, a reinforcement learning approach is used. The teacher can change pre-selections. Thus, new training data are continuously created to train the model with more data to become more generalizable. Also, the student can evaluate feedback to obtain a critical view of its applicability. The main idea is to separate teachers from the NLP approaches, that run in the background.

4 DISCUSSION

The proposed approach enables more fine-grained feedback and dynamic support compared to a pre-defined rule-based approach only. To cope with limited resources for the provision of formative feedback in higher education, the proposed approach aims at enhancing teachers' practices and reducing their workload when providing highly informative feedback on text artifacts which in turn enables learners to derive appropriate future learning activities. One major concern of using AI in educational contexts is that the algorithms are difficult to comprehend or do not provide actionable outcomes thus potentially resulting in limited acceptance by the stakeholders. This might be supported by the simplicity of the proposed approach that enables teachers creating feedback without the need for abstract technical skills plus by being grounded in the idea of TLA of having the human in the loop of an explainable approach. Furthermore, the model is updated by human feedback, emphasizing humans as active participants in this reciprocal learning process. Hence, in upcoming iterations of the project, stakeholders' perceptions on the tool and the feedback generated will be investigated. Moreover, in experimental studies the impact of the approach on students' learning processes and outcomes will be examined systematically, which is a high demand in the field.

Acknowledgements: This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DHBKI045.

REFERENCES

Chen, E. C.-F., & Cheng, W.-Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. In *Language Learning & Technology 12.2* (pp. 94-112).

Hansen, J., Rensing, C., Herrmann, O., & Drachsler, H. (2020). Verhaltenskodex für Trusted Learning Analytics. Version 1.0. Entwurf für die hessischen Hochschulen. In *Innovationsforum Trusted Learning Analytics.* Frankfurt am Main.

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

Landauer, T. K. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M. D. Burstein, AES*: A cross-disciplinary perspective* (pp. 87–113).

Lim, L.-A., Dawson, S., Gašević, D., Joksimović, S., Fudge, A., Pardo, A., & Gentili, S. (2020). Students' sense-making of personalized feedback based on learning analytics. In *Australasian Journal of Educational Technology 36(6)* (pp. 15-33).

Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., & Vigentini, L. (2018). OnTask: Delivering data-informed, personalized learning support actions. In J. o. 5(3).

Rüdian, S., Dittmeyer, M., & Pinkwart, N. (2022). Challenges of using auto-correction tools for language learning. In *LAK22* (pp. 426-431).

Ware, P. (2011). Computer-generated feedback on student writing. In *Tesol Quarterly 45.4* (pp. 769-774).

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. In *Frontiers in Psychology 10* (p. Article: 3087).

Using learning analytics to uncover emotions in collaborative learning contexts

Rogers Kaliisa[1], Jennifer Scianna[2], Jan Arild Dolonen[1], Rachelle Esterhazy[1], Crina, Damsa[1]
[1]University of Oslo, Department of Education, Oslo, Norway
[2]University of Wisconsin, Department of Curriculum & Instruction, Madison, USA
rogers.kaliisa@iped.uio.no; jscianna@wisc.edu; j.a.dolonen@uv.uio.no; rachelle.esterhazy@iped.uio.no;crina.damsa@iped.uio.no

ABSTRACT: Educational researchers have pointed to socioemotional dimensions of learning as crucial in gaining a more nuanced description of student engagement and learning. However, to date, only a handful of learning analytics researchers have focused on analysing emotions in collaborative learning environments. In this poster paper, we present our ongoing work attempting to identify sentiment in asynchronous online discussions. We also discuss future efforts to analyse emotions based on multimodal datasets and the plan to design a multimodal learning analytics system to capture students' emotions during collaborative learning activities.

Keywords: Learning analytics; Sentiment analysis, Emotions; asynchronous online discussions; teacher-facing dashboard

1. INTRODUCTION

There is increasing attention to uncovering the many facets of sentiment and emotion from online discourse to gain a more nuanced description of student engagement and learning, as well as an interpretation of the complex dynamics generated during collaborative learning activities [1]. Despite the considerable recognition of emotions as critical to understanding students' learning processes, empirical research on collaborative tasks such as online discussions has focused on social interactions by looking at patterns of relational connections between the different participants, leaving the dimension of emotions relatively untouched. To date, very few academic content-related sentiment studies have been reported in the literature.

The origins of emotion-centred work in education resulted from serendipitous findings, which related emotion as a component of studying motivation and other student beliefs and understandings [2]. This early work called for new theoretical frameworks, which would acknowledge the "person-in-context" [2]. Since then, research around emotion in educational settings has often relied on self-reported scales or think-aloud protocols [1]. While these approaches are helpful in a targeted research context, they can become extraneous if the emotion is not the direct target of inquiry. One approach to studying emotion in learning settings is through the natural language processing method, Sentiment Analysis (SA), which has predominantly been used for analysing learner affect on course structures, both as an indicator of attrition rate and student performance [3]. In this paper, we present ongoing work attempting to use sentiment analysis to analyse the sentiment polarity (positive, negative and neutral) of students' contributions within asynchronous online discussions. We also discuss future efforts to design a multimodal learning analytics dashboard that uses multiple modalities to capture students' emotions during collaborative activities.

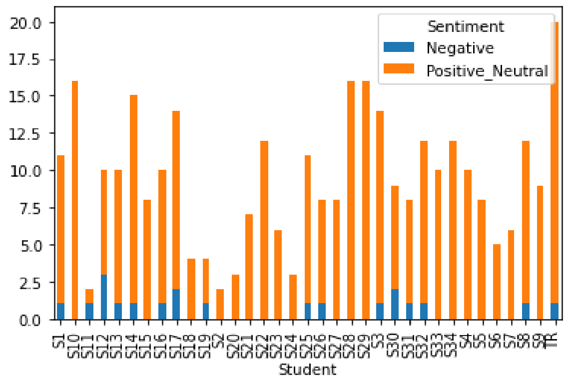
2. DATA, PARTICIPANTS, COURSE CONTEXT, AND ANALYSIS

The data used in this study was extracted from the discussion forum contributions posted on Canvas, a learning management system, within a blended bachelor's course at a public university in [blinded]. This course is a part of the university's bachelor's in pedagogy. The course had 34 students and four teachers. There were six learning modules in the course, and weekly online discussions were created based on each module. The seven discussions form the basis of the data used in this study. To prepare the data for sentiment analysis, the textual data was translated into a qualitative data table where each row contained one post. Each post was labelled as responding to either the primary forum prompt or another participant's post. Additional metadata included the week the post was made, the participant who posted it, and the position of the post in a sequence of all posts for a given week.

To perform sentiment analysis, the sentiment of each post was analysed using the Valence Aware Dictionary for Sentimental Reasoning (VADER) model [4] that is present in the Natural Language Toolkit library. VADER captures both the polarity and magnitude of overall sentiment within a text. Posts were considered holistically as documents in the model, and each word contributed to the overall valence. The number of positive, neutral, and negative words are then normalised to yield independent scores, which are aggregated to create the compound sentiment score. Using this model, an overall sentiment score was assigned to each post using the compound variable with a threshold of -.05 drawing the bound between negative and positive_neutral. This binary classification scheme is supported by prior work, demonstrating that such a system more consistently identifies negative sentiment in education-related texts [5].

3. PRELIMINARY RESULTS

The VADER analysis identified 311 of the posts as either positive or neutral and 20 posts as negative. Sixteen students in the course discussion contributed in ways that were identified as negative. Of these, four also received negative responses from other students. Six students received negative responses while not contributing themselves negatively. The main post of weeks two and seven received direct negative feedback. Furthermore, some weeks had higher occurrences of Negative posts than others did. Weeks 2 and 7 were the highest among all weeks, while Week 5 received the least number of Negative posts (See Figure 2). Moreover, a qualitative analysis of the negative posts showed that negative sentiment is connected with students connecting course content to *Personal* experiences or being critical of some topics. For example, Week 2's discussion seemingly centred on *Personal* connections to the course aligning with the prompt itself, which asked students to discuss their experiences with Educational Technology. This topic is highly relatable for the students and offers an opportunity to share everyday experiences. Negative posts in this week shared two themes: waste and teacher education (or lack thereof). Furthermore, some weeks had higher occurrences of Negative posts than others did. Weeks 2 and 7 were the highest among all weeks, while Week 5 received the least number of Negative posts (See Figure 2).



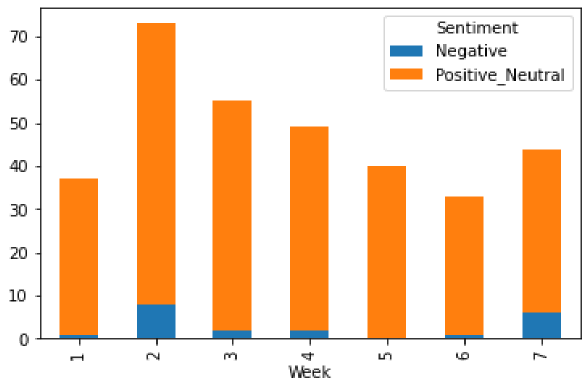


Fig. 1. Posts segmented by sentiment per participant. **Fig. 2.** Sentiment scores by the week of discussion.

4. CONCLUSION AND NEXT STEPS

We applied the VADER algorithm to explore how posts from academic discussion forums would be sorted based on their sentiment. We found that VADER could identify posts and indicate patterns of relationships to weeks and between participants. The preliminary insights provide evidence that sentiment analysis can provide valuable insights into students' attitudes during collaborative learning tasks. Such information could be leveraged to provide a more nuanced and detailed understanding of how learners respond and engage with academic content or collaborative tasks. The teachers can use this information to design or adapt the collaborative tasks based on the identified sentiment (e.g., areas of concern or conflict) to foster a more positive, productive, and engaging learning environment. To support the use of sentiment analysis for making timely learning design decisions, we have developed an automated teacher-facing dashboard (CADA) that captures the sentiment polarity of students' discussion posts in real-time (6). We plan to extend the current work by studying specific emotions (e.g., anger, frustration, joy, boredom, amusement, anxiety) expressed during collaborative learning tasks using multimodal datasets such as video recordings, student reflections, diaries, and interviews. Our future work aims to develop a multimodal learning analytics feedback system that automatically captures students' emotions during collaboration. Similar to (6), the multimodal feedback system will be pluggable into the Teams platform to offer students and teachers timely formative feedback about their emotional and epistemic states during collaborative tasks.

REFERENCES

[1] Joksimović, S., Dawson, S., Barthakur, A., Poquet, O., Wang, Y. E., Marmolejo-Ramos, F., & Siemens, G. (2022). Mapping the Landscape of Social and Emotional Learning Analytics. In *Social and Emotional Learning and Complex Skills Assessment* (pp. 27-47). Springer, Cham.
[2] Meyer, D. K., & Turner, J. C. (2002). Discovering Emotion in Classroom Motivation Research. *Educational Psychologist*, *37*(2), 107–114. https://doi.org/10.1207/S15326985EP3702_5
[3] Zheng, L., & Huang, R. (2016). The effects of sentiments and co-regulation on group performance in computer-supported collaborative learning. *The Internet and Higher Education*, *28*, 59-67.
[4] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).
[5] Borchers, C., Rosenberg, J. M., & Gibbons, B. (2021). To Scale or Not to Scale: Comparing Popular Sentiment Analysis Dictionaries on Educational Twitter Data. *Proceedings of the 14th International Conference on Educational Data Mining*, 619–624. https://educationaldatamining.org/EDM2021/EDM2021Proceedings.pdf

[6] Kaliisa, R., & Dolonen, J. A. (2022). CADA: a teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. *Technology, Knowledge and Learning*, 1-22.

Identification of Reflection Prompts that Provide Insight to Student Motivation to Persist in Engineering

Lauren Singelmann, Yuezhou Wang, and Elizabeth Pluskwik
Minnesota State University, Mankato
lauren.singelmann@mnsu.edu; yuezhou.wang@mnsu.edu; elizabeth.pluskwik@mnsu.edu

ABSTRACT: This poster presents preliminary results in identifying reflection prompts that are most likely to provide meaningful insight to undergraduate student motivation to persist in engineering. A classification algorithm was used to determine student motivation levels for a variety of prompts, and these classifications were then compared to observed motivation to persist in engineering over time. Five prompts that had the greatest alignment and five prompts that had the least alignment are presented as example cases. By identifying students in need, this work can provide them scalable and timely support.

Keywords: Written reflections, motivation, engineering

1 INTRODUCTION

Written reflection tasks provide value for both students and instructors alike; students gain value from taking part in metacognitive activities, while instructors gain value from the insight into what their students are thinking and feeling. However, it is challenging to scale up attentive and timely review of these reflections. Therefore, learning analytics work is being done to leverage the speedy pattern-finding of a computer and the context of an instructor when reviewing student written reflections.

Previous work has shown that text classification algorithms were able to differentiate between negative and positive motivation in written reflections with sufficient agreement to a human reader (Pluskwik, 2022). However, questions remain as to how the motivation level demonstrated within a specific prompt aligns with motivation to persist within a program over time. For example, some prompts may elicit student responses that all demonstrate positive motivation but do not necessarily predict persistence in engineering over time. This work intends to overcome the limitation by exploring which types of prompts are most likely to differentiate between high and low motivation over time.

2 METHODS

This study looked at written reflections from 45 undergraduate engineering students from the same engineering program. During their third and fourth years, students were assigned weekly reflection prompts. The 81 reflection prompts were grouped into 6 categories: General, Technical Competencies, Design, Professionalism, Future, and Teamwork.

Each student's response to each prompt was classified using an algorithm that assigns a probability of showing positive motivation versus negative motivation. This algorithm has been tested previously on its ability to classify student reflections by motivation level with sufficient agreement to a human reviewer (Pluskwik, 2022). This classification was then compared to multiple faculty members' observation of that student's motivation to persist in engineering over time. The faculty made these classifications by considering their interactions with the students, and they used the motivation continuum from self-determination theory (Deci, 2012) as a framework.

Each prompt was then assessed using a variety of metrics including accuracy, precision, recall, and ROC AUC score (the area under the receiver operating characteristic). This paper reports on the ROC AUC score because it quantifies how well the model distinguishes between the two classes (negative motivation versus positive motivation) by using the probability scores from the classifier for each reflection. The ROC plots the false positive rate versus the true positive rate at each decision threshold. A perfect classifier's ROC would have an area under the curve of 1, and a random classifier's ROC would have an area under the curve of 0.5.

3 RESULTS

Results are shown in Table 1 and Figure 1. Table 1 shows the five prompts that elicited responses that were most aligned with overall motivation to persist in engineering, as well as the five prompts that elicited responses that were least aligned with overall motivation to persist in engineering. Figure 1 shows the ROC AUC curves for these ten prompts.

Table 1: Example questions that most strongly aligned with long-term motivation to persist in engineering (Q1-Q5) and example questions that least strongly aligned with long-term motivation to persist in engineering (Q6-10)

Code	Question	Category	ROC AUC
Q1	Describe the professional goal [you made] and the progress you have made.	Professionalism	0.764
Q2	How did your individual contributions impact project completion?	Teamwork	0.689
Q3	What steps can you take to achieve that improvement [with your team]?	Teamwork	0.685
Q4	Describe the take-away knowledge [from your design project].	Design	0.671
Q5	Which of the actions of an engineer do you feel most qualified in doing? Why?	Professionalism	0.663
Q6	How might you learn from this experience to help others in your engineering career?	General	0.401
Q7	Which of these skills [listed] do you think are your strengths?	Teamwork	0.368
Q8	Engineers have an obligation to contribute to their communities. How do you plan to contribute to your communities after graduation?	Future	0.363
Q9	How do you see mindfulness as being of value to you in the context of your engineering career?	Future	0.333
Q10	What are the 3 most important aspects of interpersonal communication?	Professionalism	0.167

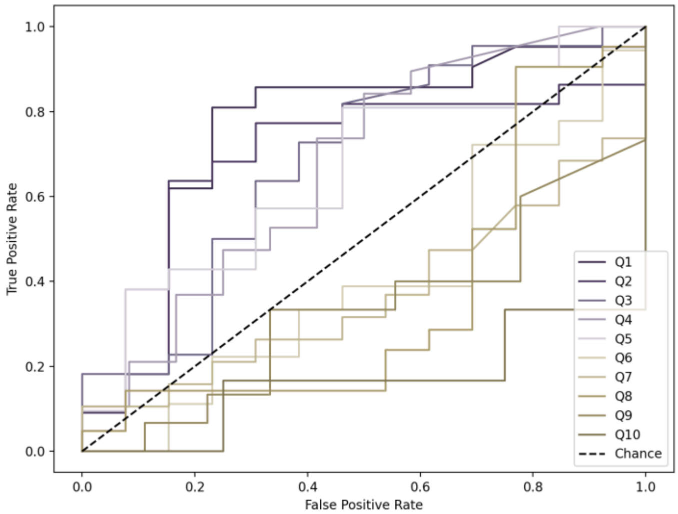


Figure 1: An ROC AUC graph showing example curves for Q1-10. Q1 has the greatest area under the curve, and therefore the greatest alignment with observed motivation over time. Q1-5 (purple) have better than chance alignment, meaning that the classifications of the students' responses to those questions were strongly aligned with their observed motivation to persist in engineering over time. Q6-10 (gold) have worse than chance alignment, meaning that the classifications of the students' response to those questions were poorly aligned with their observed motivation to persist in engineering over time.

4 CONCLUSION

This work indicates that some reflection prompts are better than others at giving insight into long-term motivation. Reading these student reflections takes time, and it can be challenging for instructors to identify trends and patterns in responses across semesters. Therefore, this approach provides a solution for instructors by analyzing student responses in real time and determining if the student is showing signs of negative motivation to persist in engineering. It should be noted that there are many factors that play into whether a student persists in engineering. Therefore, the goal of future work is to not be able to perfectly predict student outcomes, but rather to empower instructors to provide timely support to students who need it most.

REFERENCES

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd.

Pluskwik, E., Wang, Y., & Singelmann, L. (2022, August). Work-in-Progress: Understanding learners' motivation through machine learning analysis on reflection writing. In *2022 ASEE Annual Conference & Exposition*.

Contextually appropriate learning analytics; a case of activity engagement differences in geo-cultural contexts

Saman Rizvi
University of Cambridge, UK
sr2004@cam.ac.uk

Bart Rienties
The Open University, UK
bart.rienties@open.ac.uk

Jekaterina Rogaten
University of the Arts, London, UK
j.rogaten@fashion.arts.ac.uk

René F. Kizilcec
Cornell University, USA
kizilcec@cornell.edu

ABSTRACT: Open online learning spaces, such as MOOCs, have the potential to provide affordable and widely accessible education. However, research has shown that persistence and completion rates among MOOC learners are concentrated in specific, developed, predominantly English-speaking regions. The way these courses are designed using different proportions of learning activity types (learning design) potentially influences learners' persistence. A large and growing body of learning analytics research has been conducted on the analysis of voluminous behavioral data produced and stored in course logs. However, limited work has considered the variability in the contexts in which the behaviors occurred. This poster shares selected results from two linked learning analytics studies. The first study explored the association between learning designs and learners' persistence. The second study examined variations in self-reported preferences for learning designs and allowed to understand the observed differences in the first study. Both studies take into account variations in learners' geo-cultural subgroups. The behavioral patterns mirror existing inequalities which suggests that the design of courses is unable to improve upon the status quo. A careful consideration of possible heterogeneity in data evidently makes learning analytic outcomes more pervasive and contextually appropriate.

Keywords: Learning analytics, data-informed decision-making, human-centered design, equity

1 INTRODUCTION

Massive Open Online Courses (MOOCs) are often recognized for their scalability and global reach. However, regardless of their potential to facilitate affordable and widely accessible learning, there is a disparity across the globe in terms of enrolment and performance (Kizilcec et al., 2017). The way courses are designed may have a critical predictive role in keeping learners engaged. Contextual characteristics such as regional, cultural, or socioeconomics may also help us understand learners' experiences, and thereof, their persistence in a course. A MOOC learning design comprises different proportions of activities, for example, instructional videos, reading material or articles, discussion-based activities, and assessments. Prior work points to potentially distinct learning activity type priorities across distinct cultures. A large and growing body of learning analytics research remains focused on various approaches to improve the *overall* learner experience and *overall* engagement, and limited work (e.g., Sha, et al., (2022)) has explored outcome variations, for example, across disciplines or various contextual indicators.

Consideration of the cultural heterogeneity in the learners' population enables research to advance beyond a one-size-fits-all approach to analytical solutions for various subgroups. An aggregated analysis tends to favor larger, privileged groups over smaller, less privileged groups. Most learning analytics methods inspect behavioral data as a large and heterogenous collection of information to inform decisions. Such methods err more in one direction and potentially favor the majority group while leaving behind under-representative groups. We argue that any strategic plans for improvements resulting from, or guided by, overall data analysis may in fact be biased. This poster shares evidence from two studies underlining the promise of subgroup analyses in learning analytics.

1.1 Study 1: Predictive link between course learning activities and persistence

This study leveraged the log data of 49,582 learners from ten MOOCs developed by the Open University (OU) and offered via the FutureLearn platform. The courses were diverse and each week contained steps with several types of activities: articles, discussions, videos, and quizzes. Using the GLOBE framework, IP-based locations were categorized into ten geo-cultural clusters: e.g., Sub-Saharan Africa (AF), Confucian Asia (CA), Eastern Europe (EE), Germanic Europe(GE).

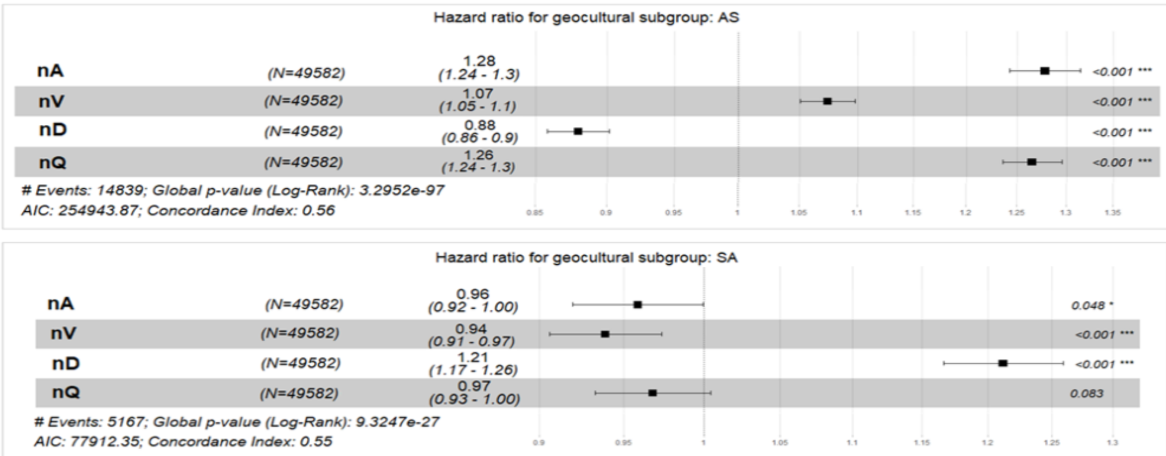


Figure 1: Contrasting behavior of learners from Anglo-Saxon (top) and South Asian (bottom) subgroups. (Note: nA, nV, nD, nQ = number of Articles, videos, discussions, quizzes)

Survival analysis was used to quantify the predictive association between the number of learning activities (e.g., videos, articles, quizzes, and discussions) and the hazard to persistence. The percentage of activities learners accessed was used to operationalize persistence. Following that, group-level analysis was performed to see if the degree of association was different for different geo-cultural subgroups. To reduce the risk of finding spurious relationships between several significant two-way interactions between the learning design constructs and geo-cultural groups, penalized Cox Regression (LASSO_Cox) was used. As illustrated in Figure 1, the group-level analysis revealed that a change in the types of learning activities is associated with opposite hazard profiles for the two largest geo-cultural contexts in the dataset: Anglo-Saxon (AS) and South Asia (SA) (source: Rizvi et al., 2021).

1.2 Study 2: Learners' perceptions of various learning design elements

The study discussed above analyzed activity engagement behavior data stored in the course logs. What remained unclear was how learners perceived the role of different learning design activities in retaining their interest in the course. Therefore, a mixed-method study was conducted that allowed to examine the reasons behind observed differences using semi-structured interviews with

FutureLearn MOOC learners (n = 22 from seven geo-cultural regions). The interview transcripts were analyzed using sentiment analysis to assess participants' perceptions of the various elements of learning designs. The R package *sentimentr* was used to approximate the sentiment polarities.

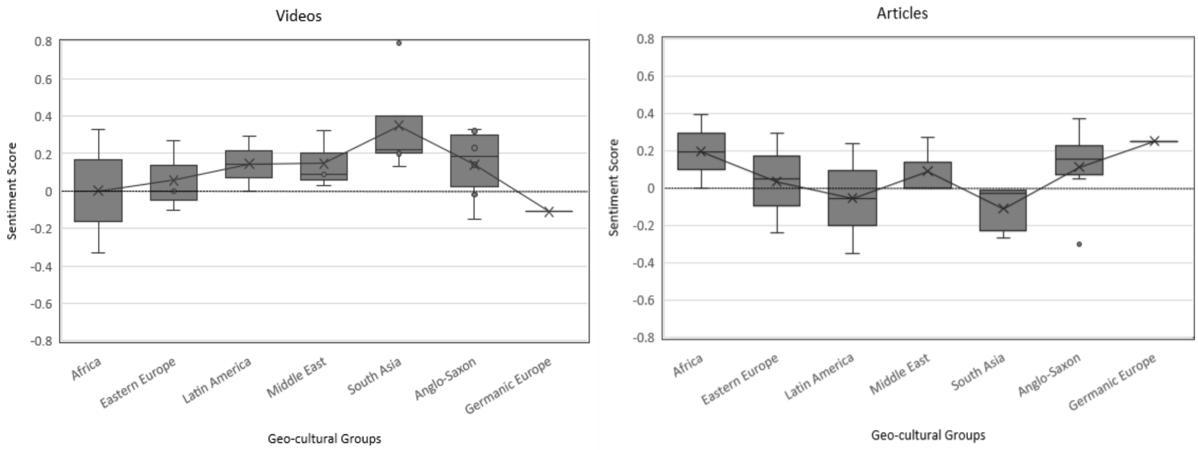


Figure 2: Variations identified in sentiment scores for the seven participating geo-cultural groups.

The nature and direction of sentiments and self-reported activity preferences varied across the geo-cultural context. Figure 2 shows an elevated positive score for instructional videos in South Asian learners but less positive sentiment in learners residing in the Anglo-Saxon region (source: Rizvi et al., 2022). The sentiment score variations would have been conflated in a single measurement for each activity type if they were not compared across the theoretically meaningful, geo-cultural contexts.

2 CONCLUSION

While learning design variations may be one factor, other factors have also contributed to global gaps in MOOCs, such as language proficiency, insufficient technical resources, and digital literacy levels in different contexts. Further research is needed to see what the learning analytics field can accomplish with similar methods while making interpretations and inferences from the behavioral data that was generated within the context in which the behavior occurred. We believe that context awareness can help make learning analytics more meaningful, transparent, and contextually appropriate.

ACKNOWLEDGMENT

This work was supported and funded by the Leverhulme Trust, Open World Learning, and UKRI ESRC.

REFERENCES

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, *355*(6322), 251-252.

Sha, L., Raković, M., Das, A., Gašević, D., & Chen, G. (2022). Leveraging class balancing techniques to alleviate algorithmic bias for predictive tasks in education. *IEEE Transactions on Learning Technologies*, *15*(4), 481-492.

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2021). Investigating variation in learning processes in a FutureLearn MOOC. *Journal of computing in higher education*, *32*(1), 162-181.

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. (2022). Culturally adaptive learning design. *Open World Learning (2022)*, *103*.

The state of teaching about algorithmic bias and fairness in Learning Analytics programs

JEANNE MCCLURE*, North Carolina State University, USA
DOREEN MUSHI, North Carolina State University, USA
SHIYAN JIANG, North Carolina State University, USA
SHAUN KELLOGG, North Carolina State University, USA

Algorithms play a critical role in education decisions, and incorrect predictions or inferences lead to discrimination towards specific groups. For this study, we look at Learning Analytic Programs to understand their teaching methods, materials used, and other content that can aid in students' understanding of algorithmic bias/fairness in relation to Learning Analytics. [Method] We conducted an online exploratory analysis, evaluating course descriptions and syllabi. Further, using a loose content analysis to code. [Results] Materials used in the course for algorithmic bias/fairness include lectures, slides, blogs, articles, and book chapters, with the core materials used to define the topic being journal articles. Most instructors assess their students through rich group discussions or reports. [Discussion] Based on such results, we provide and discuss initial findings and offer recommendations to improve the teaching of algorithmic bias and fairness in LA Programs.

Additional Key Words and Phrases: Learning Analytics, Algorithmic Bias, Algorithmic Fairness

1 INTRODUCTION

In education, we see fewer cases of algorithmic bias compared to healthcare, and criminal justice [3]. Across the literature, algorithmic bias and fairness have varying definitions, such as a socio-technical phenomenon [6], unfair over-biased [4], the absence of bias and discrimination in a system [9] and [7] description that bias is an "inequitable prediction across identity groups" (p. 228). Collective harms of Algorithmic bias/fairness include loss of opportunity, economic loss, and social stigmatization[11]. Today, we find Artificial Intelligence (AI) systems in various fields, from reading medical reports to manuscript writing [5, 10]. The literature points to an increase in the adoption of algorithmic systems developed to reduce the time for predictions in vast amounts of educational data [9]. Machine learning (ML) and deep learning systems, subsets of AI, are promising to "increase learning" through predictive analytics, Intelligent Tutoring Systems, and Dashboards [2]. [3] show that in education, the algorithmic bias appears in three categories: race/ethnicity, nationality (comparing learners' current national locations), and gender. Additionally, studies have found bias in models predicting graduation for indigenous learners [1] and bias against female learners [1, 8].

Although there is robust literature on the Ethics of Learning Analytics and Privacy, less is known about algorithmic bias, especially in LA classroom settings. However, it's critical to prepare learning analytics researchers and practitioners to understand this important topic. This paper explores the state of teaching this important topic and discusses an agenda to improve teaching in this area. This study specifically examines: "RQ1: How do Learning Analytics Programs teach Algorithmic bias/Fairness in Learning Analytics Programs (topic or subtopic)?"

In LAK '23: The 13th International Learning Analytics and Knowledge Conference, March 13–17, 2023, Arlington, TX.,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

"RQ2: What teaching-learning materials (TLM) aid in students' understanding of algorithmic bias/fairness in relation to Learning Analytics?"

"RQ3: What activities do students engage in to connect algorithmic bias/fairness to learning analytics concepts?"

2 METHOD

In this research, we specifically examined colleges and universities which offer a Learning Analytics Degree or Certificate. We did not include Learning Analytic MOOCs or single courses within University/College for this initial examination. Through an exploratory analysis, we used a Google search key words like "Learning Analytics" & "Graduate Certificate" or "Masters" or "PhD" or "Doctorate." We found sixteen LA Programs (15 US & 1 Australia). We examined the degree websites for individual course descriptions and the course syllabus (if available). If no information was located on the program website, we emailed a program coordinator requesting information using the research questions as a guide.

Each Course Description and Syllabus were scanned for Algorithmic bias or Algorithmic fairness using semi-loose content analysis. Once identified, we further analyzed whether it is taught as an individual topic or sub-topic. (We defined a topic as a week or longer with Algorithmic bias/fairness as the full topic, and a sub-topic means Algorithmic bias/fairness is mentioned within a broader topic such as predictive modeling). Once the descriptions were coded, the researchers looked to understand the teaching methods, materials, and other content used to aid in students' interaction with algorithmic bias/Fairness in relation to Learning Analytics.

3 FINDINGS AND CONCLUSION

Within the 16 LA programs, we reached out to 14 schools with a 64% response rate.

RQ1: How do Learning Analytics Programs teach Algorithmic bias/Fairness in Learning Analytics Programs (topic or subtopic)?

For this study, out of the 64% of LAK programs that responded, only 40% teach a topic or sub-topic on Algorithmic bias/fairness. We found that out of 40% of those LA programs, 84% teach Algorithmic bias/Fairness as an individual topic, and two introduce Algorithmic bias/fairness in combination with another lesson. Like NYU, which mentions Algorithmic bias/fairness while teaching predictive models.

RQ2: What types of teaching-learning materials (TLM) aid in students' understanding of algorithmic bias/fairness in relation to Learning Analytics?

We found that LA Programs use various materials to teach Algorithmic bias/Fairness. They use lectures, slides, blogs, articles, and book chapters. For instance, several LA Programs used the Baker and Hawn article "Algorithmic bias in Education" as a primary source on Algorithmic bias/fairness. For example, at NCSU, in their Intro to Learning Analytics course, students read a chapter from the book "Learning Analytics Goes To School" as part of a broader topic that includes workflow, ethics, and predictive analytics. At Brandeis University, students read a blog called "5 Algorithms that Demonstrate Artificial Intelligence Bias" and another blog by Kumar Chandrakant called "Biases in Machine Learning," which looks at types of algorithmic biases in Machine learning.

RQ3: What types of activities do students engage in to connect algorithmic bias/fairness to learning analytics concepts?

Our results show that the LA programs that teach Algorithmic bias and fairness include activities that engage students in rich discussions through online discussion groups and projects. Another course at NCSU, Machine Learning, has students read the Baker and Hawn article and then post a discussion connecting Algorithmic bias to machine learning and responding to classmates. In another example, at the University of Pennsylvania, in the Big Data, Education, and Society course, students engage in a sub-project where they conduct an assessment of the primary risks and challenges for their larger project. Connecting information learned from an earlier lecture to their real-world problem, thinking deeply about challenges or risks they may encounter in the machine learning lifecycle. Then they write a product explaining the steps needed to prevent the risks.

The initial results of this exploratory analysis indicate that LA programs recognize the importance of engaging students in understanding Algorithmic bias/fairness. However, there are inconsistent depths of how the subject is taught, if at all. We encourage Learning Analytic programs to proactively address factors contributing to algorithmic bias beyond theoretical assumptions. We suggest an initial agenda to improve the teaching of algorithmic bias and fairness in LA Programs.

First, developing a conceptual model to define teaching and learning of Algorithmic bias/fairness concepts, learning objects, and learning outcomes in the Learning Analytics courses. Working with industry leaders to understand what tools are used to mitigate and reduce algorithmic bias so that their students are industry-ready when graduating.

Second, improving the sharing of Algorithmic bias/fairness materials with open registered reports, open repositories, and open and FAIR data practices. Adoption of open source platforms such as GitHub to share materials and exchange towards benefiting both instructors and students.

Third, create a network of organizations and universities to foster sharing, open reviews, and idea exchange to make a community-level effort to prepare students to detect and mitigate algorithmic bias/fairness.

In conclusion, this paper offers emerging results of the current state of teaching about algorithmic bias and fairness in Learning Analytics Programs. From these results, we envisioned an initial agenda to improve the teaching of algorithmic bias and fairness in LA Programs. However, in the future, a deeper survey analysis that includes all LA programs would establish a richer agenda, thus contributing to the evolution of the teaching of algorithmic bias and fairness in LA Programs.

REFERENCES

[1] H. Anderson, A. Boodhwani, and R. S. Baker. Assessing the fairness of graduation predictions. In *EDM*, 2019.

[2] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai. Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, 20(2):13–29, 2016.

[3] R. S. Baker. Baker: From tracking to classroom instruction, ed tech based on biased data can make inequities worse. *America's Education News Source*, n.v.(n.n).

[4] R. S. Baker and A. Hawn. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4):1052–1092, 2022.

[5] T. Davenport and R. Kalakota. The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2):94, 2019.

[6] M. Favaretto, E. De Clercq, and B. S. Elger. Big data and discrimination: perils, promises and solutions. a systematic review. *Journal of Big Data*, 6(1):1–27, 2019.

[7] J. Gardner, C. Brooks, and R. Baker. Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 225–234, 2019.

[8] N. Gaskins. Interrogating algorithmic bias: From speculative fiction to liberatory design. *TechTrends*, pages 1–9, 2022.

[9] R. F. Kizilcec and H. Lee. Algorithmic fairness in education. In *The Ethics of Artificial Intelligence in Education*, pages 174–202. Routledge, 2022.

[10] E. Nakazawa, M. Udagawa, and A. Akabayashi. Does the use of ai to create academic research papers undermine researcher originality? *AI*, 3(3):702–706, 2022.

[11] C. Tatar, D. Culbreth, S. Jiang, C. Rose, J. Chao, R. Ellis, S. Jiang, and K. Wiedemann. High school students' sense-making of artificial intelligence and machine learning. In *Proceedings of the 22nd Koli Calling International Conference on Computing Education Research*, pages 1–2, 2022.

Tracking Student Interaction with Automated Feedback for Programming Assignments in Large-scale Computer Science Courses

Nils Rys-Recker
Western Washington University
rysrecn@wwu.edu

Qiang Hao
Western Washington University
qiangh@wwu.edu

Lu Ding
University of South Alabama
luding@southalabama.edu

Michael Tsikerdekis
Western Washington University
tsikerm@wwu.edu

ABSTRACT: This poster paper discusses tracking and analyzing student interaction with automated feedback for programming assignments on a large-scale. We described a replicable approach to provide student automated feedback for programming assignments using technology stacks that are free and accessible to everyone, as well as a preliminary study that predicts student struggles based on the described approach. This study concludes by discussing the potential of the replicable approach to implement automated feedback systems for learning analytics.

Keywords: continuous integration, automated feedback, programming assignments

1 INTRODUCTION

Automated feedback for programming assignments has been studied extensively in computing education from different angles. These include the tools that automate the process of providing feedback, how students interact with automated feedback, as well as how to design effective automated formative feedback (Chow et al., 2017; Ihantola et al., 2010). Despite the advances from different perspectives, there is still a lack of a simple and replicable approach to deploy automated feedback for programming assignments in large-scale computer science courses. The lack of replicable approaches to deploy automated feedback makes collecting data on how students interact with automated feedback challenging, which in turn reduces the possibility of replication research on this topic and limits further applications of automated feedback.

2 METHODOLOGY

We proposed and studied a scalable approach that implements automated feedback from programming assignments using widely adopted technologies, and collecting fine-grained data of student interactions with automated feedback. To achieve this goal, we combined two existing technologies that are widely adopted by programming professionals for other purposes to implement our automated feedback. The two technologies include version control systems and continuous integration tools.

Continuous integration is a development practice that emphasizes merging new code into a well-tested code base (Ståhl & Bosch, 2013). When the merging happens, all the code needs to undergo preconfigured automated tests. When continuous integration is adopted in learning and teaching of programming, students are able to continuously update their code, see the updated code being tested, receive the feedback of the test and make further improvements accordingly. A **version control system** is a specialized type of database used by developers to store different versions of the source code (Casquina & Montecchi, 2021). When version controls and continuous integration are used together in the context of computing education, students can get unlimited formative feedback on their code quality whenever they store a version of their source code. In combination, version control systems and continuous integration can generate fine-grained data that inform learning analytics researchers of how students interact with automated feedback, and make progress towards problem solving.

In this study, we implemented automated feedback using Git as the version control system and GitHub Action as the continuous integration tool. The bridge that links Git and GitHub Action is GitHub, which is a cloud system that can host Git repositories. Git, GitHub, and GitHub Action work in concert to provide students automated feedback, and the process is shown in Figure 1. Both GitHub and GitHub Action have a set of publically accessible APIs that can be used for data collection.

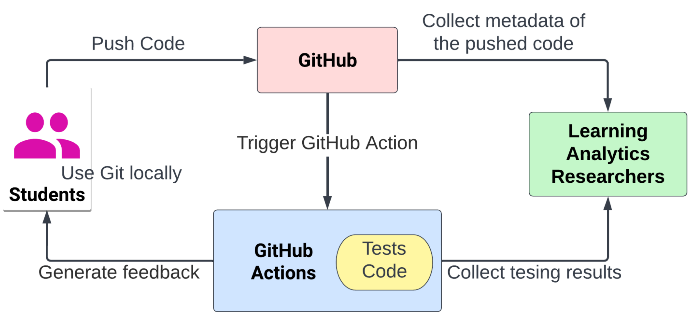


Figure 1. The architecture for data collection utilizing GitHub and GitHub Actions

To test the capability of the system for learning analytics research, we collected and analyzed the data to explore how early and accurately we can predict student struggles in programming assignments. We collected data on one programming assignment in CS2 of 167 students. A student maintained a separate GitHub repository. While working on the assignment, students periodically pushed their progress to GitHub. Automated testing was performed on each push, and feedback was sent to students through emails, and can also be viewed through the web interface. Whether a student struggles on the assignment is the target of our analysis. For the ground truth labels, we chose to define "struggling" as that by the time the assignment was due, a student failed at least one test case. We collected data on each push and computed the following features accordingly:

- Timestamp: The timestamp of a push of code.
- Lines of changed code: The number of lines added and removed.
- Test ratio: The ratio between the passed test cases and the total test cases.
- Error ration: The ratio between compilation error numbers and push numbers.
- First commit timestamp: The timestamp of the first commit.

We experimented with a single node recurrent neural network (RNN) and long short term memory network (LSTM), a wide RNN and LSTM each with one layer of three nodes, and a deep RNN and

LSTM with multiple layers of multiple nodes. To serve as a point of comparison, we constructed a baseline model that does not require pre-training. Given a sequence of pushes belonging to a student, the baseline model finds the line of best fit to approximate the test ratio as a function of the normalized timestamp. This approximated function is used to predict at what time the student will pass all test cases.

3 RESULTS AND DISCUSSION

We used the area under the ROC (AUC) and how early a model successfully predicts student struggles as the metric to assess predictive power. To better evaluate our methodology and mitigate potential effects of heterogeneity in our dataset, we utilized 10-fold stratified cross validation. Beyond ROC, Our preliminary results are summarized in the following Table 1.

Table 1: The earliest a model can predict student struggles in a 14-day assignment.

Model type	Mean AUC >= 0.7	Mean AUC >= 0.8
RNN	6.81 days	13.97 day
LSTM	4.27 days	7.88 days
Wide RNN	3.99 days	4.51 days
Wide LSTM	3.65 days	3.78 days
Baseline model	6.57 days	-

Our findings indicate that it is possible to build efficient machine learning models that identify student struggles on the data collected in just a few days. Additionally, our findings demonstrated the potential of the automated feedback system implemented using the combination of a continuous integration tool and a version control system. GitHub and GitHub Action, in particular, provide detailed documentation on how to use their APIs to collect data on student interaction with the automated feedback at a fine-grained level. Such a replicable approach to implement automated feedback systems and collect data has significant value for learning analytics researchers, such as making it possible to conduct replication studies on this topic, and building automated systems that identify and reach out to struggling students.

REFERENCES

Chow, S., Yacef, K., Koprinska, I., & Curran, J. (2017). Automated data-driven hints for computer programming students. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, ACM, 5–10.

Ihantola, P., Ahoniemi, T., Karavirta, V., & Seppälä, O. (2010). Review of recent systems for automatic assessment of programming assignments. In Proceedings of the 10th Koli calling international conference on computing education research, ACM, 86–93.

Casquina, J. C., & Montecchi, L. (2021, September). A proposal for organizing source code variability in the git version control system. In Proceedings of the 25th ACM International Systems and Software Product Line Conference-Volume A (pp. 82-88).

Ståhl, D., & Bosch, J. (2013). Modeling continuous integration practice differences in industry software development. In Accelerating Digital Transformation (pp. 23-49). Springer, Cham.

Students' Domain Confidence and their Participation in Optional Learnersourcing Activities

Steven Moore[1], Huy Nguyen[2], John Stamper[3]
Carnegie Mellon University, USA
StevenJamesMoore@gmail.com[1], hn1@cs.cmu.edu[2], jstamper@cs.cmu.edu[3]

ABSTRACT: Generating multiple-choice questions (MCQs) is a popular form of learnersourcing that benefits both the students' higher-order thinking and the instructors' collection of assessment items. However, student participation with such learnersourcing activities has often varied when the activities are optional. To better understand how student confidence impacts their engagement with learnersourcing activities, we deployed multiple optional MCQ generation activities across three courses at two community colleges. In an effort to measure if these interventions were reaching all students, we analyzed how students' perceived confidence in the course domain influenced their participation in a set of optional MCQ generation tasks. We found that these optional learnersourcing activities were attempted by students from a wide range of confidence levels.

Keywords: Confidence, Participation, Optional Activities, Learnersourcing

1 INTRODUCTION

The confidence a student has in their own ability to succeed in a domain impacts the motivation they have towards the course, subsequently influencing their participation (Barak et al., 2016). When students are motivated to participate and engage, it leads to higher learning gains, particularly when learning activities are optional and student participation might be decreased. For instance, students may be asked to develop an assessment question, such as a multiple-choice question (MCQ), which will then be answered by another student. This is known as a form of *learnersourcing*, where students complete activities that produce new content that can be leveraged by future students (Moore et al., 2022). These types of question generation activities give us rich student data and help to improve courseware (Moore et al., 2022). However, they are sometimes presented to students in an optional manner depending on the context of the course and learning platform. In this research, we look to better understand how students' confidence in a domain influences their participation with such optional activities.

Previous research has demonstrated that students with greater confidence in their academic abilities often have higher performance and participation rates in courses (Akbari & Sahibzada, 2020). There are several verified methods used by these studies to measure student confidence, such as a survey instrument which asks students to self-report their beliefs in how well they can do in the course domain (Honicke & Broadbent, 2016). These surveys provide an effective and low-stakes method to measure perceived confidence, which often correlates with academic performance. Student confidence has been linked to participation, impacting how often students complete optional activities found in courses (Makhija et al., 2018). In the present study, we investigate: How does a student's self-perceived confidence in the domain affect their participation with optional activities?

2 METHODS

This study was conducted in three different courses at two 2-year community colleges located on the west coast of the United States. All three courses took place online during the fall 2021 semester. The three courses were introductory chemistry, advanced chemistry, and introductory statistics. There were a total of 64 unique students across all three courses. We utilize data that came from four to five week-long units that were used towards the beginning of each course. All three courses were deployed on the same learning platform, which has been used in previous studies involving online learning at community colleges (Moore at al., 2021). Each unit in these courses consists of five to ten related topics and takes roughly one week to cover. The units contain multiple pages of instructional content featuring text and brief instructional videos, along with formative assessments interspersed throughout intended as practice opportunities. They include multiple-choice, short answer, essay, matching, and fill-in-the-blank style questions. In addition to these standard activities, we placed a learnersourcing activity towards the end of each unit in each course that prompts students to generate an MCQ targeting any concept they learned from the unit. The MCQ generation activity can viewed here[1]. Finally, at the conclusion of each unit, students completed a summative quiz that assessed them on the topics covered in the given unit. At the beginning of each course, students were prompted to answer a set of 5-point Likert scale questions assessing their confidence in the course's domain, followed by a brief demographic survey. The five confidence questions were adapted from a verified instrument for measuring confidence in different domains that has been utilized by previous research (Honicke & Broadbent, 2016).

3 RESULTS

The average student rankings (on a scale from 1 to 5) from the five confidence questions were 3.31, 3.85, 4.15, 3.00 and 3.76 respectively. An unpaired two sample t-test revealed no significant difference in the average reported confidence between female ($M=3.59$, $SD=.32$) and male ($M=3.69$, $SD=.14$) students; $t(57)=.61$, $p=.548$. There was likewise no significant difference in the confidence between first-gen students ($M=3.61$, $SD=.28$) and others ($M=3.62$, $SD=.29$); $t(57)=.101$, $p=.921$. A Kruskal-Wallis test was conducted to examine the differences of self-reported confidence and student ethnicities, which revealed no significant difference, $H(2)=2.03$, $p=.363$. To determine if more confident students were more likely to participate in the learnersourcing activities, we conducted an unpaired two-tailed t-test. There was not a significant difference in confidence between students that participated in the activities ($M=3.62$, $SD=.19$) and those that did not participate ($M=3.61$, $SD=.46$); $t(57)=-.05$, $p=.961$. For the 37 students that did participate in some of the learnersourcing activities, there was no significant correlation between their average confidence score and the percentage of learnersourcing activities done out of the five available opportunities they had, $r(35)= -.005$, $p=.976$. We then investigated if a students' self-reported confidence had an impact on their participation and performance with the formative and summative assessments found in the courses. We found no significant correlation between student confidence and quiz scores, $r(57)= -.15$, $p=.267$, or between their confidence and the number of formative assessments they worked on, $r(57)=-.02$, $p=.869$.

[1] https://github.com/StevenJamesMoorer/LAK2023/blob/main/learnersourcing_interface.png

4 DISCUSSION & CONCLUSION

We originally expected that more confident students would both participate in a greater number of formative assessments and learnersourcing activities, in addition to performing higher in the course overall. However, our analysis revealed no significant effect of confidence on any of these participation or performance variables. It is unclear if the disconnect observed with the results of the confidence measures reflect discrepancies between self-reported measures, which have been used for years, and behavioral measures, which have more recently become the focus with improvements made to learning analytics (Quick et al., 2020). Students' responses to survey items about their domain confidence are indirect measures, relying on introspective reports of one's own beliefs and behavior, rather than direct measures of it. While being verified instruments for measuring student confidence, it is possible that the self-report measures themselves systematically biased the responses of the students, encouraging them to report higher levels of confidence in line with a more socially-desirable response (Podsakoff et al., 2003). Future learnersourcing efforts should continue to collect information regarding student confidence, which could be supported by incorporating participation and performance analytics, in addition to self-report measures, to gain a better understanding of how they might influence students' contributions to learnersourcing tasks.

REFERENCES

Akbari, O., & Sahibzada, J. (2020). Students' self-confidence and its impacts on their learning process. American International Journal of Social Science Research, 5(1), 1–15.

Barak, M., Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: Examining aspects of language and social engagement. Computers & Education, 94, 49–60.

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. Educational Research Review, 17, 63–84.

Makhija, A., Richards, D., Haan, J. de, Dignum, F., & Jacobson, M. J. (2018). The influence of gender, personality, cognitive and affective student engagement on academic engagement in educational virtual worlds. Conference on Artificial Intelligence in Education, 297–310.

Moore, S., Nguyen, H. A., & Stamper, J. (2021). Examining the effects of student participation and performance on the quality of learnersourcing multiple-choice questions. In Proceedings of the Eighth ACM Conference on Learning@ Scale, 209-220.

Moore, S., Stamper, J., Brooks, C., Denny, P., & Khosravi, H. (2022). Learnersourcing: Student -generated Content@ Scale. In Proceedings of the Ninth ACM Conference on Learning@ Scale, 259-262.

Moore, S., Nguyen, H. A., Bier, N., Domadia, T., & Stamper, J. (2022). Assessing the quality of student-generated short answer questions using GPT-3. In the European conference on technology enhanced learning, 243-257.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879.

Quick, J., Motz, B., Israel, J., & Kaetzel, J. (2020). What college students say, and what they do: Aligning self-regulated learning theory with behavioral logs. Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, 534–543.

What Happened When a University Retained a COVID-era Flexible Course Withdrawal Policy? Exploring Relationships Between Grading Policy and Equity

Rebecca L. Matz, Caitlin Hayward, Barry Fishman
University of Michigan
rlmatz@umich.edu, cholma@umich.edu, fishman@umich.edu

ABSTRACT: In response to the pressures of the COVID-19 pandemic, colleges and universities made substantial changes to grading-related policies, typically with the goal of providing flexibility to students. As the pandemic changed and waned, institutions often returned to pre-pandemic policies. Here, we explore equity-related outcomes from a university that has thus far retained a COVID-era course withdrawal policy that is more flexible than the pre-pandemic policy. Relying on records from more than 1.6 million course enrollments across nearly 90,000 undergraduate students, we find that the percentage of course enrollments resulting in withdrawals increased significantly from 1.8% in "pre-COVID" terms to 2.9% in "post-COVID" terms. Controlling for student and course characteristics with binary logistic regression including students' typical course performance, we find that the shift to a more flexible course withdrawal policy was associated with an increased likelihood that female (vs. male) students, first-generation (vs. continuing-generation) students, and higher income (vs. lower income) students would withdraw from courses. The policy change was also associated with an increased likelihood of students withdrawing from STEM compared to non-STEM courses.

Keywords: course withdrawal, curricular analytics, grading policy, higher education, STEM

1 INTRODUCTION

The pressure to earn high grades, especially in science, technology, engineering, and mathematics (STEM) courses, has weighed on undergraduate students for too long (Seymour & Hunter, 2019). While moderate academic stress can be a useful motivator for performance, on balance, the traditional grading system is fraught; inequitable policies, content overload, and other factors contribute to excessive pressure and focus on grades instead of learning (Kirschenbaum et al., 2021). At the University of Michigan (UM), the negative outcomes associated with an overemphasis on grades and strict grading policies were recognized well prior to the COVID-19 pandemic but the urgent needs posed by the pandemic provided fuel for making swift grading policy changes. In the name of flexibility, alleviating stress around grades, promoting a focus on learning over penalties, and supporting students from marginalized communities that were more often negatively affected by the pandemic, UM and nearly every other institution of higher education made substantial changes to the traditional grading system in response to COVID-19.

In the interim, higher education institutions have been eager to return to "normal" modes of behavior. Indeed, many universities fully returned to pre-COVID grading policies as soon as Fall 2020, imparting no flexibility for students beyond the first crisis term (Lederman, 2020). At UM, most grading practices reverted to pre-pandemic policies, however, a key COVID-era flexible course withdrawal policy has been retained to date. Here, we evaluate some equity-related outcomes from this policy comparing pre-pandemic results to the most recent academic year. Given the significant effort required to implement grading policy changes at a large university, we contend that earnest assessment of such changes is imperative and that it provides valuable information for other institutions.

2 COURSE WITHDRAWAL POLICY CHANGES

The University has long employed a course withdrawal (also known as drop) policy that is common for semester-based institutions in the United States. Within the first three weeks of the term, students can withdraw without consequence (assuming they remain full-time students)—this type of course withdrawal is not considered herein because of the many reasons that students withdraw early on during the term. Instead, we consider students who withdraw from a course “late” meaning after the “drop date” three weeks into the term. These students receive a “W” as the grade on their transcript.

In Fall 2019 and earlier, students could initiate late course withdrawals between weeks four and nine of the fourteen-week term. Withdrawal requests after week nine were discouraged for all but the most serious circumstances, and they were rarely approved. In Winter 2020 (the initial COVID-impacted term at UM) and the following academic year (Fall 2020 and Winter 2021), students could withdraw up through the last day of the course without consequence and these were not recorded on the transcript. In Fall 2021, UM implemented a version of the withdrawal policy that is substantially like the pre-COVID policy (“W” grades were again recorded on the transcript) with one significant distinction leaving more power in the hands of students: students are now able to initiate late course withdrawals up through the last day of classes. This change added five weeks to the time for students to decide whether to withdraw from a course.

In this study, we identify the groups of students that made the greatest changes in their use of the more flexible withdrawal policy and in which types of courses the greatest changes were observed.

3 METHODS

Deidentified course enrollment, term, and demographic information was collected for undergraduates who enrolled in one or more full-term courses during the timeframes that we defined as “pre-COVID” (Fall 2014 to Fall 2019; 11 terms total) and “post-COVID” (Fall 2021 to Winter 2022; 2 terms total).[1] Summer terms and those most affected by COVID-related grading changes—Winter 2020, Fall 2020, and Winter 2021—were excluded. The dataset consists of 1,698,497 course enrollments from 88,475 students. The data transformations and analyses were carried out primarily with SPSS 28 and the study was considered exempt research by the Institutional Review Board.

4 RESULTS

Withdrawals represent a small but relatively stable percentage of course grades from term to term. Overall, in these pre-COVID terms, 1.8% of course enrollments resulted in withdrawal grades. In the

[1] We use the term “post-COVID” herein solely as a convenient contrast to “pre-COVID” in full recognition that the post-COVID term is imprecise and insufficient to convey the complexities of the timeline of the pandemic yet that most universities had made considerable shifts towards hybrid and even fully in-person modalities by Fall 2021 with concomitant returns to pre-COVID grading policies by then if not before.

post-COVID terms, this percentage increased significantly to 2.9% (χ^2 (1) = 1303, p < .01); the odds of students taking a withdrawal were 1.6 times higher in post- than pre-COVID terms. Indeed, in exploring how different groups of students by demographic characteristics (sex, race/ethnicity, first- versus continuing-generation status, and socioeconomic status) changed their use of the policy, we found that every subgroup increased the rate at which they took withdrawal grades.

We used binary logistic regression to model the relationships between both demographic and course characteristics of interest (STEM vs. non-STEM and upper- vs. lower-division) and the likelihood of students withdrawing from their courses both pre- and post-COVID, including interaction terms, using grade point average in other courses to control for student performance (Huberth et al., 2015). All the following characteristics were associated with higher odds of withdrawing from a course in general: identifying with an underrepresented racial/ethnic group, being a lower-income student, the post-COVID timeframe, the lower-division level, and STEM disciplines. Of these, the most impactful covariate is discipline; the odds of a student withdrawing from a STEM course were 1.9 times higher than those for a student withdrawing from a non-STEM course. Importantly, the interaction terms in the model showed that with the more flexible post-COVID withdrawal policy, withdrawals were even more likely post- as compared to pre-COVID in STEM courses (1.3 times higher), for female students (1.2), for first-generation students (1.2), and for higher-income students (1.1).

5 DISCUSSION

The contrasting results with respect to first- versus continuing-generation status and socioeconomic status are particularly stark. The policy change itself—the addition of several weeks of flexibility for students to withdraw—is associated with a greater likelihood of both first-generation students and higher-income students withdrawing from their courses. Under the broad goal of educational equity, these results are pitted against each other in a way. While the change perhaps provides additional flexibility for first-generation students, we may also be observing unintended consequences; the additional flexibility may have provided a behavioral incentive for more privileged (higher-income) students to exploit the educational system in their favor by facilitating multiple retakes of the same course. Collecting rich qualitative information from students via interviews and surveys centered on their rationale for withdrawing from a course and how the withdrawal has affected or might affect their subsequent academic plans would help explicate the affordances and constraints of the more flexible withdrawal policy. Future questions of interest include the effects of withdrawal on students' other course grades and withdrawal rates and on course retakes in subsequent terms.

REFERENCES

Huberth, M., Chen, P., Tritz, J., & McKay, T. A. (2015). Computer-Tailored Student Support in Introductory Physics. PLOS ONE, 10(9), e0137001. doi.org/10.1371/journal.pone.0137001

Kirschenbaum, H., Napier, R., Simon, S., & Fishman, B. (2021). Wad-Ja-Get?: The Grading Game in American Education, 50th Anniversary Edition. Michigan Publishing.

Lederman, D. (2020, August 12). Grading in a Pandemic (Still). Inside Higher Ed. insidehighered.com/digital-learning/article/2020/08/12/many-colleges-will-return-normal-grading-fall-will-semester-be

Seymour, E., & Hunter, A.-B. (2019). Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education. Springer International Publishing.

Learning Analytics for Last Mile Students in Africa

Christine Kwon, Ren Butler, John Stamper, Amy Ogan
Carnegie Mellon University
{ckwon2, ddbutler, jstamper, aeo}@andrew.cmu.edu

Angela Forcier, Erin Fitzgerald, Samson Wambuzi
Yiya Solutions, Inc.
{angela, erin, samson}@yiyasolutions.org

ABSTRACT: Although technology is ubiquitous in the homes and classrooms in the Global North, access to educational technology in the Global South is still out of reach for much of the population. This work describes an educational technology system, Yiya Air Science, to reach "last mile" students in Africa (specifically rural Uganda) where access to basic computers or smartphones is rare. Courses are deployed on a system of radio broadcast and basic texting phones (USSD). Yiya Air Science has deployed three courses focused on STEM skills and entrepreneurship as key learning outcomes. In this research, we are adopting state of the art learning science and engineering principles with fine-grained data collection and a continuous improvement loop.

Keywords: Global South, Rural Africa, Phone, USSD, Multimodal Learning

1 INTRODUCTION

Yiya Solutions, Inc.[1] is building and studying an offline virtual classroom created for adolescent learners in rural Uganda that provides STEM education through radio broadcasts and basic keypad phones called Yiya Air Science. Their remote learning model does not require access to the internet, smart devices, or other advanced technologies. Thus, using simple interactive technologies, radio broadcasts, and keypad phones eliminates both physical and financial barriers that prevent many African children from receiving any sort of educational experience. Yiya Air Science ultimately aims to provide equitable access to higher quality education through interactive technologies for all Ugandan children living in under-resourced and remote communities. The curriculum develops resilience and problem solving in African youth through its advanced STEM educational content. Our primary goal in this research is to develop a research driven implementation with rich learning analytics and strong feedback loops for learning centered system improvement.

2 THE EDUCATIONAL TECHNOLOGY

The STEM content within Yiya Air Science aims to increase learner competencies and passion in pursuing STEM education. The course is structured to provide learners with skills to solve problems in their local communities and to make a change within their developing communities. The STEM content within the curriculum centers on students performing hands-on technological experiments at their homes, creating innovative quotidian products, such as solar food dryers and pedal powered

[1] https://www.yiyasolutions.org/

washing machines. Yiya Air Science's course content is taught to students on a daily basis, each week focusing on a step to construct the final innovative product. Every week, instructors inform students on the science and engineering methods behind each step in building the final product, interacting with students through daily questions, product building exercises, and creative examples related to the product. The duration of each course is about twelve weeks, beginning with an initial baseline survey and finishing with a final assessment and endline survey.

Even today, radios and basic keypad phones are the most commonly used and present technologies in Ugandan households, with 74% of these households possessing radios and eighty-seven percent possessing keypad phones (BBC Media Action, 2019). These technological items are also shared amongst neighboring households, making them easily accessible for all learners in Yiya Air Science. The modality of how learners utilize the interactive technologies, radios and keypad phones, used in Yiya Air Science is shown in Figure 1, and can be informed by the use of multimodal learning analytics (Liu et al., 2018). Each day, students will turn on the radio to access and listen to the channel that provides the lesson. Students prepare the materials needed to conduct a step in building the final product of the course, and follow the instructions on completing the step given by the instructor speaking in the radio broadcast. Instructors ask questions for listeners to answer through live call-ins or through the USSD application. Learners also complete weekly formative quizzes on phones.

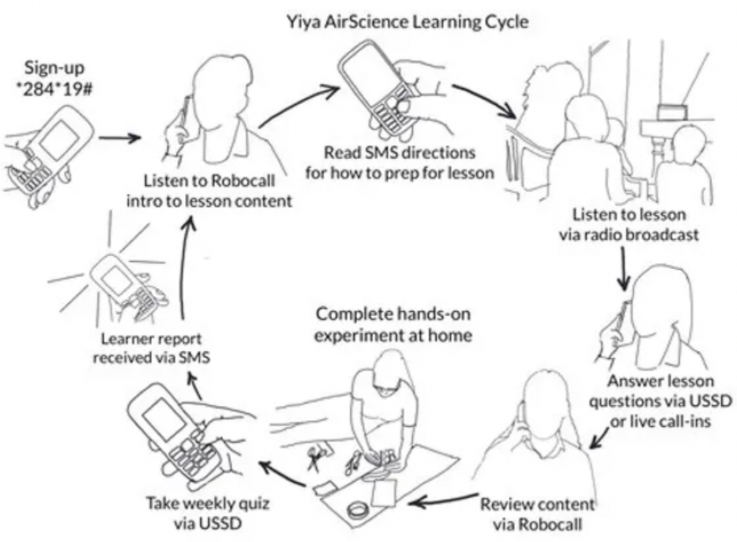


Figure 1: Students listen to lessons on the radio and interact with the course through USSD texting.

At the end of the course, students are given a final summative assessment to review their knowledge and developed skills from the course. Instructors give the final assessment on radio by reading out each question and the answer options. Students provide their answers through the USSD application. Students who receive a passing grade of 75% or higher receive a certificate. Students are also encouraged to complete the endline survey to provide data on the impact of the course, compared to the baseline taken at the start of the course.

While courses are available to anyone, Yiya Air Science is designed for youth in low-resourced regions that are typically ruralistic, which prevent access to a multitude of basic educational opportunities. Youth living in rural areas must overcome cumbersome obstacles in order to receive the most basic educational experience. These obstacles include long physical distances from homes to the closest school, which may be hazardous for these children to travel, and expensive schooling fees (Uchidiuno et al., 2018). Children's labor is also unavoidably needed for subsistence farming, as it is the main source and income for many families living in rural regions, and other core responsibilities at their homes. Girls are more likely than boys to be kept home to fulfill household responsibilities. Living in rural areas also restricts youth from accessing technical resources, including internet connectivity, let alone basic educational resources (Uchidiuno et al., 2018). A survey conducted in 2017 showed that only 4% of Ugandans had used a computer in the previous 3 months (UNHS, 2018). Yiya Air Science is centered around youth in Northern Uganda who struggle from these impediments, compounded by decades of internal conflict. There are twenty-two million children in Uganda, of which only thirty-five percent of those children complete primary school (UNHS, 2018). The lack of education for younger Ugandan children is concerning, as forty-six percent of Ugandan children of ages six through twelve have never attended school. The impact of COVID-19 has only significantly decreased the number of children who attend schooling institutions.

3 LEARNING ANALYTICS

Three different courses have been created and run in Yiya Air Science. Each course focuses on building a STEM based tool that can be used by the student. These include a solar food dehydrator, solar cell, and pedal-powered washing machine. All USSD phone data is logged to a SQL database backend that allows for robust analytics on the fine-grained interaction data. In addition, we apply multi-modal analyses that line up the student interactions with the radio broadcast similar to research on log data and video (Lui et al., 2019). This allows us to understand which broadcast the student is listening to and provide interventions to students through texts or robocalls. We examine a number of learning analytics including tracking learning by demographics and are building models to predict engagement and dropout utilizing existing model designs (Chatterjee et al., 2020) with the goal of adding interventions to encourage students to continue in the course.

REFERENCES

BBC Media Action Uganda - Media Landscape Report (February 2019). Retrieved from: https://www.communityengagementhub.org/wp-content/uploads/sites/2/2019/09/Uganda-Media-Landscape-report_BBC-Media-Action_February-2019.pdf

Chatterjee, R., Madaio, M., & Ogan, A. (2020). Predicting Gaps in Usage in a Phone-Based Literacy Intervention System. In *International Conference on AIED* (pp. 92-105). Springer, Cham.

Liu, R., Stamper, J., Davenport, J., Crossley, S., McNamara, D., Nzinga, K., & Sherin, B. (2019). Learning linkages: Integrating data streams of multiple modalities and timescales. *Journal of Computer Assisted Learning*, 35(1), 99-109.

Liu, R., Stamper, J. C., & Davenport, J. (2018). A novel method for the in-depth multimodal analysis of student learning trajectories in intelligent tutoring systems. Journal of Learning Analytics, 5(1), 41-54.

Uchidiuno, J., Yarzebinski, E., Madaio, M., Maheshwari, N., Koedinger, K., & Ogan, A. (June 2018). Designing Appropriate Learning Technologies for School vs Home Settings in Tanzanian Rural Villages. In *ACM Conference on Computing & Sustainable Societies (COMPASS '18).* (pp. 1-11).

UNHS (Uganda National Household Survey) 2016/2017 Report (2018). Uganda Bureau of Statistics. www.ubos.org/wp-content/uploads/publications/03_20182016_UNHS_FINAL_REPORT.pdf

Classroom Assessment Ecosystem: Exam Analytics for Closing Group Performance Gaps

Bo Pei[1,2], G. Alex Ambrose[2], John Behrens[2], Alison Cheng[1,3], Eva Dziadula[4]

[1]Lucy Family Institute for Data & Society, University of Notre Dame, IN 46556, United States
[2]Office of Learning Analytics, University of Notre Dame, IN 46556, United States
[3]Department of Psychology, University of Notre Dame, IN 46556, United States
[4]Department of Economics, University of Notre Dame, IN 46556, United States
{bpei, gambrose, jbehrens, ycheng4, eva.dziadula.1}@nd.edu

ABSTRACT: In this poster, we report an exam analysis with the purpose of identifying and closing student performance gaps. Specifically, we first conducted the difficulty and discrimination level analysis to identify the low-quality questions that are not fulfilling the testing purposes. Next, on each question, we explicitly investigated the performance gaps of the predefined performance groups. Finally, we conducted a text analysis to analyze the differences in the knowledge mastery levels of different groups. Our analytic results, on the one hand, specified the low-quality questions from different metrics for instructors to revise in future instructional practice. On the other hand, we also highlighted the instructional significance in conducting course level assessments for improving teaching and learning efficiency.

Keywords: Item analysis, group performance discrimination analysis, performance gap, text analysis

1 BACKGROUND

Classroom testing is an important way for measuring student learning achievements in the teaching and learning environments. Testing results can both demonstrate the level of students' knowledge proficiency and indicate the effectiveness of instructors' teaching practices. A high-quality test is expected to discriminate different student performance groups, collect accurate evidence related to their learning achievements, and offer targeted insights for instructors to provide timely and personalized feedback (Behrens, 1997; Palimbong et al., 2018). In this study, we conducted a holistic analysis to examine the quality of questions in the final exam of a course called "Principles of Microeconomics" instructed at the University of Notre Dame during the 2021/22 academic year.

Particularly, this poster reports the collaborative efforts of the course instructor as a domain expert and the learning analytics group for analyzing student learning outcomes with the goal of optimizing the instructional practice to close the performance gap. As such, we first conducted question level analysis, such as difficulty level analysis (Lourdusamy & Magendiran, 2021) and discrimination analysis (Manalu, 2019). With this analysis, we can specifically identify the questions that need to be revised or do not fulfill the testing purpose. Then, based on the predefined student performance groups (i.e., Thriving Group, NonThriving Group, and DFW Group) with the course instructor, we investigated how different student groups performed on each question and identified the most difficult topics for each group via group performance gap analysis and topic analysis. Specifically, in this poster, we investigated the answers to the following questions:

(1) What are the difficulty and discrimination levels of each question in the final exam?
(2) What are the performance gaps of the three groups on each question and the corresponding differences in knowledge proficiencies?

2 METHOD AND RESULTS

Our analysis focused on the questions in the final exam adopted in the course. The final exam includes 50 questions, 46 of them are multiple selections and 4 of them are filled in blank questions. To obtain the completed information about the knowledge tested in the exam, we analyzed both the question stems and options. There were 200 students participating in the exam in total.

2.1 Student performance groups clustering

We first clustered students into three different performance groups based on their final performance. After consulting with the course instructor, we define students with a score above B- as Thriving Group, between B- and C- as NonThriving Group, and below C- as DFW group. Figure 1 shows the basic statistical description of student performance in the final exam, from which we can find that there is a clear difference in the final exam performance of students in the three different groups. Specifically, Figure 1(a) describes the histogram of the student final exam performance, Figure 1(b) examines the distributions of all the students' performance, and Figure 1(c) compares the performance differences across different performance groups.

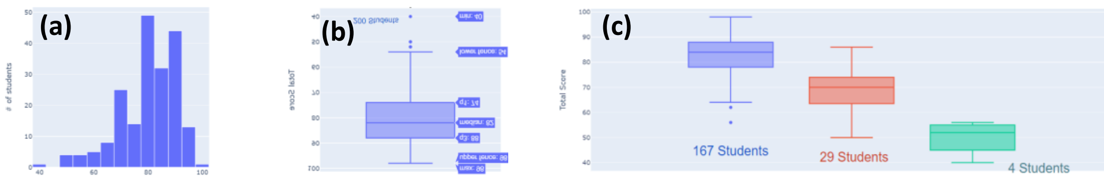


Figure 1: The basic statistical description of student performance in the final exam.

2.2 Question Difficulty level analysis and Discrimination analysis

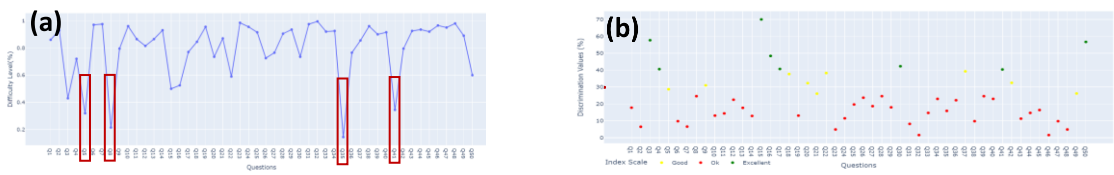


Figure 2: Question level analysis. (a) Question difficulty level analysis. (b) Question discrimination level analysis

From Figure 2(a), we can find that question5, question8, question35 and question41 have higher difficulty levels with less than 40% of students answering correctly. As shown in Figure2(b), the question discrimination index was calculated as the difference on the average exam score of top 27% of students and bottom 27% students (DeVellis, 2006). If the difference of the scores is in the range of 40% to 100%, we treat the question as having an excellent discrimination ability; between 25% to

39% with a good discrimination ability; and lower than 24% with an OK discrimination ability. As such, the questions with OK discrimination ability below the red line may need close examination to improve their corresponding discrimination abilities.

2.3 Performance gap analysis and topic analysis

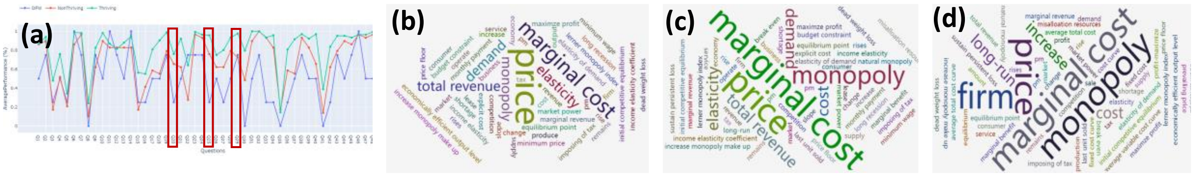


Figure 3: Group performance analysis. (a) group performance gap on each question. (b)-(d) are the top 10 most difficult topics for students in Thriving, NonThriving and DFW Group, respectively

From Figure 3(a), we can find that the large performance gaps among the three performance groups occur at the questions such as question20, question25, question30, and so on. According to Figure 3 (b)-(d), we can find that there are some topics all the three groups had difficulties with such as marginal cost, price setting, etc. And for some topics like those related Monopoly topics, the students in NonThriving and DFW groups seem to have more difficulties than those in the Thriving group.

3 CONCLUSION

Our study conducts a holistic analysis on the final exam questions with a purpose of analyzing the differences in knowledge mastery levels of student performance groups to support instructors to generate targeted interventions to close the performance gap. Particularly, our analysis allows instructors to identify questions that need to be better designed in terms of both difficulty level and discrimination ability to fulfill the assessment purpose. Moreover, by employing text analysis, in combination with the identified differences in knowledge mastery levels, our study also allows instructors to provide specific learning interventions for each group of students to close the performance gaps.

Acknowledgement: We appreciate the support from Kevin Abbott and Andrew Craker for collecting data on our project.

REFERENCES

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*, 131–160. https://doi.org/10.1037/1082-989X.2.2.131

DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, *44*(11), S50–S59.

Lourdusamy, R., & Magendiran, P. (2021). A systematic analysis of difficulty level of the question paper using student's marks: A case study. *International Journal of Information Technology, 13*(3), 1127–1143. https://doi.org/10.1007/s41870-020-00599-2

Manalu, D. (2019). An Analysis of students Reading Final Examination By Using Item Analysis Program on Eleven Grade of SMA Negeri 8 Medan. *JETAL: Journal of English Teaching & Applied Linguistic, 1*(1), Article 1. https://doi.org/10.36655/jetal.v1i1.98

Palimbong, J., Mujasam, M., & Allo, A. Y. T. (2018). Item Analysis Using Rasch Model in Semester Final Exam Evaluation Study Subject in Physics Class X TKJ SMK 2 Manokwari. *Kasuari, 1*(1), 43–51. https://doi.org/10.30862/kpej.v1i1.807

ITM-Rec: An Open Data Set for Educational Recommender Systems

Yong Zheng
Illinois Institute of Technology, USA
yzheng66@iit.edu

ABSTRACT: With the development of recommender systems (RS), several promising systems have emerged, such as context-aware RS, multi-criteria RS, and group RS. However, the education domain may not benefit from these developments due to missing information, such as contexts and multiple criteria, in educational data sets. In this paper, we announce and release an open data set for educational recommender systems. This data set includes not only traditional rating entries, but also enriched information, e.g., contexts, user preferences in multiple criteria, group compositions and preferences, etc. It provides a testbed and enables more opportunities to develop and examine various educational recommender systems.

Keywords: recommender system, data set, context-aware, group, multi-criteria

1 INTRODUCTION

Nowadays, recommender systems (RS) have become one of the major tools in technology-enhanced learning. RS have been introduced to the education domain (Santos et al., 2011) to deliver personalized learning, recommend textbooks or informal learning materials, assist decision-making in group studies or teamwork, and provide better adaptive learning in mobile environments, etc.

Various RS were well-developed with enriched information, such as contextual variables and multi-criteria user preferences. However, the education domain may not benefit from these RS due to a lack of corresponding data sets with necessary information. In this paper, we fill this gap by announcing and releasing the ITM-Rec data set, which can be utilized as a testbed to develop and examine different types of RS (Zheng & Mobasher, 2018; Zheng, 2018; Zheng et al., 2022) in educations.

2 ITM-REC: DATA COLLECTIONS, STATISTICS AND USAGE

2.1 Data Collections

The data set was collected from a questionnaire using Qualtrics from 2017 (Zheng, 2018). The subjects were graduate students enrolled in the specialization of data management and analytics at the ITM department in Illinois Tech. The questionnaire was designed to collect student preferences on the topics of the final projects in three courses: database (DB), data analytics (DA), and data science (DS).

More specifically, students in the DA and DS classes were given a list of Kaggle data sets as candidates for analysis in their final projects. Each student was asked to select at least three data sets they liked and three that they did not like. They were asked to rate their selections by giving an overall rating and additional ratings for three criteria: App (their liking of the data's application domain), Data (their liking of the data's processing or storage), and Ease (their liking of the degree of ease in using the data for the final project). In the DB class, students were asked to build web information systems with connections to relational databases, and were asked to rate different potential project topics (e.g., hotel booking systems, hospital appointment systems) by giving overall ratings and the same multi-criteria ratings. The questionnaire was first assigned to each student to collect individual preferences. In addition to user and group preferences, we also collected student demographic information (e.g., age, gender, marriage status) and item content features (e.g., URL of Kaggle data, title of the data, textual descriptions of the data).

Regarding the final projects, students had the option to work independently or as part of a team. If a team was formed for the final project, each group of students were asked to complete the questionnaire again as a second survey. It is worth noting that this second survey was assigned to each group and they were required to submit a single copy of the questionnaire after group discussions. Their input can be considered as group preferences on the items, rather than individual preferences.

2.2 Data Description and Statistics

We have collected the data from 2017 to 2022. The user information has been well anonymized, and data set will be released on Github[1] and Kaggle[2]. The whole data set is composed of five major files:

- Users.csv, where we provide meta data (ID, gender, age, etc.) about 476 unique students.
- Items.csv, where we provide meta data (ID, title and descriptions) about 70 unique items.
- Ratings.csv refers to individual preferences and it contains 5,230 ratings given by 476 users over 70 items. An example is shown in Table 1. In addition to overall and multi-criteria ratings, context information such as the course (DB, DA or DS), semester (Spring, Fall) and COVID-19 lockdown periods (PRE, DUR, POS) is provided. PRE refers to the timeframe from 2017 Fall to 2019 Fall, DUR refers to the timeframe from 2020 Spring to 2021 Spring, and the POS period refers to the timeframe from 2021 Fall to 2022 Fall.
- Group_ratings.csv refers to group preferences and contains 1,117 ratings given by 143 groups.
- Group.csv describes the composition of groups. There are 143 groups, where 88, 42, 9, 4 groups have a group size of 2, 3, 4, 5, respectively.

Table 1: Example of Individual Ratings by Students

UserID	*ItemID*	*Rating*	*App*	*Data*	*Ease*	*Course*	*Semester*	*Lockdown*
1173	28	5	4	4	4	DA	Fall	PRE
1175	41	5	4	4	4	DS	Spring	POS
…	…	…	…	…	…	…	…	…

2.3 Data Usage

Due to the enriched information, this data can be utilized as a testbed to develop and examine various recommender systems. Below are existing or possible examples by using this educational data set:

[1] https://github.com/irecsys/RecData

[2] https://www.kaggle.com/datasets/irecsys/itmrec

- *Traditional RS with side information* (e.g., user demographic information and item features), or identification of *grey sheep users* in RS (Zheng et al., 2017).
- *Context-aware RS* (Zheng & Mobasher, 2018; Zheng, 2014) which adapt the recommendations to different contexts (e.g., course, semester, lockdown periods), or context suggestions (Zheng et al., 2016) which recommend contexts, rather than items.
- *Multi-criteria RS* (Zheng & Wang, 2022) which enhance recommendations by taking advantage of user preferences in multiple criteria (e.g., App, Data, Ease).
- *Group RS* (Zheng, 2018) which recommend items to each group, rather than individuals.
- *Multi-objective RS* (Zheng & Wang, 2022) which optimize multiple objectives in RS.
- *Integrated RS with multi-factors*, where researchers can build RS by integrating various factors together, e.g., integrating the context information in multi-criteria RS (Zheng et al., 2019), utilizing multi-criteria preferences towards group RS (Zheng, 2019), etc.

3 CONCLUSIONS & FUTURE WORK

In this paper, we announce and release an open data set for educational recommender systems, where enriched information (e.g., contexts, multi-criteria preferences, group compositions and preferences) are available. As a result, this data set can be utilized to develop and examine different types of educational recommender systems. We may continue the process of data collections, and release another version of this data in the next few years.

REFERENCES

Santos, O. C., & Boticario, J. G. (Ed.). (2011). Educational Recommender Systems and Technologies: Practices and Challenges, IGI Global.

Zheng, Y. (2018). Personality-aware decision making in educational learning. In ACM IUI companion.

Zheng, Y. and Mobasher, B. (2018). Context-Aware Recommendations. In Book: Collaborative Recommendations: Algorithms, Practical Challenges and Applications, World Scientific Publishing, 173-202

Zheng, Y. (2014). Deviation-based and similarity-based contextual SLIM recommendation algorithms. In Proceedings of the 8th ACM Conference on Recommender Systems.

Zheng, Y., Mobasher, B., & Burke, R. (2016). User-oriented context suggestion. In ACM UMAP.

Zheng, Y., & Wang, D. (2022). Multi-Criteria Ranking: Next Generation of Multi-Criteria Recommendation Framework. IEEE Access, 10, 90715-90725.

Zheng, Y. (2018). Exploring user roles in group recommendations: A learning approach. In Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization.

Zheng, Y., & Wang, D. X. (2022). A survey of recommender systems with multi-objective optimization. Neurocomputing, 474, 141-153.

Zheng, Y., Agnani, M., & Singh, M. (2017). Identification of grey sheep users by histogram intersection in recommender systems. In 13th Conference on Advanced Data Mining and Applications.

Zheng, Y., Shekhar, S., Jose, A. A., & Rai, S. K. (2019). Integrating context-awareness and multi-criteria decision making in educational learning. In ACM/SIGAPP Symposium on Applied Computing.

Zheng, Y. (2019). Educational group recommendations by learning group expectations. In IEEE TALE.

Addressing Promotion System Based on Student Data to Support Desk-to-Desk Instruction by Teaching Assistants

Shin Ueno, Takahiro Yoshino and Hironori Egi
The University of Electro-Communications
u2130010@edu.cc.uec.ac.jp, hiro.egi@uec.ac.jp

ABSTRACT: Students are often believed to be solitary in programming exercise classes where they have to work on programming assignments alone. Teaching assistants (TAs) can reduce the sense of loneliness among students. Thus, in this study, we propose a system to support TAs in addressing students in desk-to-desk instruction. The proposed system uses students' data from questionnaires on student attendance, homework submission rate, review tests, and sense of loneliness to determine the priority in which to address the students. The system then automatically specifies which students should be addressed. In addition, the system presents random scenarios for TAs to address students along with hints on how the students should be addressed. The proposed system was introduced in a programming exercise class for first-year undergraduate students at a science and engineering university. Our qualitative findings showed that the system facilitated the TAs in addressing of students, which may have reduced the students' sense of loneliness.

Keywords: Teaching Assistant, Student data, Loneliness, Addressing, Scenarios, Tablet Computer

1 INTRODUCTION

Social isolation and loneliness are important aspects of the sustainable development goals established by the United Nations based on the basic principle of "No one will be left behind" (RISTEX, 2021). In general, students in programming exercise classes tend to feel isolated when they are required to do programming assignments alone. In such cases, teaching assistants (TAs) can play an important role in reducing the sense of loneliness among the students. In addition, the friendly behavior of TAs may positively impact the learning experience of students and contribute to improving their retention rate (C.O' Neal et al., 2007). For these reasons, we consider that TAs are necessary for students even today, when there exist many systems that support learning among students without TAs.

A system to support TAs is proposed based on a learning support strategy that defines which students should the TAs support during class (Imamura et al., 2020). This system lists out a number of students who require additional support from TAs based on their learning status, and the TAs themselves select the students to be guided and go to support. Our experimental results suggested that the system may encourage the TAs to address students. However, this system has two issues. First, which particular students need support is not provided. Second, the system does not provide how the TAs should support the students.

Thus, this study proposes a system to support TAs in addressing students in desk-to-desk instruction. The proposed system automatically indicates which students should be addressed. The system also presents random scenarios for addressing students and provides hints on how to address them. When TAs are equipped with these details, the number of times they address students will increase. We considered that the increase in the number of addressing and the increase in the number of communications between TAs and students could be an opportunity to reduce students' sense of loneliness.

2 METHODOLOGY

In this study, we propose a system that acquires data on individual students and presents scenarios for addressing each student. Figure 1 shows an overview of the system and a photo of TA holding a tablet.

Figure 1: Overview of the method & Photo of TA holding with tablet

The upper left corner of the tablet's screen displays the identification number and the name of the student with the highest priority to be addressed. The upper right corner of the screen displays the seating chart of the students, and the seat in which the student is sitting is shown in red. The lower right portion of the screen retrieves and displays the student's attendance rate, homework submission rate, percentage of correct answers to review tests, score in the previous class review test, and score of the day's review test. Random scenarios for addressing are displayed in the left center of the screen. In the lower left part of the screen is a button to register the result of addressing that student. The blue button denotes that the TAs have approached the student and determined that the student can work on their own during the class. The red and yellow buttons denote that the TAs find it necessary to address the student again during the class. When the TA presses one of the buttons, the next student to be addressed is displayed in the same manner.

The queue is intended to determine which student should be addressed (A, B, C..., address a student). The results of a review test given to each student at the beginning of the class and the students' responses to a questionnaire about loneliness allow the TAs to make the decision to address a student. In the programming practice class in which this system is to be used, the students were tested in the review test at the beginning of the class. Based on the results of the review test, the students are categorized into three groups considering that the result of the review test indicates the student's level of understanding of the class content. For each group, the students are sorted in order of their scores on the loneliness questionnaire for feeling lonely (6-point scale of disagree to agree) as a queue (Priority Que). The first student in the group with the highest priority in the queue created above is displayed on the screen (student B is presented first in the figure). With respect to the reinsertion of the corresponding student into the queue, three buttons of blue, red, and yellow colors are provided and if the blue button is pressed, the student is placed at the end of the queue in the group with the lowest priority. In comparison, the student is placed at the end of the queue in the group with the original priority if the red or yellow button is pressed. Thus, we established the above rules for reinsertion into the queue, from the perspective of equality in teaching students.

The scenarios for addressing students were determined by interviewing undergraduate and graduate students with TA experience and excluding duplicate responses. We prepared 20 sets of scenarios multiplied by four of the determined scenarios and designed the system so that the scenarios would be randomly displayed on pressing the buttons to display other scenarios.

3 PRACTICE IN CLASSROOM LESSONS AND RESULTS

The system was introduced in a programming exercise class for first-year undergraduate students at a science and engineering university. The classes were taught by one teacher and two TAs, and four TAs were the subjects of analysis in two classes dealing with the same content. In class 1, the system was introduced in the 8th lesson. Our findings showed that the number of addressing times for the 8th session was 11 for TA1 and 7 for TA2. In addition, the score of questionnaires on student's sense of loneliness was improved at lesson #08 (n = 52, Ave. = 2.15, S.D. = 1.35) when compared with lesson #07 (n = 52, Ave. = 2.15, S.D. = 1.15). In class 2, the system was introduced in the 8th and 9th lessons. The number of addressing times for the 8th session was 6 for TA3 and 4 for TA4, and the score of questionnaires on student's sense of loneliness was found to be worsened at lesson #08 (n = 60, Ave. = 2.36, S.D. = 1.20) compared with lesson #07 (n = 58, Ave. = 2.36, S.D. = 1.27). For the 9th session, the number of addressing times for the 9th session was 17 for TA3 and 8 for TA4, and the score of questionnaires on student's sense of loneliness was found to be improved at lesson #09 (n = 60, Ave. = 2.10, S.D. = 0.85) compared with lesson #08 (n = 60, Ave. = 2.36, S.D. = 1.20). TA1 commented that he had previously been unable to address those students who had not completed review tests or homework because he did not know how to address such students, but this system enabled him to do so. Based on the above results, the system may have promoted TAs to talk to students, thereby reducing the students' sense of loneliness. On the other hand, the strength and timing of the effect on the sense of loneliness may differ depending on the class and TA. Therefore, it is necessary to conduct additional research needs to be conducted by increasing the number of classes and number of experiments in the future.

REFERENCES

Research Institute of Science and Technology for Society. (2021). *SOLVE for SDGs: Social Isolation & Loneliness (FY2021-).* Retrieved from https://www.jst.go.jp/ristex/en/funding/solve-koritsu/index.html

O' Neal, C., Wright, M., Cook, C., Perorazio, T., Purkiss, J. (2007). The Impact of Teaching Assistants on Student Retention in the Sciences: Lessons for TA Training? *Journal of College Science Teaching*, 36(5), 24–29.

Imamura, R., Terui, Y., Yokoyama, Y., Egi, H. (2020). Evaluation of the Learning Support Strategies for Teaching Assistants. *Proc. of IEEE TALE2020,* 120–126.

Uncovering Features of Discourse that Increase Interactions

Jenna Matthews, Ha Nguyen, Hillary Swanson
Utah State University
jenna.matthews@usu.edu, ha.nguyen@usu.edu, hillary.swanson@usu.edu

ABSTRACT: Isaac Asimov discussed *the parable of the cubic meter*, in which a group of individuals are unable, through their own unaided strength, to lift a cubic meter of platinum due to its weight. He compares this to a problem that cannot be solved because it isn't possible to get enough people around it (Asimov, 1989, p.62). Some problems are larger than what a single mind can solve alone; these problems require collaboration or collective intelligence to arrive at a solution. As learning scientists, improving our understanding of collaborative problem solving – and what makes that problem solving effective – can help us design collaborative processes that more effectively lead to solutions. In this paper, we introduce a technique for automating the identification of a key moment in group discourse (an interruption) that leads to an increase in interactions among group members, thus enriching the group's thinking.

Keywords: Interactions; Automation; Text Analysis; Collective Intelligence; Collaboration

1 WHY INTERACTIONS MATTER

Interactions have been shown to have a powerful and causal relationship with what is learned and retained (Koedinger et al., 2018; Van Campenhout et al., 2021). In theories of distributed cognition (i.e., the idea that intelligence is distributed across people, tools, and contexts), these interactions are regarded not as evidence of thinking, but as thinking itself (Hutchins, 2008). Building on these concepts, it seems reasonable to expect that increasing interactions increases the thinking that occurs in the group, which can have a positive effect on learning. To help learning communities become more interactive we need to better understand what factors lead to a change – especially an increase – in interactions. By describing key changes in interactional patterns, we can lay the groundwork to develop computational methods for analyzing larger data sets of group interactions.

1.1 The Polymath Project

In prior research, we analyzed data from the first Polymath Project (Cranshaw & Kittur, 2011) to understand productive moves in distributed knowledge-construction activities (Matthews & Swanson, in review). In the Polymath project, a group of individuals with a range of mathematical expertise was able to solve, within a few months, a problem that had remained unsolved for some time (Polymath, 2012). A proof that seemed an impossible target was created by this group in a relatively short time. This collaborative work was facilitated by technological tools and artifacts - from computers connected through the internet to a blog facilitating discussion threads. In our prior work, we evaluated this robust data set (consisting of blog posts and response threads) to better understand the processes involved in distributed knowledge construction.

2 THE INTERRUPTION AND ITS IMPACT ON GROUP INTERACTIONS

Through a qualitative coding process, we identified a key moment where a question served as an interruption that changed the interactional pattern of the group. Recalling from Hutchins (2008) that interactions are the thinking process itself, we interpret a change in interactional patterns as a change in thinking. In our data, the group went from a monologue (i.e., a large word count with a single, distinct speaker) to a conversation (i.e., additional speakers with smaller individual word counts), following the interruption. Figure 1 represents the change in the group's interactions from left to right. Each circle represents a speaker's contribution; its radius represents the word count per comment. The single grey circle represents a lengthy contribution made by a single speaker. The smaller multi-colored circles represent the shorter contributions of seven speakers, including the speaker from before the *interruption*~~.~~ Our qualitative examination revealed that the post-interruption comments addressed the interrupting question in relevant and diverse ways. This is most noticeable in the way that the interrupting question challenged assumptions from earlier in the conversation. Indeed, many post-interruption comments reconsider these earlier assumptions. Recognizing the importance of interruptions, we examine automatic ways to detect these moments in this paper.

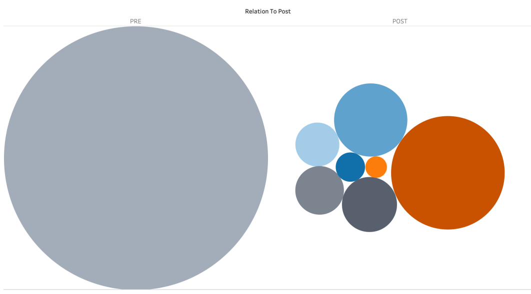


Figure 1: The impact of interruption – more speakers with fewer words per speaker

3 DETECTING INTERRUPTIONS

3.1 Manual Process

3.1.1 Data Set

We started with a data set of 765 comments posted on the blog of a leading mathematician near the beginning of the first Polymath project. We selected this data set to investigate the nature of distributed knowledge-construction in the early phase of the Polymath project. Furthermore, the data set provided multiple interesting interactions between individual mathematicians, as revealed by our pilot analyses. We removed comments that happened well after the initial set of interactions. This left us with a subset of 552 comments, all of which happened between January 27, 2009 at 4:47 pm and February 15, 2009 at 10:59 pm (times are taken from blog post data).

For each remaining post, we calculated the word count, then a running total of the word counts and speakers for the previous 10 posts. These running totals were used to create a ratio of words to speaker. For each post that ratio was compared to the ratio for the next 10 posts (placing the post in the center of a before-after picture). A decrease in ratio (before - after > 0) means that there were more interactions after the post (more speakers, lower word:speaker ratio) than before. Because we are comparing the 10 posts before with the 10 posts following, the first and final 10 posts in the data set are not included in the analysis. We are therefore working with a data set of 532 comments.

3.1.2 Posts Increasing Interactions

Out of the 532 comments from the previous step, we were left with 289 posts which led to a decrease in the word:speaker ratio, meaning those which led to an increase in interactions. These decreases in ratio varied significantly, from 3.05 to 17,638.67. Sorting from largest to smallest variation can allow us to focus our time on the most impactful interruptions, or those which change the word:speaker ratio - and therefore the number of interactions - the most.

4 NEXT STEPS – AUTOMATING THE DETECTION OF INTERRUPTIONS

By automating this process – detecting those moments where *something* happens to cause an increase in interactions – we can focus our time as researchers on understanding what it is that increases the interactions and the discourse that follows. We agree with Hutchins (2008) that these interactions *are* the thinking process, and with Koedinger et al., (2016) and Van Campenhout et al., (2021) that the interactions increase understanding and retention. It follows that any automation that helps researchers understand how to increase interactions allows us to better design the collaborative work around our difficult problems - our cubic meters - to increase a group's thinking and their progress toward a solution.

REFERENCES

Asimov, I. (1989). *Prelude to Foundation*. Random House Worlds.

Cranshaw, J., & Kittur, A. (2011). The polymath project: Lessons from a successful online collaboration in mathematics. CHI 2011, May 7–12, 2011, Vancouver, BC, Canada. 1865–1874. http://dx.doi.org/10.1145/1978942.1979213

Hutchins, E. (2010). Cognitive Ecology. Topics in Cognitive Science, 2(4), 705–715. https://doi.org/10.1111/j.1756-8765.2010.01089.x

Is massively collaborative mathematics possible? (2009, January 27). *Gowers's Weblog*. https://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/

Koedinger, K. R., McLaughlin, E. A., Jia, J. Z., & Bier, N. L. (2016). Is the doer effect a causal relationship? How can we tell and why it's important. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, 388–397. https://doi.org/10.1145/2883851.2883957

Matthews, J., Swanson, H. (in review). *The Pivot: Codifying Emergent Tactics in Distributed Epistemic Games*. Manuscript submitted for publication.

Polymath, D. (2012). A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, 1283–1327.

Van Campenhout, R., Johnson, B., & Olsen, J. (2021). The Doer Effect: Replicating Findings that Doing Causes Learning. Proceedings of the Thirteenth International Conference on Mobile, Hybrid, and On-line Learning, eLmL 2021, pp.1-6, 2021, Online. ISBN: 978-1-61208-873-0

Designing for analytic actionability: Temporality and plurality as strategies for human-centered learning analytics

Yeonji Jung, Juan Pablo Sarmiento, Alyssa Friend Wise
New York University
yeonji.jung@nyu.edu, jps651@nyu.edu, alyssa.wise@nyu.edu

ABSTRACT: Analytic actionability refers to the ways that information provided to students and educators affects what they do in learning environments and is a pressing issue for learning analytics to impact practice. As a way to achieve better actionability in our tools and their implementation, this poster describes the strategies of temporality and plurality to intentionally encourage attention to actionability during human-centered processes of learning analytics design. Presented in the context of a design process to create student-facing analytics for collaborative annotation, *Learning Cycle* and *Persona-Based Storyboarding* activities are described as two concrete instantiations of these strategies.

Keywords: Analytic actionability, human-centered learning analytics, participatory design

1 INTRODUCTION

Analytic actionability refers to the ways that information provided to students and educators affects what they do in learning environments. Actionability is a pressing issue for learning analytics to make an impact on practice (Dimitriadis et al., 2021). Despite the widespread development and adoption of analytic tools, common challenges have been reported for both teachers and students in drawing insights from analytic data and translating awareness of learning status into actions for improvement (Dimitriadis et al., 2021). This is a critical issue from the perspective of human-centered learning analytics (Buckingham Shum et al., 2019) and one which highlights the importance of the socio-technical nature of analytics use, as well as the need to plan for the complex webs of educational contexts, routines, and relationships into which an analytic tool will be introduced (van Harmelen & Workman, 2012). One way to plan for actionability is to involve stakeholders in the analytic design process (Buckingham Shum et al., 2019). While simply including the intended users of analytics in their creation may naturally elicit some discussion of factors related to actionability (or lack thereof), such efforts are more effective if attention to actionability is intentionally prompted. However, there has been little documentation in the literature of what a human-centered learning analytics design process specifically targeting actionability might look like. This poster fills that gap by describing how temporality and plurality were used as strategies to encourage attention to actionability in a process to design student-facing analytics for collaborative annotation.

2 DESIGN CONTEXT & PROCESS

The goal of the design process was to create a tool to support university students in finding places to contribute to collaborative annotation activities. Collaborative annotation is an educational activity that engages students in discussing learning materials by making comments and reactions directly in the relevant section of the reading or video (Novak, Razzouk, & Johnson, 2012). In collaborative annotation, shallow interaction (i.e. comments which fail to engage with the material and each other in substantial and important ways) is a crucial issue, which can benefit from external support, such as learning analytics, to help students become more attentive, critical, and reflective in producing deep interaction (Novak et al., 2012). To design a tool to support deep engagement in collaborative annotation, a participatory design approach (Sanders, 2002) was taken to involve students and instructors as co-designers. A total of 11 undergraduate and 13 graduate students with experience using collaborative annotation in their courses, and three instructors who taught those courses were involved in iterative cycles that moved from understanding needs to ideating and specifying solutions and finally testing. For each cycle, a workshop session was conducted with students, a synthesis of the insights was crafted by the researchers, and these findings were reviewed with instructors. This poster focuses on the strategies of temporality and plurality that were used in Workshop 3.

3 TEMPORALITY & PLURALITY AS STRATEGIES FOR ACTIONABILITY

Workshop 3 of the design process focused on ideating and specifying analytic solutions that could help with one of the key challenges students identified in Workshop 1: finding places to meaningfully contribute to a collaborative annotation activity. The strategies emerged through conversation with the course instructors guided by the insight synthesis for Workshop 2 in which they emphasized that the final tool should reflect different kinds of student profiles and engagement patterns that could change over time. Specifically, the instructors suggested that an effective tool would potentially need to allow for varied versions, rather than one-size-fits-all, to meet different needs at different points in the learning cycle. This led to the use of two emergent strategies for design: (1) Temporality (consider the flow of the learning activity) and (2) Plurality (be open to the potential for a need for more than one solution). Each of the strategies shaped how the design activities were conducted and through this had an identifiable impact on the resulting analytics product.

3.1 Temporality

Learning is an activity that unfolds over time, involving changes in dimensions such as knowledge, skills, and modes of participation (Reimann, 2009). In this sense, learners' informational needs and their ability to act upon analytic data can vary at different points in the learning process. Considering temporality and the flow of the learning activity in the process of learning analytics design can provide a critical context for understanding, if, how, and when learners (or educators) are likely to act on information. In the current design effort, attention to temporality was introduced in Workshop 3 through a Learning Cycle activity, designed to elicit information about how students participated in collaborative annotation activities during the course of the week. The Learning Cycle activity used group whiteboarding as a tool for students to share the different ways they and their classmates engaged in collaborative annotation at three different points in time: early-, mid-, end-week when there were no, some, or many comments already made in the tool. This provided the foundation for persona-building storyboarding in a subsequent activity described below.

3.2 Plurality

The basic notion that different learners can have different needs and benefit from support tailored to meet these extends back well over a century (e.g. Thorndike, 1911). Nonetheless, all too often learning analytics are developed as a one-size-fits-all solution, with the presumption that students or instructors can handle this need for variation on their own. In contrast, van Harmelen and Workman (2012) suggested the possibility that more than one analytic tool solution may be needed to meet the diversity of users and diverse contexts with learning analytics needs. In the current design effort, the possibility for more than one analytic solution was engaged with in Workshop 3 through a Persona-Based Storyboarding activity. This activity first characterized the profiles of different personas based on the different activities elicited in the Learning Cycle activity (e.g. learners who prepared their work outside of the annotation platform vs. those who wanted to be influenced by the community). This categorization was then used by the co-design student participants to create storyboards for the different challenges (and emotions) each persona might experience in collaborative annotation at different points in time throughout the week and what information could be useful (and actionable) to help address these challenges. The activity inherently allowed for the possibility that different students might find different information helpful and actionable at different points in time.

4 IMPACT ON THE DESIGN PRODUCT

The emergent strategies of temporality and plurality were clearly found to impact the design of the analytics solution product and its potential actionability. The final product was an email-based recommendation system that offers students suggestions about where they might usefully contribute to a collaborative annotation activity, based on (1) if the recommendation is provided early or late in the assignment week (temporality) and (2) if the student has already participated substantially or not (plurality). All recommendations are driven by an underlying analytics engine that presents different possibilities to different students. Each of the four versions is expected to serve a different role in learning; for example, an email late in the week for those who have not yet posted can act as a reminder and provide an entry point to "crowded canvas" (e.g. do you want to respond to this comment from a peer you haven't connected with recently?) while an email late in the week may support class preparation (e.g. do you want to review this active/controversial comment stream?). By meeting the needs of particular students at specific time points in the learning activity these recommendations are posited to support greater actionability than a one-size-fits-all tool that makes students search for the information useful to them. This proposition will be tested in the coming term by implementing the tool as course activity and examining student reactions and actions to the tool.

REFERENCES

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

Dimitriadis, Y., Martínez-Maldonado, R., & Wiley, K. (2021). Human-centered design principles for actionable learning analytics. *Research on E-Learning and ICT in Education* (pp. 277-296).

Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education, 15*(1), 39-49.

Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of CSCL, 4*(3), 239-257.

Sanders, E. B. N. (2002). From user-centered to participatory design approaches. In Design and the social sciences (pp. 18-25). CRC Press.

Thorndike, E. L. (1911). Individuality. New York: Houghton, Mifflin.

Van Harmelen, M., & Workman, D. (2012). Analytics for learning and teaching. CETIS Analytics Series, 1(3), 1-40.

Student-facing Learning Analytics for Data Literacy: Findings from an Integrative Review

Nora Patricia Hernández López, Xiao Hu
The University of Hong Kong
noraphl@connect.hku.hk, xiaohu@hku.hk

ABSTRACT: The vast amount of data generated by digital interactions has made evident the need for data knowledge and skills in every field, education being no exception. Data from teaching and learning are collected and analysed– for offering information that benefits students' outcomes and teachers' performance, among others. For maximising the potential benefits offered by the analysis of learning data, it is necessary that the information can be retrieved by the relevant stakeholders so that it can be converted into actions that improve learning. Such task requires knowledge and self-efficacy with data, i.e., the complete skillset offered by data literacy. We are interested specifically on how students can benefit from the use of student-facing learning analytics to improve their data literacy skills. By conducting an integrative review of the literature, we identify novel and promising practices that benefit students' understanding of how data is collected, processed, and presented in an authentic learning environment. We discuss the characteristics of such learning designs and the opportunities for future research.

Keywords: Data literacy, student-facing learning analytics, literature review

1 INTRODUCTION

Data literacy can be understood as the ability to reason, communicate, and make decisions with data, and it has been widely recognised as one of the essential competences for daily interactions (Papamitsiou et al., 2021). In the context of education, data from teaching and learning is collected and analysed with the purpose of understanding learning and improving educational practices. This is the field of learning analytics (LA), where research recognises two main skills demanded for the success of the LA-enabled classroom, namely 1) ability for interpreting data, or data literacy; and 2) ability for informing pedagogic practice based on information provided by the data (Papamitsiou et al., 2021). LA tools specifically targeted for students, or student-facing LA (SFLA), track students' learning behaviours from online learning environments and provide feedback back to students to improve students' autonomy over their learning. However, this has been found to be overwhelming for most students (van Leeuwen et al., 2022), and mediated by a variety of cultural and environmental factors, and students' own DL skills (Xing & Wang, 2021). Therefore, we believe that a crucial step into promoting the effective use of SLFA is improving students' data literacy. In this study, we aim to review how students' DL can benefit from interactions with LA or SFLA. We conducted an integrative literature review (Torraco, 2016) to synthesise representative perspectives of emerging discussions and practices at the intersection of these topics. Thus, the research question guiding the review is: how has students' data literacy skills have been addressed in the context of LA or SFLA? In the following sections we present the method executed for conducting the review and the preliminary findings.

2 METHOD

To investigate the relationship between data literacy (DL) and student-facing learning analytics (SFLA), we conducted an integrative literature review (Torraco, 2016), which is best suited for an early conceptualisation of topics with rapid and dynamic growth. We searched three major databases: Web of Science, Scopus, and the Association for Computing Machine Digital Library (ACM-DL). The search strategy included keywords such as “data literacy,” “learning analytics,” and “student-facing learning analytics;” and the query matched title, abstract, and keywords. No restrictions were applied on publication year or language. The search yielded a total of 129 records, and after deleting duplicates, 85 were left~~.~~ After excluding 6 conference proceedings and workshops, 10 review articles, 1 article in Spanish, and 1 not in the field of LA, 67 papers were screened based on title and abstract to assess their relevance. Relevance of each article was assessed based on their inclusion of DL within the context of LA/SFLA, or vice versa. 41 excluded students from the discussion, but addressed other stakeholders, e.g., teachers. From the reviewed articles, we identified 13 that exclusively address data literacy (DL) for students in the context of learning analytics (LA) or student-facing LA (SFLA). Based on the findings of these studies, we present the most relevant findings.

3 PRELIMINARY RESULTS

The critical need for data literacy (DL) to grow the development and adoption of learning analytics (LA) systems across different educational levels and stakeholders was identified in 13 of the articles reviewed. Four of these articles discussed the need of DL among students, all of them in higher education. However, discussion is limited to identifying such need, and do not include more specific strategies aimed at using LA-mediated learning for strengthening DL. Quantitative approaches to assess the possible factors that mediate LA usage remark that university students' perceived data autonomy, digital identity, and reflectiveness is significantly mediated by students’ data literacy (Xing and Wang, 2021).

We identified two student-facing LA (SFLA) implementations that resemble the students’ practices when trying to beat the system but that benefit their understanding of how data is collected, processed, and analysed. Pedagogic designs with writing analytics tools take the lead in engaging students in critical practices with SFLA. In the work proposed by Shibani et al. (2022), automatic assessment is provided by the system to students’ essays. Then, students must answer whether they agree on the output, argue why, and adjust their writing if necessary. Similarly, Kitto et al. (2018) engage students in active learning and critique about a machine learning (ML) model that classifies students’ comments in an online forum. They invite students to challenge the output of the system, which is later used for improving accuracy, while motivating them to increase their participation in the online discussion. While both approaches are effective for engaging students in a critical awareness for using the LA tools, the key difference between the two is the closed feedback loop and support provided for students to understand the role of data including production, collection, and analysis emphasised by Kitto et al. (2018).

We distinguish two driving factors in the interventions described: 1) the notion ML models behind LA could be wrong, and 2) learning designs that promote critical engagement with LA systems. By offering opportunities to learn with SFLA that do not blindly evaluate, rank, or categorise students, DL skills are learned in an authentic environment as students create the data and reflect on them. As if they were trying to game the system, students are required to learn how the system process data to produce its output and are encouraged to beat it. DL becomes an essential learning outcome that empowers learners to be critically aware and escape the mindless repetition of patterns as they must reflect on what could be incorrect and how to make it better. This critique process also benefits disciplinary learning outcomes by requiring an active student participation to generate the data that fuels the SFLA system.

4 CONCLUDING REMARKS

Results of this integrative review have shown the interdependence of data literacy (DL) and learning analytics (LA). While this relationship has been dominated by discussions about the need for data literacy for implementation and adoption of LA, we identified practices with SFLA that improve DL among students. Despite this intersection being not widely studied yet, current examples offer innovative and promising practises for further study. Future empirical interventions should aim to measure the effect that SFLA have over students DL skills to understand how students learn about data in authentic learning environments. Designs that serve this purpose should have solid theoretical and pedagogical foundations that allow critical engagement while developing SFLA systems that protect students' data uses, privacy and ownership.

REFERENCES

Kitto, K., Buckingham Shum, S., & Gibson, A. (2018). Embracing Imperfection in Learning Analytics. Proceedings of the 8th International Conference on Learning Analytics and Knowledge, 451–460. https://doi.org/10.1145/3170358.3170413

Papamitsiou, Z., Filippakis, M. E., Poulou, M., Sampson, D., Ifenthaler, D., & Giannakos, M. (2021). Towards an Educational Data Literacy Framework: Enhancing the Profiles of Instructional Designers and e-Tutors of Online and Blended Courses with New Competences. Smart Learning Environments, 8. https://doi.org/10.1186/s40561-021-00163-w

Shibani, A., Knight, S., & Buckingham Shum, S. (2022). Questioning Learning Analytics? Cultivating Critical Engagement as Student Automated Feedback Literacy. LAK22: 12th International Learning Analytics and Knowledge Conference, 326–335. https://doi.org/10.1145/3506860.3506912

Torraco, R. J. (2016). Writing Integrative Literature Reviews: Using the Past and Present to Explore the Future. Human Resource Development Review, 15(4), 404–428. https://doi.org/10.1177/1534484316671606

van Leeuwen, A., Wise, A. F., & Teasley, S. D. (2022). Teacher and Student Facing Learning Analytics. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), The Handbook of Learning Analytics (2nd ed., pp. 130–140). SoLAR. https://www.solaresearch.org/publications/hla-22/hla22-chapter13/

Xing, W., & Wang, X. (2021). Understanding students' effective use of data in the age of big data in higher education. Behaviour and Information Technology. https://doi.org/10.1080/0144929X.2021.1936176

iLearn Insights: supporting educators to re-engage and motivate students in digital learning

Author: Shamim Joarder
Macquarie University, Australia
Shamim.joarder@mq.edu.au

ABSTRACT: iLearn Insights is an online analytical application that helps educators address the challenge of engaging, motivating and supporting students by identifying who is engaging, triggering a range of personalised communications to re-engage students with their unit (subject), thus allowing early intervention and ultimately improved student retention. By providing data-driven insights into student performance, engagement, and behavior, iLearn Insights can help educators and administrators identify areas for improvement in the delivery of the unit and make data-driven decisions to improve student outcomes.

Keywords: iLearn Insights, Learning Analytics, Student engagement, Learning and Teaching

1 CONTEXT OF USE

iLearn Insights anaylses student learning data such as unit access patterns, forum activity, media views, learning activity submissions and grades allowing educators to easily visualise student activity and engagement in their unit. Educators can then take action using automated or targeted communication to commend high-achieving students, offer additional assistance to lower performing students or recapture disengaged students. iLearn Inisights visulaisations are also available across a course, department or faculty level to aid more strategic decision making.

2 UNIT ENGAGEMENT

Overall unit engagement is calculated by analysing a range of variables such as log in activity, assignments submitted on time, level of grades achieved, and forum participation. Educators select the variables relevant for their unit/subject to calculate the engagement score. The default setting uses data on student access to the unit within the last 7 days and on-time submissions.

3 EVALUATION OF THE SYSTEM

The effectiveness of sending timely and targeted email communications to students is evidenced by the percentage of students who re-engage with the unit, submit an assignment, or access a learning resource within 24-48 hours after receiving a personalised email through iLearn insights. User driven enhancements ensure iLearn Insights continues to evolve and meet the needs of educators to support students.

4 HOW ILEARN INSIGHTS WORKS

The video is available at: https://youtu.be/KKfCqLTWDvI

Demo of GrouPer: Group-based Personalization Application

Tanya Nazaretsky, Yael Feldman-Maggor, Giora Alexandron
Weizmann Institute of Science
{tanya.nazaretsky, yael.feldman-maggor, giora.alexandron}@weizmann.ac.il

ABSTRACT: This demo presents a walkthrough of GrouPer (**Grou**p-based **Per**sonalization), an AI-for-Teacher tool that supports personalized instruction in blended learning K-12 environments. GrouPer combines machine learning algorithms, rich student response data, and expert knowledge to perform a multidimensional analysis of student responses. Based on this analysis, GrouPer discovers knowledge profiles and assign students according to their interactive assessment results. Unlike a teacher, GrouPer has access to 'big data' (e.g., the entire student responses to a particular interactive assessment instrument, meta-data on the corresponding items, etc.) and has the computational power to process these data in real-time (e.g., through K-Means clustering). So, it can promptly discover response patterns that may not be identifiable in 'small' data and highlight to teachers student knowledge profiles that they may not see or be aware of looking into their own students' data only. Then, the resulting assignments of students to the knowledge profiles (aka groups) are presented in an interactive dashboard enabling teachers to examine each group's performance and assign learning activities adapted to the needs of students based on their strengths and weaknesses in terms of knowledge skills and competencies. GrouPer does not decide on the follow-up activities. On the contrary, it allows teachers to craft their own learning sequences based on the presented analysis (e.g., create homogeneous teams from the same group and adapt focused treatment based on missing skills, or create heterogeneous student teams from different groups, so the students can use their completing skills to work together on a follow-up task with similar characteristics).

The tool was conceptualized in a co-design process that involved learning analytics researchers, science educators, teachers, and instructional designers (Nazaretsky et al., 2022). It is integrated into PeTeL (**Pe**rsonalized **Te**aching and **L**earning) – a free online learning environment developed within the Department of Science Teaching at the Weizmann Institute of Science. PeTeL serves more than a thousand science teachers from schools with varied socio-demographic profiles. In the last year, it was piloted by a hundred physics and chemistry teachers who used its insights. We are now enhancing the system to collect the teachers' creative follow-up activities and map them to the knowledge profiles that they address as the basis for a social recommender system that we are developing. Another active direction of our research is improving GrouPer's usability by introducing Explainable AI methods and means (e.g., Feature importance, SHAP, LIME, etc.) into the system for the automated creation of semantically meaningful descriptions of the resulting knowledge profiles.

Keywords: Personalized Instruction; Blended Learning; Teacher Dashboards; Participatory Design; Learning Analytics; Explainable AI

REFERENCES

Nazaretsky, T., Bar, C., Walter, M., & Alexandron, G. (2022). Empowering Teachers with AI: Co-Designing a Learning Analytics Tool for Personalized Instruction in the Science Classroom. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 1–12.

Video: https://youtu.be/TxyR0fjW9YU

The iSAT Collaboration Analytics Pipeline

John Weatherley[1], Rachel Dickler[1], Peter W. Foltz[1], Aravind B. Srinivas[1], Samuel Pugh[1], Nikhil Krishnaswamy[2], Jacob Whitehill[3], Michal Bodzianowski[1], Maggie Perkoff[1], Rosy Southwell[1], Jeff Bush[1], Michael Chang[4], Leanne Hirschfield[1], Daeja Showers[6], Ananya Ganesh[1], Zeqian Li[3], Elita Danilyuk[2], Xinlu He[3], Ibrahim Khalil Khebour[2], Indrani Dey[5], Sidney K. D'Mello[1]

[1]University of Colorado, Boulder
[2]Colorado State University
[3]Worcester Polytechnic Institute
[4]University of California, Berkeley
[5]University of Wisconsin, Madison
[6]Brandeis University
Peter.foltz@colorado.edu

ABSTRACT: The NSF AI Institute for Student-AI Teaming (iSAT) is an interdisciplinary community dedicated to transforming classrooms into more effective, engaging, and equitable learning environments through developing interactive AI partners to support collaboration in the classroom. We have developed and are using a data analysis pipeline for analyzing multimodal interaction data for small group collaborations. The pipeline supports real-time collection of student interactions and integration of AI-based analyses for use by interactive partners to support teachers and students in collaborative activities.

Keywords: Collaboration, Collaborative Problem Solving, Collaboration Analytics, AI, Data Analysis Architectures, Natural Language Processing.

1 DEMONSTRATION OF THE COLLABORATIVE ANALYTICS PIPELINE

The iSAT Collaborative Analytics Pipeline demonstrates an architecture in which multiple streams of classroom speech and video is captured and stored into our data repository. Then multi-faceted analyses are performed on the data using a variety of AI-based models including automated speech recognition, speaker identification, discourse and content analysis, detection of on/off task and topic language, use of collaborative problem-solving skills like constructing shared knowledge and negotiating, and use of academic productive talk. The resulting analyses are stored back in the data repository which notifies the interactive partners to display information or provide feedback to teachers, students, or researchers. The pipeline serves as a foundational data analytic service which allows flexible collection of multimodal classroom data and incorporation and testing of novel AI-based techniques for assessing collaboration. It further allows the development and testing of new user interfaces of AI-based partners to develop effective interactions with students and teachers in the classroom

We provide a video to demonstrate the collaborative analytics pipeline at:
https://www.colorado.edu/research/ai-institute/videos

CoTrack: A Multimodal Learning Analytics tool to guide teachers during collaborative learning activities with intervention suggestions in classroom

Pankaj Chejara, Reet Kasepalu, Luis P. Prieto, María Jesús Rodríguez-Triana, Adolfo Ruiz-Calleja
Tallinn University
pankajch@tlu.ee, reetkase@tlu.ee, lprisan@tlu.ee, mjrt@tlu.ee, adolfo@tlu.ee

ABSTRACT: Teachers are often expected to monitor and support students during collaborative learning activities in the classroom which becomes extremely difficult as the number of groups increases. Multimodal Learning Analytics researchers have developed tools to support teachers. However, these tools mainly focus on monitoring tasks and lack the actionability aspect (i.e., what teacher should do when collaboration quality is detected as Low). This paper presents CoTrack, a Multimodal Learning Analytics tool to guide teachers with monitoring and potential interventions strategies to improve quality of collaboration. Furthermore, CoTrack also caters to the needs of researchers in terms of providing pre-processed multimodal data in CSV format for their research. CoTrack is a web-based application with an easy to use interface that allows teachers (and researchers) to create collaborative learning activities with multimodal data tracking. This tool also has a prototypical dashboard consisting of predictive analytics. This dashboard visualizes participants' contributions in terms of speaking and writing, and also shows predicted levels of collaboration quality with theory-informed intervention strategies to support students (Kasepalu et al., 2022). The dashboard has been codesigned together with teachers in two iterations involving a total of 58 teachers (Chejara et al., 2022). With CoTrack, we envision bringing MMLA research's benefits to the practice.

Keywords: Multimodal Learning Analytics, Collaborative Learning, Guiding Dashboard

Demonstration Movie:https://www.youtube.com/watch?v=l-B2hXGRvek

REFERENCES

Kasepalu, R., Prieto, LP., Ley, T. & Chejara, P. (2022). Teacher Artificial Intelligence-Supported Pedagogical Actions in Collaborative Learning Coregulation: A Wizard-of-Oz Study. *Front. Educ.* 7:736194. doi: 10.3389/feduc.2022.736194.

Chejara, P., Kasepalu, R., Prieto, L.P., Rodríguez-Triana, M. J., & Ruiz-Calleja, A., (2022). Co-Designing a Multimodal Dashboard for Collaborative Analytics. Poster presented at: *The 15th International Conference on Computer-Supported Collaborative Learning (CSCL) 2022,* Online.

Development of a Learning Analytics Dashboard to Complement the Training of Procedural Skills in Physiotherapy Education

Rocío Hernández, Ignacio Villagrán, Javiera Fuentes, Loreto Larrondo, Elisa Margozzini, María Trinidad Hurtado, Zoe Iriarte, Julián Varas, Andrés Neyem, Isabel Hilliger
Pontificia Universidad Católica de Chile
{rbhernandez, invillagran, jfuentesi, lorelarrondov, elisa.margozzini, t.hurtado, ziriarte, jevaras}@uc.cl, {aneyem, ihillige}@ing.puc.cl

Constanza Miranda
Johns Hopkins University
constanzamiranda@jhu.edu

ABSTRACT: The COVID-19 pandemic stressed the need to use online learning environments to continue procedural skills training. Given its psychomotor domain, learning procedural skills requires deliberate practice, along with opportunities for repetition, reflection, and improvement; which can be approached from a self-regulated learning (SRL) perspective that allows the learner to have control over the actions they perform in a structured environment and be active in their learning experience in a metacognitive, behavioral and motivational way. To meet this need, we created a learning analytics dashboard (LAD) that seeks to complement a web-based platform for remote teaching of procedural skills through videos. The design of the LAD followed a design-based research approach guided by lessons learned from Human-Centered Learning Analytics, and its indicators are based on SRL strategies like planning (e.g., steps to be completed), self-evaluation (e.g., feedback inputs), elaboration (e.g., instructional material), and help-seeking (e.g., collaborative forum). Likewise, the LAD was designed to make visible the feedback inputs of the different stages that the students had to complete, along with the instructional material of the next step and a text box for self-assessment so that they could reflect and incorporate the feedback. During the first cycle, an initial version was developed with 8 indicators used in LADs with operation and cognitive-condition approaches, like overall performance, performance by stage and by item; performance line over time, feedbacks by stage, self-evaluation mailbox, among others. This was tested through cognitive walkthroughs with 8 students. Findings triggered the promulgation of a second version, in which specific data visualizations were changed to show student performance on each rubric item, along with modifying usability aspects such as the size of titles and legends, and extending the time taken for an animation to display new information. This second version was implemented in 6 courses of a physical therapy degree program (see Demo video). At the end of the academic period, SRL and sensemaking surveys were applied; and according to the majority of the respondents (74,7%), the results showed that the indicators valued as "extremely relevant" were those related to the student's final grade and their performance in comparison with the maximum possible.

Keywords: Learning analytics dashboard; self-regulated learning; procedural skills; feedback.

1 Demo video

The video is available at https://drive.google.com/file/d/demo.

Data and Evidence-Informed Educational Practice with LEAF System

Hiroaki Ogata, Brendan Flaganan, Rwitajit Majumdar and Izumi Horikoshi
Kyoto University
hiroakiogata@gmail.com

ABSTRACT: The LEAF system is a Learning Analytics infrastructure that supports the collection, analysis, and utilization of learning logs. LEAF system consists of a Learning Management System (LMS), an eBook reader (BookRoll), Learning Record Store (LRS), and a Learning Analytics tool (Log Palette). BookRoll works as a behavior sensor and records student log data related to reading interactions and artifacts generated within the content. Log Palette analyzes and visualizes the log data obtained from BookRoll. The log data can be further used for interactive lectures, reflection, recommendations, and class improvement. LEAF system has been used in over 120 educational institutions, from elementary to higher education, in eight countries and regions. Our goal is to scientifically analyze those data, support each teacher and student, and transform "education based on intuition and experience" into "education based on data and evidence." In our video and demo, we will introduce the followings: (1) LEAF system functions to support data-and-evidence-based education, and (2) the use cases and comments of teachers who have been utilizing LEAF system for their class activities at the K-12 level.

Keywords: Data-and-Evidence-Informed Education, K-12, Higher Education, LEAF system

Video URL:
https://drive.google.com/file/d/12OsuIHTlVtdO3eBJm8XLZ59AuBJy2-rD/view?usp=share_link

Interpretable Code-Informed Learning Analytics for CS Education

Yang Shi
North Carolina State University
yshi26@ncsu.edu

ABSTRACT: The development of deep learning technologies brings opportunities to computer science education (CSEd). Recent CSEd datasets provide actual code submissions that could potentially offer more information on students' learning for learning analytics, but adding the information needs special considerations. My research focuses on leveraging these code submissions to inform learning analytics. Specifically, it addresses two problems: 1) finding and detecting a meaningful representation of knowledge components (or skills) in students' programming and 2) using them to provide formative feedback. Previous works have defined or discovered skills in CSEd, but these skills do not follow certain properties of learning. My research proposes to incorporate these properties into data-driven model design, and improves the knowledge components so that they are consistent with theoretical properties, and meanwhile also provide better interpretability than typical deep learning models. Following this, my second project will focus on providing personalized support in student learning.

Keywords: Computer Science Education, Educational Data Mining, Code Analysis, Interpretable Deep Learning

1 INTRODUCTION

Computer science (CS) education faces unprecedented opportunities and challenges when meeting the fast development of deep learning (LeCun et al., 2015) technologies these years. With these tools empowered by big data analysis, teachers and students are expected to enjoy better learning and teaching environments, while researchers will have more detailed, accurate, and efficient learning analytics. For example, tools based on deep learning can be used for tracing students' learning progress of knowledge (Piech et al., 2015), and thus students can get formative feedback (Shute, 2008) on problems that they are predicted to be wrong and reflect on the knowledge gaps in certain concepts. Teachers can also use data analysis to learn their students' progress, and inform their pedagogical decisions (Ju et al. 2019). However, using deep learning to analyze CS education data has additional challenges.

Many challenges reside in learning analytics (LA) or educational data mining (EDM) when using deep learning technologies, even without domain-specific data. For example, deep learning models are known as black boxes (Castelvecchi, 2016) despite their advantages in task performance. The uninterpretable models can harm the trustworthiness of the conclusions the model provides, especially in the educational domain (Cohausz, 2022). Recent research in CSEd has been collecting submitted code data along with the typical performance data (Leinonen, 2022), and these datasets add more opportunities to derive interpretability from code data. This adds more information to the models but needs careful design. Furthermore, even with an interpretable model, leveraging interpretations and propagating intelligent feedback to students are still unaddressed challenges. My research focuses on providing solutions to these challenges. In the first project, I will design a deep neural network to leverage code information for knowledge component (KC) (Koedinger et al., 2012a) detection. The model learning process is guided by learning theory. While the model is designed as a performance predictor, trying to take historical submissions and performances and use them to predict future performance, the middle layer of the model can be interpreted as a representation of students' practiced skills. Learning theory is designed as a constraint on the middle layer, so that it follows properties such as the power law of practice (Cen et al., 2006). The middle layer can be tracked to actual code input, attributing the coding concepts students are capable of or incapable of. The second project addresses the usage of the detected KCs. For a student with detected knowledge gaps on certain concepts, code examples (Brandt et al., 2009) will be shown to students as formative feedback.

2 RELATED WORK

Student Modeling in CS Education: There are many works around the topic of student modeling for CSEd. As more student code datasets are made accessible to the public, student modeling in CS education has seen fast development in recent years. It is a special domain, as programming code is rich in information, but also often bears much noise. Some recent works have been manually analyzing small to medium size student code datasets for student modeling. For example, Paul et al. have created instruments for manually detecting misconceptions from student code (Paul & Vahrenhold., 2013). Davies et al. manually analyzed programming trace data and constructed a library of knowledge gaps, showing that programming logs would reveal more misconceptions than single submission data (Davies et al., 2015). While expert analysis is more detailed and offers insights, they are also expensive, time-consuming, and often vulnerable to "expert's blind spots" (Nathan et al., 2003). Data-driven models, on the other hand, don't have these disadvantages, however, often suffer from low performance when evaluated on expert-assigned labels. They can achieve good performance with little help from experts. For example, Marwan et al. developed a subgoal detector in students' programming data and showed that adding a few expert constraints helped the detection outperform data-driven models (Marwan et al., 2021). However, this would still need explicit expert help. Deep learning models can be leveraged as an alternative way to improve performance, but they suffer from low available labeled datasets, and limited interpretability and thus caused low trustworthiness in the modeling results. My previous work used a semi-supervised learning method to address the limited label issue (Shi et al., 2021b), and my proposed work looks into how to add more interpretability to deep learning models, while they provide accurate modeling results.

KC Discovery and Refinement: KCs in my proposed work refer to programming skills that students learned and should be observable from the programming dataset in the CS domain. Traditionally, KCs are manually analyzed and defined by cognitive task analysis (CTA, Clark et al., 2008), but they are time-consuming and prone to experts' blind spots. As KCs are mapped to problems by Q-matrices (Barnes, 2005), many methods are developed to refine these Q-matrices or to discover new KCs (Cen et al., 2006; Koedinger et al. 2012b). These methods require less human effort, but the resulting Q-matrices can be also less interpretable. For example, one key method is to use learning curves for refinement. Cen et al. proposed that the error rate of students' practice on certain KCs should follow the power law of practice (Cen et al., 2006). They calculated the score of fitting to evaluate the quality of KCs specified by a Q-matrix. Their method is further extended to an automatic search algorithm for Q-matrix refinement, using A* search (Koedinger et al. 2012b). However, these methods do not have domain information involved and thus cannot attribute errors to the exact skills in students' code. In the computing education domain, Rivers et al. (2016) proposed to use of abstract syntax tree (AST) nodes to represent KCs and found that some KCs do not fit the expected properties (e.g. the power law). My proposed method is inspired by but differs from theirs. It leverages programming code to inform the discovery process and does not require the initialization of the Q-matrix. Instead, it uses a randomly initialized matrix for modeling the KC-problem relationships, updated in the deep learning model training process.

Deep Code Learning: Deep learning has evolved to solve more complex problems in more disciplines recently, including program code analysis. There are many models designed to process code in recent years. Earlier models (e.g. GRU, Bi-LSTM, etc.) only treat code as a sequence of tokens (Reyes et al., 2016), leaving tree-structured information of code unused. Later models such as code2vec (Alon et al., 2019) and ASTNN (Zhang et al., 2019) take this structural information into account, and achieved higher performance in tasks such as function name classification and code summary. In an educational context, recent papers have discussed the usage of such models in student performance prediction (Mao et al., 2021; Shi et al., 2022), bug detection (Shi et al., 2021a; Shi et al. 2021b), code classification (Fein et al., 2021), etc. tasks. However, one key issue still remains in these applications, and that is the lack of interpretability of these models. This undermines the trust of models, as the prediction process of models cannot be explained to teachers or students. For example, in performance prediction, these models only make the predictions when students fail without telling why they will fail on certain problems. Moreover, as recent research suggests, data-centric methods are prone to biases that lead to inequity (Ocumpaugh et al. 2014), which further requires interpretability for the proper application of these models. My work proposes to use learning theory to guide the model training process, introducing a possible way to interpret the middle layer vectors as the learning progress of KCs and thus add more trustworthiness to the prediction outcomes produced by deep learning models.

Intelligent Feedback: There are many previous works using intelligent feedback in tutoring systems. Research shows that formative feedback helps students' learning, but it needs to be specific, on-time, corrective, and positive (Shute, 2008). Due to these specific properties of feedback, it is difficult to provide automated feedback to students and meet these requirements at the same time. Prior works have been only fulfilling partial properties. For instance, autograders can only provide corrective feedback, namely only giving feedback about which test cases the students failed or passed. More importantly, it does not provide feedback on which exact KC the student is not able to achieve. My research entails a solution to provide feedback not only on the problem correctness, and the corresponding test cases, and also on the actual knowledge gaps in students' submitted programs.

3 RESEARCH QUESTIONS

The limitations of code-informed learning analytics inspired my proposed projects. An overview of the projects is shown in Figure 1. Specifically, these two projects are focused on solving these two research questions:

RQ1: How to enable interpretable deep code learning for performance predictions, guided by learning theory? This will be addressed in Project 1.

RQ2: How to use the interpreted skills to create formative and intelligent feedback to propagate to students? This will be addressed in Project 2.

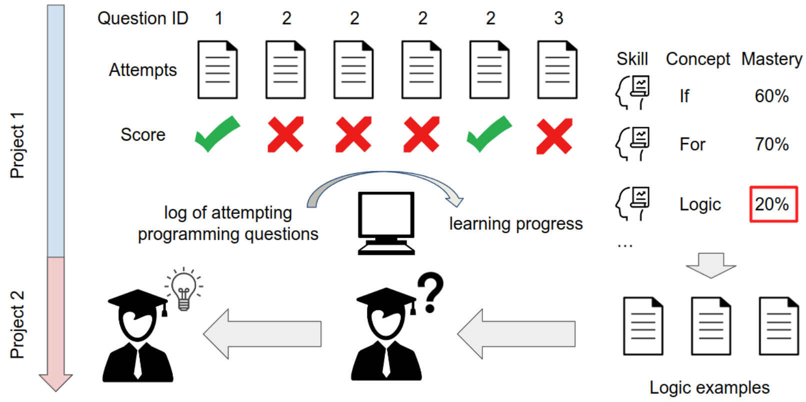


Figure 1: Timeline and concepts of proposed projects

4 Proposed Solution

In Project 1, I will introduce educational context into deep learning models and add interpretability to the models. Originally, the models are only capable of making predictions on students' next problem performance, with massive middle layer parameters uninterpretable. I propose to leverage the layer before prediction as the error rate of certain KCs. To this end, constraints will be needed to guide the learning process of not only the prediction output, but also this layer. Since KCs are often evaluated using learning curves, I propose to hypothesize ideal learning curves for KCs, and calculate the fitness of the layer to the expected learning curve for every KCs. This fitness is added to the loss function to guide the learning of the model. This process will lower the weight of prediction loss, and thus will produce less accurate predictions, and thus hyperparameters should be carefully tuned to find a balance between the fitness of the learning curves for the candidate KCs, and the prediction accuracy of students' performance. The evaluation of the model should also be considered carefully. The candidate KCs should have the property that generates ideal learning curves, but this cannot guarantee that they will represent actual concepts in programming code. Over different problems, they may inconsistently represent code components as well. The values of the candidate KCs should also be examined to check the actual code, and evaluated that 1) if they are consistent within the problem, representing that they are corresponding to certain code components, and 2) if they are consistent throughout different problems, showing if they can be used for unseen problems. More constraints and special design considerations should be added if the concepts represented by the KCs are inconsistent. There are three phases for me to solve this problem: The first one is to create a data-driven model to detect the KC performance from labeled correct submissions. In the second phase, I will use the model and infer the KC performance on incorrect submissions and detect which KCs students don't practice well. Finally, I will incorporate the learning curve analysis into the model and evaluate whether this addition will improve the performance of the KC discovery problem.

Project 2 will be a succeeding project from Project 1. The artifact of Project 1 is a model that given any code submissions, will predict the correctness of students' code submissions, and will show the scores of every KCs and the corresponding code concepts. Low scores on certain concepts can be seen as a lack of skill, and examples will be given to students to learn specific ways to correct their code. A set of examples from various problems will be also processed by the model, and corresponding scores are collected. Students will receive code examples with high scores on the skills that their code achieved a low score, and thus address the specific concepts in their code.

5 PROGRESS AND FUTURE WORK

During my first 3 years of Ph.D. study, I have published 5 papers in related areas (3 of them are first-authored papers). In LAK'21, I published a paper about how to integrate the code2vec model into CS education, using the model for auto-grading and at the same time discovering student errors from clusters created by the middle layer information of the model. In EDM'21, I presented my paper about using a semi-supervised learning method for bug detection, showing that deep learning models may perform better even without a large dataset, if more unlabeled data is available. In EDM'22, my paper about Code-DKT introduced how to add code information to the DKT model for performance prediction, but since skill-problem mapping is unclear in the CSEd domain. I will submit a paper to AIED'23 to solve address the first research question. The proposed projects lay the foundation for automated teaching and learning support in CS education. While the technical side still has a lot to improve (e.g. performance prediction accuracy, methods to improve code representations in a deep neural model, etc.) as possible future work, there are multiple aspects alongside these projects that are awaiting. The evaluation of interpretability and the quality of the interpretation are yet to be evaluated. Although we could use the additive factors model (AFM) to evaluate the Q-matrices generated, there's a lack of direct evaluation of how the concepts are corresponding to actual knowledge taught in class. For project 2, future evaluations on how students improve using specific feedback are also required. More generally, there is a lot to work on in this field and I plan to work on them after my Ph.D. study.

REFERENCES

Alon, U., Zilberstein, M., Levy, O., & Yahav, E. (2019). code2vec: Learning distributed representations of code. *In POPL 2019.*

Barnes, T. (2005, July). The Q-matrix method: Mining student response data for knowledge. In *AAAI 2005 educational data mining workshop* (pp. 1-8). AAAI Press, Pittsburgh, PA, USA.

Brandt, J., Guo, P. J., Lewenstein, J., Dontcheva, M., & Klemmer, S. R. (2009, April). Two studies of opportunistic programming: interleaving web foraging, learning, and writing code. In *CHI 2009*.

Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, *538*(7623), 20.

Cen, H., Koedinger, K., & Junker, B. (2006, June). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International conference on intelligent tutoring systems* (pp. 164-175). Springer, Berlin, Heidelberg.

Clark, R. E., Feldon, D. F., van Merriënboer, J. J., Yates, K. A., & Early, S. (2008). Cognitive task analysis. In *Handbook of research on educational communications and technology* (pp. 577-593). Routledge.

Cohausz, L. (2022). Towards real interpretability of student success prediction combining methods of XAI and social science. In *EDM 2022*.

Davies, R., Nyland, R., Chapman, J., & Allen, G. (2015, March). Using transaction-level data to diagnose knowledge gaps and misconceptions. In *LAK 2015*.

Fein, B., Graßl, I., Beck, F., & Fraser, G. (2022). An Evaluation of code2vec Embeddings for Scratch. *In EDM 2022*.

Ju, S., Zhou, G., Azizsoltani, H., Barnes, T., Chi, M. (2019). Identify critical pedagogical decisions through adversarial deep reinforcement learning. *In EDM 2019*.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science*, *36*(5), 757-798.

Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2012). Automated Student Model Improvement. *International Educational Data Mining Society*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Leinonen, J. (2022). Open IDE Action Log Dataset from a CS1 MOOC. In *Proceedings of the 6th Educational Data Mining in Computer Science Education (CSEDM) Workshop.*

Mao, Y., Shi, Y., Marwan, S., Price, T. W., Barnes, T., & Chi, M. (2021). Knowing" When" and" Where": Temporal-ASTNN for Student Learning Progression in Novice Programming Tasks. In *EDM 2021*.

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. In *EDM 2021*.

Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American educational research journal*, *40*(4), 905-928.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, *45*(3), 487-501.

Paul, W., & Vahrenhold, J. (2013, March). Hunting high and low: Instruments to detect misconceptions related to algorithms and data structures. In *SIGCSE TS 2013*.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, *28*.

Reyes, J., Ramírez, D., & Paciello, J. (2016, December). Automatic classification of source code archives by programming language: A deep learning approach. In CSCI *2016*.

Rivers, K., Harpstead, E., & Koedinger, K. R. (2016, September). Learning curve analysis for programming: Which concepts do students struggle with? *In ICER 2016.*

Shi, Y., Shah, K., Wang, W., Marwan, S., Penmetsa, P., & Price, T. (2021, April). Toward semi-automatic misconception discovery using code embeddings. *In LAK 2021*.

Shi, Y., Mao, T., Barnes, T., Chi, M., & Price, T. W. (2021, January). More with less: Exploring how to use deep learning effectively through semi-supervised learning for automatic bug detection in student code. *In EDM 2021*.

Shi, Y., Chi, M., Barnes, T., & Price, T. (2022). Code-DKT: A Code-based Knowledge Tracing Model *In EDM 2022*.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Zhang, J., Wang, X., Zhang, H., Sun, H., Wang, K., & Liu, X. (2019, May). A novel neural source code representation based on abstract syntax tree. In *ICSE 2019.*

ConCICE Support For Online Algebra I Learning: Conceptual Change Oriented Conversational Artificial Intelligence Using Induction, Concretization, and Exemplification

Chenglu Li, PhD Candidate
University of Florida
li.chenglu@ufl.edu

ABSTRACT: Algebra I is a challenging topic due to its demanding nature that disconnects students' prior mathematical knowledge and the new paradigm in algebra. Extensive studies have examined students' misconceptions using a conceptual change approach to support their algebra learning. However, these studies may not scale well in online settings due to their dependence on manual support or independent learning contexts. A promising approach to automatically support students' algebra learning through an organic integration to existing learning environments is conversational artificial intelligence (ConvAI). This study aims to design and develop a mixed ConvAI using both rule-based and generation techniques to support students' online Algebra I learning following the conceptual change framework. Specifically, this study seeks to create a theory-driven, socially responsible, and culturally responsive ConvAI to support students in online algebra learning using induction, concretization, and exemplification. The significance, methods, and current status of this study are discussed in the proposal.

Keywords: conceptual change, conversational AI, natural language generation, math learning, online discourse

1 INTRODUCTION

Algebra I is a challenging topic. Students' passing rate of Algebra I in Florida consistently stayed below 50% from 2015 to 2021, with the most recent result being 37% (Brown, 2021). As the gateway to more advanced mathematical tools (Kieran, 2004) in science, technology, engineering, and math (STEM), Algebra I directly influences students' performance and motivation in STEM learning, career directions, and ultimately national STEM workforce development (Cirino et al., 2019). The demanding nature of algebra mainly resides in the disconnections between students' prior mathematical knowledge and the new paradigm in algebra (Litke, 2020), including abstract reasoning, algebraic literacy, and mathematical structure.

The detachment of prior knowledge from algebra has been extensively studied with students' alternative conceptions using a conceptual change approach. Alternative conceptions, also referred to as misconceptions, are thinking models to understand, interpret, and apply concepts in ways that do not align with the generally accepted explanations; Conceptual change is a theoretical framework envisioning that learning happens as students restructure or replace their existing alternative conceptions when interacting with new experiences (DiSessa & Sherin, 1998; Limón, 2001). Although there are numerous studies examining and addressing students' misconceptions in STEM education through the lens of conceptual change, these studies mainly delivered interventions manually in the forms such as one-on-one human tutoring (e.g., Muldner et al., 2015) and curricula enhanced with scaffolding (e.g., Maharani & Subanji, 2018), or utilized separate software such as Cognitive Tutors for training students (e.g., Booth et al., 2013). In an online learning environment, such manual support and independent contexts might not be scalable and fit to support students in a large-scale.

A promising approach to automatically support students' algebra learning through an organic integration to existing learning environments is conversational artificial intelligence (ConvAI). ConvAI and chatbot are often used interchangeably, which are defined as human-developed software powered by natural language processing (NLP) techniques to respond to human languages (Li et al., 2022). There are two distinct ways of constructing ConvAI. The first is a rule-based agent that requires manual engineering with classical NLP methods, and the other uses automatic data-driven inferences to generate responses. The former extracts keywords, intents, and emotions from students' input, producing responses with predefined templates. The latter utilizes deep neural networks trained with big data to "learn" to respond to student input with human-like texts. Both forms have advantages and downsides. Rule-based ConvAI can easily integrate pedagogical strategies to support learning. However, the lack of variations and finite states in inputs and outputs can compromise students' learning experience. While generation-based ConvAI can conduct open-domain conversations with students, the generated content might not be pedagogical or distracting for learning (Li et al., 2022). Therefore, there seems to be an opportunity to integrate both approaches to build ConvAI organically.

2 PURPOSE STATEMENT

This study aims to design and develop a mixed ConvAI using both rule-based and generation techniques to support students' online Algebra I learning following the conceptual change framework. Using Math Nation as the platform, the study adopts a learning engineering approach for the ConvAI development and its evaluation on the affordances of algebra learning from students who took Algebra I in Math Nation. Math Nation is an online learning environment covering broad topics on math in K-12 settings, which was developed by the Lastinger Center for Learning at the University of Florida and Study Edge. A preview of the ConvAI prototype can be found at https://youtu.be/nz8DfKWxmfQ. I grounded this study with the following research questions (RQs):

1. What are students' acceptance levels with the ConvAI in Math Nation?
2. To what extent does the ConvAI influence students' academic motivation in Math Nation?
3. To what extent does the ConvAI influence students' conceptual change in Math Nation?

3 CONCEPTUAL FRAMEWORK & RELEVANT STUDIES

This study adopts the epistemological standpoint of conceptual change, where knowledge inquiries restructure alternative conceptions to be aligned with publicly acknowledged conceptions when students actively interact with learning environments. Figure 1 demonstrates the conceptual framework derived from the literature review, details of which are discussed in the following sections.

3.1 Alternative Conceptions in Algebra

Alternative conceptions can be cognitive and motivational. Previous studies have shown that students' common alternative conceptions in Algebra conceptual and procedural knowledge include but are not limited to (Bush & Karp, 2013; Egodawatte, 2011): (1) Misunderstanding a variable's role in an

Algebraic structure (e.g., labels, constants, unknowns, etc.); (2) Incorrectly simplifying or evaluating Algebraic expressions or equations using inapplicable prior knowledge; (3) Struggling with finding inverse values in Algebraic equations; and (4) Mistakenly generalizing conditional properties in functions to unconditional ones.

Apart from the alternative cognitive conceptions, a common motivational misconception affecting students' learning is students' mindsets. Students having difficulty learning STEM are often found to have a fixed mindset (Sun et al., 2021). Studies have shown that students' growth-mindset and fixed mindset in STEM learning are significantly related to their learning outcomes (Yeager & Dweck, 2020). A growth mindset is "the belief that personal characteristics, such as intellectual abilities, can be developed, and a fixed mindset is the belief that these characteristics are fixed and unchangeable" (Yeager & Dweck, 2020, p. 1270). The two mindsets can influence students' reactions to and handling of challenges, praises, and success of others, which can further impact students' learning behaviors such as learning goal setting, agenda implementation, and reflection (Campbell et al., 2020).

3.2 Strategies to Address Alternative Conceptions

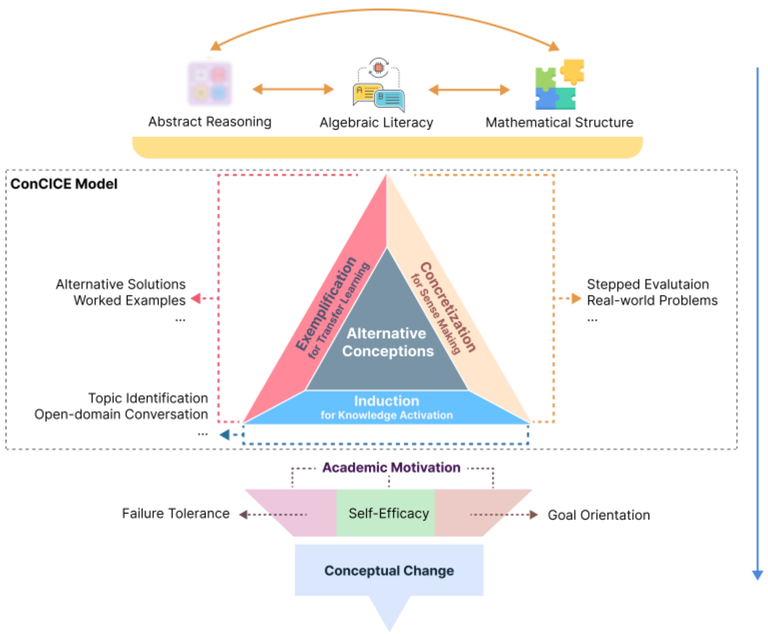


Figure 1. Visual representation of the conceptual framework—conceptual change using induction, concretization, and exemplification (ConCICE) model. Solid arrows indicate influential relationships, while dotted arrows

Studies have shown that three strategies can effectively help students achieve conceptual change in math learning. The first is to **induce** students of prior procedural and conceptual knowledge (e.g., terms in Algebra) embedded in a problem. The process allows students to activate their prior knowledge within or outside a subject to be prepared for problem-solving and to notice the connections and disconnections between prior knowledge and Algebra problems (Asghar et al., 2012; Higgins, 1996). The induction to knowledge activation also supports students' help-seeking behaviors, assisting them with problem recognition and providing potential resources to systematically ask for help. The collaborative procedure of help-seeking creates interactive experiences that potentially help students accommodate existing knowledge. Examples of induction for knowledge activation include identifying learning topics in Algebra problems, providing students with tools to effectively and efficiently connect to prior knowledge, and conducting thinking aloud or collaborative discussions.

Second, **concretizing** Algebra problems can support students in building personal meanings of abstract math problems (Fyfe et al., 2014). The concretization of problems can be procedural or conceptual. Procedurally, concretizing problem solving with stepped evaluations of math problems can offer students opportunities for self-monitoring and reflections, which can lead to moments of conceptual change (van Gog et al., 2020). Conceptually, mathematical concepts can have multiple representations and studies have shown that students could better benefit from more concrete representations such as texts and graphs than mathematical notations. Therefore, connecting abstract variables, expressions, or functions to real- world word problems and visualizing them with graphs can be effective in supporting students to fully understand Algebra problems, triggering conceptual change.

Finally, **exemplification** in Algebra learning can be helpful to clarify goals of problem-solving (e.g., what results are expected), reduce students' cognitive load, develop long-term memory for procedural fluency, and potentially support students to transfer what has been gained in the examples to problem-solving (Kapur, 2014; Ridwan et al., 2021). Strategies of exemplification include providing worked examples to a current problem, demonstrating alternative solutions to help students see the flexibility of Algebraic problem-solving and connections of prior knowledge, and recommending relevant learning resources such as videos and worksheets for further practice and clarifications. Although the aforementioned strategies have been shown to be valuable to support students' conceptual change, their effects are subject to individual differences. The next section discusses students' constructs that can influence the effects of these strategies on Conceptual Change.

3.3 Moderators of Conceptual Change

Studies have shown that students' failure tolerance, achievement goal orientation, and self- efficacy could significantly influence their achievement of conceptual change, three of which are important constructs of students' academic motivation. Students with low levels in these constructs can choose to avoid conflicts in their cognition, ignore their misconceptions in learning, or reject replacing their existing cognitive schema (Potvin, 2021). First, advancement in learning needs exposure to challenging tasks relative to students' abilities. However, such challenges can lead to setbacks and stress in learning. Failure tolerance describes students' attitudes towards setbacks in learning. Students with high failure tolerance tend to better adapt to uncertain, confusing, and frustrating situations in learning. Second, achievement goal orientation demonstrates students' beliefs on what matters in learning, where previous studies showed that mastery orientation could be essential to yielding conceptual change as the orientation is related to students' embracement of challenges (e.g., Kang et al., 2005). Third, students' self-efficacy is closely related to their effectiveness in adopting learning strategies for problem-solving and persistence toward goals (e.g., Simamora & Saragih, 2019), where students with a higher self- efficacy tend to demonstrate more conceptual change (Potvin, 2021).

4 SIGNIFICANCE AND ORIGINALITY

To the best of my knowledge, this study is the first of its kind to examine the affordances of ConvAI using both rule-based and language generation techniques to support online Algebra I learning. Specifically, this study grounded its scaffolding with the ConCICE framework to systematically and effectively provide support for algebra learners. The language generation component is extended based on my prior investigations on natural language generation to generate socially responsible, inspiring, and supportive conversations (see Li & Xing, 2021; Li et al., 2022). The findings of this study can provide instructional design insights, methodological innovations, and pedagogical implications for learning designers, researchers, and instructors. Meanwhile, the developed ConvAI is open sourced (repo1 & repo2) to allow further extensions and reusable integration with other learning platforms.

5 METHODS VIGNETTE

This study will adopt a mixed methods approach using a sequential explanatory design. The iteration conducts a two-week pre-experimental study with pre- and post-test design without a comparison group to measure students' constructs on academic motivation and conceptual change. At the end of the iteration, I will also measure students' acceptance levels with the ConvAI. The pre-experimental study intends to recruit two classes of 8th-grade students (n = 40-50) who are taking Algebra I in Math Nation. Besides quantitative data collected with scales and behavioral logs in the platform, the study also collects students' qualitative data of open-ended questions and discourses with the ConvAI. The iteration starts with a quantitative analysis to understand the affordances of the ConvAI on students' motivation and conceptual change. Qualitative analysis will then be conducted to explain and elaborate on the findings. All constructs (e.g., motivation, conceptual change) will be measured with validated instruments.

There were ethical concerns about applying a technological intervention to middle school students. However, most of them would be addressed through the IRB. One consideration was that the effects of AI on students could be subject to individual differences, which could yield educational equity issues. This study would investigate how individual differences interacted with AI technologies. Implications on the equity examination would be provided in the study to inform researchers and practitioners.

6 CURRENT STATUS

Currently, I have finished the first two chapters of this study to elaborate on its significance, theoretical foundations, and relevant studies. I have also conducted a Delphi study to inquire and synthesize EdTech experts' opinions on the design of the proposed ConvAI. A prototype has been developed based these experts' collective feedback. Two teachers have agreed to participate in the study and help with students recruitment, upon the approval with IRB.

REFERENCES

Asghar, A., Ellington, R., Rice, E., Johnson, F., & Prime, G. M. (2012). Supporting STEM education in secondary science contexts. *Interdisciplinary Journal of Problem-Based Learning, 6*(2), 4.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25*, 24–34.

Brown, D. J. (2021, August 3). Only 37 percent of FL 8th graders passed a crucial statewide math exam. Florida Phoenix. https://floridaphoenix.com/2021/08/03/only-37-percent-of-fl-8th-graders-passed-a-crucial-statewide-math-exam/

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior, 32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

Campbell, A., Craig, T., & Collier-Reed, B. (2020). A framework for using learning theories to inform 'growth mindset' activities. *International Journal of Mathematical Education in Science and Technology, 51*(1), 26–43. https://doi.org/10.1080/0020739X.2018.1562118

Cirino, P. T., Tolar, T. D., & Fuchs, L. S. (2019). Longitudinal algebra prediction for early versus later takers. *The Journal of Educational Research, 112*(2), 179–191.

DiSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10), 1155–1191.

Egodawatte, G. (2011). Secondary school students' misconceptions in algebra. University of Toronto, Toronto.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review, 26*(1), 9–25.

Higgins, E. T. (1996). *Activation: Accessibility, and salience*. Social Psychology: Handbook of Basic Principles, 133–168.

Kang, S., Scharmann, L. C., Noh, T., & Koh, H. (2005). The influence of students' cognitive and motivational variables in respect of cognitive conflict and conceptual change. *International Journal of Science Education, 27*(9), 1037–1058. https://doi.org/10.1080/09500690500038553

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022.

Kieran, C. (2004). Algebraic thinking in the early grades: What is it. *The Mathematics Educator, 8*, 139–151.

Li, C., & Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. *International Journal of Artificial Intelligence in Education, 31*(2), 186-214.

Li, C., Xing, W., & Leite, W. (2022). Building socially responsible conversational agents using big data to support online learning: A case with Algebra Nation. *British Journal of Educational Technology*, *53*(4), 776-803.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*(4), 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Litke, E. (2020). Instructional practice in algebra: Building from existing practices to inform an incremental improvement approach. *Teaching and Teacher Education, 91*, 103030. https://doi.org/10.1016/j.tate.2020.103030

Maharani, I. P., & Subanji, S. (2018). Scaffolding Based on Cognitive Conflict in Correcting the Students' Algebra Errors. *International Electronic Journal of Mathematics Education, 13*(2), 67–74.

Muldner, K., Wixon, M., Rai, D., Burleson, W., Woolf, B., & Arroyo, I. (2015). Exploring the impact of a learning dashboard on student affect. 307–317.

Potvin, P. (2021). Response of science learners to contradicting information: A review of research. *Studies in Science Education*, 1–42.

Ridwan, M. R., Retnawati, H., & Hadi, S. (2021). The Effectiveness of Innovative Learning on Mathematical Problem-Solving Ability: A Meta-Analysis. *International Journal of Research in Education and Science, 7*(3), 910–932.

Simamora, R. E., & Saragih, S. (2019). Improving Students' Mathematical Problem Solving Ability and Self-Efficacy through Guided Discovery Learning in Local Culture Context. *International Electronic Journal of Mathematics Education, 14*(1), 61–72.

Sun, X., Nancekivell, S., Gelman, S. A., & Shah, P. (2021). Growth mindset and academic outcomes: A comparison of US and Chinese students. *Npj Science of Learning, 6*(1), 1–9.

van Gog, T., Hoogerheide, V., & van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review, 32*(4), 1055–1072.

Yeager, D. S., & Dweck, C. S. (20201231). What can be learned from growth mindset controversies? *American Psychologist, 75*(9), 1269. https://doi.org/10.1037/amp0000794

Doctoral Consortium Research Summary

Kimberly Williamson
Cornell University
Khw44@cornell.edu

ABSTRACT: As more decisions in higher education are being justified with data, it becomes critical to research how equity, diversity, and inclusion (EDI) are affected by these data-driven decision points. This research plan explores how to use learning analytic dashboards (LADs) to inject EDI-related data to aid instructors in making course decisions to create equitable and inclusive course environments. Using informational interviews and an exploratory field, this research will examine how different data storytelling elements will affect instructors' understanding of the data and, eventually, the decisions they make from those interpretations.

Keywords: Learning Analytic Dashboard, Equity, Diversity, Inclusion, Data-Driven Decision-Making

1 BACKGROUND AND GOALS

Across post-secondary institutions, decisions related to strategic planning, pedagogy, curriculum, admissions, and promotion and tenure, occur continuously (Fairweather, 2002) and have consistently remained essential to higher education. Increasingly, these decisions are guided by data (Hora et al., 2017), offering researchers new opportunities to investigate decision-making processes empirically. Amid these ongoing decisions, many higher education institutions have espoused a commitment to advancing equity, diversity, and inclusion (EDI). Enacting this commitment creates the potential not only to gain insights into the decision-making processes in higher education but to identify interventions in the decision-making processes to improve the educational climate for minoritized and underrepresented students. While there are many decision-making points in higher education, this research will concentrate on decisions made by instructors to improve the classroom environment. This study continues the robust research direction of using a Learning Analytics Dashboard (LAD) to communicate data to instructors for decision-making.

Previous research on LADs has shown them to be instrumental in understanding the effects of how instructors make sense of data and adapt these data insights into their teaching pedagogy (Echeverria et al., 2018; Li et al., 2021; Wise & Jung, 2019). Even with the vast amount of research on instructor-facing LADs, there has been minimal exploration into incorporating EDI into LADs to create more inclusive and equitable classrooms (Williamson & Kizilcec, 2022). This research will test different features of an instructor-facing LAD to understand how instructors use equity-related data and, ultimately, how they use it to improve their courses. The research questions guiding this high academic quality research are:

- RQ1: What behavioral and performance course data influences instructors' decisions to promote equity and inclusion in their courses?

- RQ2: What data storytelling techniques in LADs can be used to facilitate understanding equity and inclusion-related information?
- RQ3: To what extent does including equity and inclusion-related information in course dashboards influence instructors' (a) perceived usefulness of the dashboard, (b) perceived usability of the dashboard, (c) number of concrete proposed actions, (d) equity-relatedness of proposed actions?

2 CURRENT KNOWLEDGE, EXISTING APPROACHES, AND PROPOSED SOLUTION DIFFERENTIATION

2.1 EDI Dashboards

Dashboards are a common tool used to inform decision-making because they communicate data in various ways and have the potential for far reach. However, research on EDI dashboards in higher education is limited (Williamson & Kizilcec, 2022). There have been specialized areas of education, such as nursing and radiology, that have explored the use of EDI in dashboards using case studies (Oates et al., 2022; Schmidt & MacWilliams, 2015). In higher education, the Equity Scorecard (Harris & Bensimon, 2007) explored using dashboards for improving equity and laid the groundwork for emphasizing the importance of quantitatively measuring inequities and putting the results into the hands of university decision-makers. While the Equity Scorecard and other case studies improved equity at individual institutions equipped with the resources to sustain a team, they have not provided a large-scale solution for displaying equity-related information for decision-making.

2.2 Visualizations, Instructional Data-Driven Decision-Making, and Data Storytelling

Within education, continuous improvement refers to monitoring policy, processes, and outcomes to identify and correct problems (Hora et al., 2017) but rarely includes how to translate data into actionable knowledge (Hora et al., 2017). Though the educational arena has begun using learning dashboards to give students and faculty feedback, recent research has begun to concentrate on moving beyond "usable" visualizations toward those effective for teaching and learning (Charleer et al., 2018; Echeverria et al., 2018; Schwendimann et al., 2017).

Data visualizations tell a story about data; however, the complexity of the story and how it is communicated varies (Wojtkowski & Wojtkowski, 2002). *Narrative* visualizations are methods intended to tell a story and have been included in systems via functionalities like user-created annotations (Wojtkowski & Wojtkowski, 2002). Data Storytelling (DS) encompasses techniques for communicating the story of data by utilizing narrative tools like plots or structures (Charleer et al., 2018; Echeverria et al., 2018), yet existing research has not established which visual elements support the narrative. Echeverria et al. (2018) build from these ideas and propose "Educational Data Storytelling," an approach that designs interfaces for visual learning analytics focused on clearly explaining student data that aligns with the teacher's intended learning design (Echeverria et al., 2018). To be effective, data storytelling design must consider motivating and informing the user of appropriate actions (Echeverria et al., 2018).

One technique that DS employs to motivate and inform users is messaging (Hullman & Diakopoulos, 2011). While messaging has been shown to have a significant impact on influencing people's motivation to accept EDI initiatives and adopting pro-EDI behaviors (Dover et al., 2016; McClanahan et al., 2022; Plaut et al., 2011), there has been limited exploration into these effects when messaging is combined with data visualizations (Jarke & Macgilchrist, 2021). Studying those effects is even less common for educational dashboards displaying EDI-related data (Jarke & Macgilchrist, 2021; Williamson & Kizilcec, 2022). This research proposes using Echeverria et al.'s (2018) Educational Data Storytelling in conjunction with psychological theories of motivation related to accepting and adopting pro-EDI behaviors.

3 RESEARCH METHODOLOGY

To understand how equity-related data can best be displayed to instructors, this research will utilize a.) informational semi-structured interviews to understand what data instructors need to help improve equity and inclusion in their courses and b.) an exploratory field study to understand how instructors react when presented with equity-related data about their courses.

3.1.1 Context

This study will be conducted within an academic department at a selective research university in the United States. The data displayed in the dashboards will come from the student information system (SIS) and the learning management system (LMS). The participants will be a mix of tenure-track and lecture faculty who teach undergraduate large lecture courses in the information science discipline. The instructors currently do not receive any data about their courses in the form of a dashboard.

3.1.2 Informational Semi-Structured Interviews

Prior to showing instructors a dashboard for addressing equity and inclusion in their courses, semi-structured interviews will be conducted with five instructors. These interviews will start by asking instructors about their prior experiences addressing equity and inclusion in their courses, including questions about prior ideas about gaps in their courses. Then the instructors will be asked about general data/analytic needs, followed by specific questions regarding what data would be helpful for equity and inclusion. The interview will end by addressing data presentation and asking questions regarding what type of messaging would be most helpful.

3.1.3 Exploratory Field Study

The field study will be conducted using ten instructors who are different from those who participated in the informational interviews. Two dashboards will be designed and built to conduct a within-subjects field study using the information gathered from the informational interviews. Since this study will also introduce dashboards to the participant pool, using a within-subject design helps to parse out findings related to general dashboard use versus the usage of a dashboard displaying equity-related data. One dashboard will display general course information in the form of historical data related to past student performance and withdrawals and current semester information about the currently enrolled students. While there will be some data disaggregation by major or year in school, the data will not be disaggregated by any socio-demographic indicators. The second dashboard will be designed to show similar historical and current semester information similar to the first dashboard described above. However, it will also allow and use DS elements to encourage the disaggregation of the data based on various socio-demographic indicators such as race/ethnicity, gender, and first-generation status.

To focus the instructors for the field study, they will be tasked with generating changes they can make to their courses to improve equity and inclusion. Each instructor will first be presented with dashboard one, followed by a short survey, and then they will be presented with dashboard two, followed by another short survey. While reviewing each dashboard, the researcher will ask each instructor to provide commentary on their interpretations of the data and dashboard elements. The short survey following each dashboard will ask questions about usability and usefulness, along with providing a space for instructors to detail the changes they would make to their course. After they have reviewed both dashboards, a semi-structured interview protocol will be used to gather participants' summarizing thoughts on the dashboard experience and to ask about additional data they would need to make decisions about improving equity and inclusion in their course.

REFERENCES

Charleer, S., Moere, A. V., Klerkx, J., Verbert, K., & De Laet, T. (2018). Learning Analytics Dashboards to Support Adviser-Student Dialogue. *IEEE Transactions on Learning Technologies*, *11*(3), 389–399. https://doi.org/10.1109/TLT.2017.2720670

Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, *62*, 58–67. https://doi.org/10.1016/j.jesp.2015.10.006

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). https://doi.org/10.18608/jla.2018.53.6

Fairweather, J. S. (2002). The Mythologies of Faculty Productivity: Implications for Institutional Policy and Decision Making. *The Journal of Higher Education*, *73*(1), 26–48. https://doi.org/10.1353/jhe.2002.0006

Harris, F., & Bensimon, E. M. (2007). The equity scorecard: A collaborative approach to assess and respond to racial/ethnic disparities in student outcomes. *New Directions for Student Services*, *2007*(120), 77–84. https://doi.org/10.1002/ss.259

Hora, M. T., Bouwma-Gearhart, J., & Park, H. J. (2017). Data driven decision-making in the era of accountability: Fostering faculty data cultures for learning. *The Review of Higher Education*, *40*(3), 391–426. https://doi.org/10.1353/rhe.2017.0013

Hullman, J., & Diakopoulos, N. (2011). Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

Jarke, J., & Macgilchrist, F. (2021). Dashboard stories: How narratives told by predictive analytics reconfigure roles, risk and sociality in education. *Big Data & Society*, *8*(1), 205395172110255. https://doi.org/10.1177/20539517211025561

Li, Q., Jung, Y., & Friend Wise, A. (2021). Beyond First Encounters with Analytics: Questions, Techniques and Challenges in Instructors' Sensemaking. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 344–353). Association for Computing Machinery. https://doi.org/10.1145/3448139.3448172

McClanahan, K. J., Birnbaum, H., & Shih, M. (2022). *Diversity Messages That Invite Allies to Diversity Efforts* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/n9fb7

Oates, M. E., Coakley, F., Cohen, G. S., Goodman, T. R., & Yee, J. (2022). Using an Annual Diversity, Equity and Inclusion Dashboard to Accelerate Change in Academic Radiology Departments. *Academic Radiology*, S1076633222003804. https://doi.org/10.1016/j.acra.2022.07.007

Plaut, V. C., Garnett, F. G., Buffardi, L. E., & Sanchez-Burks, J. (2011). "What about me?" Perceptions of exclusion and Whites' reactions to multiculturalism. *Journal of Personality and Social Psychology*, *101*(2), 337–353. https://doi.org/10.1037/a0022832

Schmidt, B. J., & MacWilliams, B. R. (2015). Evaluating an Accelerated Nursing Program: A Dashboard for Diversity. *Journal of Professional Nursing*, *31*(2), 82–88. https://doi.org/10.1016/j.profnurs.2014.09.001

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Williamson, K., & Kizilcec, R. (2022). A Review of Learning Analytics Dashboard Research in Higher Education: Implications for Justice, Equity, Diversity, and Inclusion. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 260–270. https://doi.org/10.1145/3506860.3506900

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

Wojtkowski, W., & Wojtkowski, W. G. (2002). Storytelling: Its role in information visualization. *European Systems Science Congress*, *5*.

Interaction opportunities: A design strategy for visualisations for meaningful learning

Author(s): Miguel Canizares
Queensland University of Technology
m.canizaresmena@qut.edu.au

ABSTRACT: Learning analytics systems often use dashboards to display relevant information. But who can, and who should decide which information is displayed? Deciding what is meaningful prior to designing a visualisation constrains users to the designers' view and can limit the users' ability to make meaning from the underlying data. This research explores whether *interactive* visualisations provide users with greater opportunities for agency in meaning-making than static visualisations. The research adopts a mixed-methods approach to investigate teachers' and students' use of static and interactive visualisations.

Keywords: Visual learning analytics, user experience, meaning, meaning-making.

1 BACKGROUND

In Learning Analytics (LA), data visualisations and dashboards are often used to deliver insights to different stakeholders (Verbert et al., 2020). The extensive use of data visualisations stems from the need to represent complex data analytics succinctly so they can be easily understood (Sosulski, 2019). Interactive visualisations (e.g., zoom, filter, search, etc.) have become a popular design strategy to visualise high-dimensional data (Yi et al., 2007), addressing a key limitation of static visualisation by allowing users to select which data are displayed and the form they are presented (Few, 2009).

Typically, visualisation techniques are selected according to user experience criteria such as usability. However, these techniques are dominated by an assumption that the role of the visualisation is to *transmit pre-defined insights to users* in a one-way channel. The users' role is relegated to one of "information consumer", with minimal opportunity for agency or consideration of context and environment. This can be problematic in dynamic learning contexts where concepts and meanings are constructed over time and include opportunities for learning in the moment.

Learning which is meaningful to the learner is seldom limited to receiving or recalling pre-packaged information but involves active meaning-making through interactions; seeking novelty and integrating new ideas with previous experience; recognizing possibilities for application, critiquing, challenging, extending; and allowing the process to change their actions or perspectives. This active engagement is a process of meaning-making that relies on the learner's socio-cultural context, previous knowledge and immediate learning environment (Martinez-Maldonado et al., 2017). The designer cannot know many of these factors in advance. Thus, pre-defined visualisations will always be deficient to some extent, limiting meaning-making opportunities for the learner.

2 RESEARCH QUESTIONS

This research holds the notion that interactivity in visualisations can support a richer meaning-making process for the learner than static visualisations. In short, we hold that meaning arises from interaction and can be found through action in the world (Dourish, 2004). The user's meaning emerges from interactions between the user and the world around them relying on the socio-cultural context and technological opportunities for interaction (de Souza et al., 2016). Such a position proposes an interesting design challenge: how might visualisations be designed to allow meaning-making within the learner's context if that context is not fully known in advance of the interaction? This question is further explored by the following sub-questions:

- What is the role of user-system interactions concerning the user's meaning-making?
- What is the impact of user-system interactions in reflective learning?
- Can user-system interactions support different learners' contexts? If so, how?

3 WHO DECIDES WHAT IS MEANINGFUL

Data visualisations often employ a set of pre-established techniques, practices and guidelines to create meaningful artifacts (Rosner, 2018). Typically, information flows in one direction: from data to viewer. The use of visualisations as a channel to convey a message brings the assumption that *meaning is a property of the visualisation*. However, new avenues for research in visualisation design may be found in re-thinking information flows as more like a negotiation, where meaning unfolds through interaction opportunities (de Souza et al., 2016). A negotiation approach acknowledges that *meaning is not a property of the visualisation*, but rather meaning emerges from user-system exchanges.

Exposing the assumption of *meaning being a property of the visualisation* raises the question, who decides what is meaningful? Traditionally, designers are in charge of understanding the users' context and deciding which view of the data is going to be meaningful to the users. To improve the designers' knowledge of the users' context, human-centred design (HCD) aims to include diverse users, interpretations and a broader socio-cultural context (Jivet et al., 2018). Meaning is crafted with users which improves technology use, but it is still pre-defined in advance.

Meaning-making through interactive visualisation is consistent with theories of learning as more than a cognitive process (Goodyear et al., 2018; Kress & Selander, 2012; Shepard, 2000). These theories posit that people *actively engage* in making sense of a situation by drawing from their history, cultural resources, identities, emotions and environment including technology (Fenwick et al., 2011). This view is widely accepted in education research (Goodyear et al., 2018). Yet designing visualisations that support dynamic learning contexts with unknown socio-cultural elements that can facilitate or hinder the learning process is challenging (Shepard, 2000). To accommodate the flexibility needed to support dynamic learning contexts, visualisation techniques need to be rethought, moving from *pre-defining meaning* in advance towards *negotiating meaning* through interaction opportunities.

HCD has advanced visualisation design by including the human-in-the-loop and reducing the gap between design and intended use (Sarmiento & Wise, 2022). However, computational interactivity in design can have a greater role to play than their usual cognitive portrayal and can potentially extend the micro-contexts that limit HCD artifacts to be transposed to other contexts.

4 INTERACTION OPPORTUNITIES

This research uses the term "interaction opportunities" to refer to interactive elements of a visual display that afford users the ability to explore data and make meaning that has not been pre-determined by the visualisation designer. They are not merely event-response pairs attached to screen elements (see Table 1). Interaction opportunities serve to allow users to reach the interpretation that is most important to them (Sengers & Gaver, 2006). Users can have different intentions and motivations for the same interaction (Huta, 2016). However, by focusing on designing computational interactivity based on interaction opportunities instead of designing just for transmitting information, visualisations can support more intentions and motivations with the same interaction. The purpose of design is not to predict what users need to know, rather it is to provide opportunities to engage in meaning-making through computational tools given certain data. Proposed opportunities for interaction explored in this paper include *comparison* and *association* as desired tasks from a teacher/learner perspective (Sedrakyan et al., 2019) and *transition* for awareness of change in the interface. The differences in meaning-making opportunities, when designed for interactivity as compared to those designed in advance, are explored in Table 1.

Table 1: Comparison and association techniques are commonly used when exploring data. Designers decide on the users' behalf what is most important based on their interpretation of the users' intentions. However, *interaction opportunities* hand in the power to explore the data to the user. The role of the designers is to facilitate opportunities for meaning-making to emerge.

Interaction opportunities	*Computational interactivity*	*Designing in advance*	*Designing opportunities*
Comparison	Interface capabilities to compare the data	The designer decides which data stays and not	The users can add/remove data based on their comparison needs
Association	Interface capabilities to associate the data	The designer shows important associations	The user can associate data from overview to fine details
Transition	Interface capabilities of awareness of change	The designer animates important elements	The users' actions activate animations on most elements

These interaction opportunities are evaluated by the notion of meaning arising from interaction (de Souza et al., 2016). We adopt Martela & Steger's (2016) view of meaning as having three components: 1) *coherence* as the cognitive component to understanding experience, 2) *purpose* as the goal-oriented component and 3) *significance* as the value worth component.

Interaction opportunities supported by computational interactivity are just one piece of the meaning-making puzzle in dynamic learning contexts. The experiments in interactive visualisations under this conceptualisation aim to provide opportunities that will invite users to interact while at the same time supporting diverse motivations and intentions by curating an environment where meaning-making is more likely to occur.

5 METHODOLOGY

Mixed-methods experiments are designed to explore how the presence or absence of interaction opportunities affects participants' meaning-making in a learning analytics dashboard. The experiments follow an embedded mixed methods design with a quantitative element as the primary design and qualitative elements as a secondary design (DeCuir-Gunby & Schutz, 2017).

The experiments use data from the GoingOK platform for reflective writing http://goingok.org[1]. The platform offers a large quantity of reflective writing data including pseudonyms, timestamps, group codes, reflection text and reflection state points (0 - distressed to 100 - soaring). The platform groups students according to logical learning groups (e.g., units, classes, courses). Students can write a series of short reflections over time, typically the duration of the class (e.g., a 13-week semester).

Quantitative data are collected via a survey that captures participants' orientations to interactions using the Hedonic and Eudaimonic Motives for Activities-Revised scale (HEMA-R) (Huta, 2016) used in previous HCI research focused on meaning (eg., Mekler & Hornbæk, 2016). The qualitative component is composed of two parts to investigate how the participants meaning and meaning-making were affected. First, is a think-aloud component where teachers narrate why they are doing specific actions while interacting with the interface. Second, a semi-structured interview after the experiment. The semi-structured interview has at its core the qualitative aspects of the concept of meaning - coherence, purpose and significance (Martela & Steger, 2016). The hypotheses to be tested are a) interaction opportunities do not have an effect on the aspects of the concept of meaning and b) interaction opportunities do not have an effect on orientations.

The participants are current users of the GoingOK platform in the role of teacher (group administrator) or student (author). Each participant's role is associated with a different experimental protocol.

The artifacts are four dashboards (two for teachers and two for students) developed and designed using the D3.js[2] visualisation library in a Typescript environment. Both static and interactive visualisations share a common typescript class ensuring that both dashboards include the same key components. The experimental dashboard extends the typescript classes with functions that facilitate a range of interaction opportunities. The source code for these visualisations is open source and available on GitHub[3]. The visualisations and content on the dashboards for each participant's role are the same; the difference is in the interaction opportunities available to the users. The experimental dashboard has comparison, association, and transition interaction opportunities while the control dashboard lacks these. Data familiarity effects were not minimized as meaning-making acknowledges the importance of the relationship between data and user.

5.1 Teachers experiment

This experiment is A/B (i.e., control and experimental settings) with repeated measures where the independent variable is interaction opportunities. The participants are current lecturers or sessional academics at an Australian University and have access to student data. The participants were asked to explore the data in the dashboard as if the semester was recently finished. To date, 4 participants have participated in the study: 1 male and 3 female, 2 academics work in teaching and education and

[1] http://goingok.org
[2] https://d3js.org
[3] https://github.com/maciiv/goingok-interactive-visualisations

the other 2 in information technology. This group will have three iterations of the protocol described in Figure 1. Each iteration will incorporate feedback into the experimental dashboard.

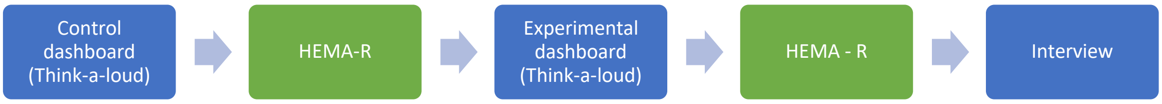


Figure 1: Experimental protocol. Blue rectangles collect qualitative data and green rectangles collect quantitative data. The protocol has a duration of one hour in total

5.2 Students experiment

The participants are students undertaking/finished a unit where the GoingOK platform is offered as a reflective journal. This experiment will be composed of two phases: 1) A/B with repeated measures where the independent variable is interaction opportunities. It is expected to recruit a total of four participants. This phase will follow the protocol described in Figure 1. The participants will be asked to explore the data as if they just finished the unit. This phase will serve as a pilot study to gather detailed insights into the student's needs. 2) A/B with independent measures where the independent variable is interaction opportunities. This phase will follow a different protocol as it will be deployed completely online by embedding a link to the HEMA-R scale, which will include an additional textbox to collect short user experiences. It is expected to reach hundreds of students with this. The participant's task will be to explore the data as desired while undertaking a unit. The students' dashboard using natural language processing (NLP) analysis will find phrases associated with reflective writing and show them in a timeline to identify trends and a network to identify associations. This data will be available to both control and experimental groups only differing in the interaction opportunities available to them.

6 PROGRESS

To date, the first iteration of the teacher's experiment has been finalised and the data analysed. Based on the feedback received the teacher's experimental dashboard has been updated. The student's dashboard development is ready for the first iteration of the experiment waiting only on the results of the NLP model.

REFERENCES

de Souza, C. S., Fontoura de Gusmão Cerqueira, R., Marques Afonso, L., Rossi de Mello Brandão, R., & Soares Jansen Ferreira, J. (2016). Software Developers as Users : Semiotic Investigations in Human-Centered Software Development. In *Software Developers as Users* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-42831-4_1

DeCuir-Gunby, J., & Schutz, P. (2017). *Developing a Mixed Methods Proposal: A Practical Guide for Beginning Researchers*. https://doi.org/10.4135/9781483399980

Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. MIT Press.

Fenwick, T., Edwards, R., & Sawchuk, P. (2011). Why sociomateriality in education? An Introduction. In *Emerging Approaches To Educational Research: Tracing the Sociomaterial* (1st Edition, pp. 1–17). Routledge. https://doi.org/https://doi.org/10.4324/9780203817582

Few, S. (2009). *Now you see it : simple visualization techniques for quantitative analysis*. Analytics Press.

Goodyear, P., Ellis, R., & Marmot, A. (2018). Learning Spaces Research: Framing Actionable Knowledge. In R. A. Ellis & P. Goodyear (Eds.), *Spaces of Teaching and Learning: Integrating Perspectives on Research and Practice* (pp. 221–238). Springer Singapore. https://doi.org/10.1007/978-981-10-7155-3_12

Huta, V. (2016). Eudaimonic and Hedonic Orientations: Theoretical Considerations and Research Findings. In J. Vittersø (Ed.), *Handbook of Eudaimonic Well-Being* (pp. 215–231). Springer International Publishing. https://doi.org/10.1007/978-3-319-42445-3_15

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 31–40. https://doi.org/10.1145/3170358.3170421

Kress, G., & Selander, S. (2012). Multimodal design, learning and cultures of recognition. *The Internet and Higher Education*, *15*(4), 265–268. https://doi.org/10.1016/j.iheduc.2011.12.003

Martela, F., & Steger, M. (2016). The three meanings of meaning in life: Distinguishing coherence, purpose, and significance. *The Journal of Positive Psychology*, *11*(5), 531–545. https://doi.org/10.1080/17439760.2015.1137623

Martinez-Maldonado, R., Goodyear, P., Carvalho, L., Thompson, K., Hernandez-Leo, D., Dimitriadis, Y., Prieto, L., & Wardak, D. (2017). Supporting collaborative design activity in a multi-user digital design ecology. *Computers in Human Behavior*, *71*, 327–342. https://doi.org/10.1016/j.chb.2017.01.055

Mekler, E., & Hornbæk, K. (2016). Momentary Pleasure or Lasting Meaning? *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4509–4520. https://doi.org/10.1145/2858036.2858225

Rosner, D. (2018). *Critical fabulations : reworking the methods and margins of design*. The MIT Press.

Sarmiento, J. P., & Wise, A. F. (2022). Participatory and Co-Design of Learning Analytics: An Initial Review of the Literature. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 535–541. https://doi.org/10.1145/3506860.3506910

Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, *50*, 19–38. https://doi.org/10.1016/j.jvlc.2018.11.002

Sengers, P., & Gaver, B. (2006). Staying open to interpretation. *Proceedings of the 6th ACM Conference on Designing Interactive Systems - DIS '06*, 99–108. https://doi.org/10.1145/1142405.1142422

Shepard, L. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, *29*(7), 4–14. https://doi.org/10.3102/0013189X029007004

Sosulski, K. (2019). *Data Visualization Made Simple*. Routledge. https://doi.org/10.4324/9781315146096-1

Verbert, K., Ochoa, X., de Croon, R., Dourado, R. A., & de Laet, T. (2020). Learning analytics dashboards: The past, the present and the future. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 35–40. https://doi.org/10.1145/3375462.3375504

Yi, J., Kang, Y., Stasko, J., & Jacko, J. (2007). Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1224–1231. https://doi.org/10.1109/TVCG.2007.70515

Investigating how Australian university teachers use learning analytics as part of their teaching practice

Author: David Fulcher
University of Wollongong
dfulcher@uow.edu.au

ABSTRACT: In recent years, the increased investment put into learning analytics by universities has been criticized for being disconnected from the lived experience of educators. This is important as ultimately claims about effective use of learning analytics depend on how it is integrated into teaching. This study seeks to address this gap by providing an in-depth view of teacher practices concerning learning analytics. Featuring a qualitative multiple case study design with a focus on examining practice, of which we know very little, the study captures richness in the actions and circumstances surrounding the use of learning analytics. The study will contribute to knowledge by providing an in-depth view of teachers' use of learning analytics. This will help illuminate how best to support teachers as they look to make the most of learning analytics to improve teaching and learning.

Keywords: “learning analytics”, “higher education”, “teacher practices”

1 BACKGROUND

The multifaceted nature of university teaching means educators think about and act in relation to multiple aspects that comprise a classroom. Learning analytics is but one aspect that can be brought into the environment to help enrich what is already known to ultimately improve possibilities for meaningful action by the teacher. There has been relatively little focus on the perspectives of university teachers and how learning analytics fits with their existing practice in the research to date (Heilala et al., 2022; Hilliger et al., 2020; Klein et al., 2019; Kollom et al., 2021; Shibani et al., 2020) Previous studies that have explored teacher interaction with learning analytics tend to be in research settings rather than authentic classroom settings (Ifenthaler & Yau, 2020; Knobbout & Van Der Stappen, 2020; Valle et al., 2018). In addition to this, the learning analytics tools used in these studies were often purpose built for the research project and not widely available to the majority of teachers at an institution (Corrin et al., 2016; Knobbout & Van Der Stappen, 2020; van Leeuwen et al., 2015). Limited attention has been paid towards authentic classroom settings and how the teacher integrates learning analytics into their practice.

Existing studies about learning analytics also tend to be primarily quantitative in nature with a focus on the technical intricacies involved to explain causes for student attrition or the effectiveness of certain learning designs (Avella et al., 2016; Papamitsiou & Economides, 2014; Viberg et al., 2018). There are few qualitative studies that investigate in detail university teachers' use of readily available learning analytics in their everyday practice. This is important because the insights that can be derived from qualitative studies allow for a broader consideration of the significant factors in student learning. This includes things such as the teaching philosophy of the individual teacher and related conceptual models of higher education teaching and learning (Biggs & Tang, 2011; Laurillard, 2002; Prosser &

Trigwell, 1999)(Biggs & Tang, 2011; Laurillard, 2002; Prosser & Trigwell, 1999; Ramsden, 2003). The premise here is that human utilisation of learning analytics is just as important as the technical design aspects. Learning analytics is not an autonomous force outside human control that of itself can result in improvements to student learning. It is but one tool in a suite of tools available to teachers. If the potential of learning analytics is to be realised it is crucial that contextual factors are considered and understood. The teacher and their everyday practices are a key part of this.

While student learning can be supported through teachers' use of learning analytics (Ifenthaler & Yau, 2022; Knobbout & Van Der Stappen, 2020), the practical problem informing this study is that such benefits are not reaching students because there is no widespread integration of learning analytics into teaching practice (Kaliisa et al., 2021; Klein et al., 2019; Viberg et al., 2018). Closer attention needs to be paid to settings where learning analytics is actually being used by teachers as part of authentic practice. If we accept that learning analytics is valuable, then it is vital that universities focus their investment so that it gets used by the right people for the things that are most effective. This involves more than student attrition. Learning analytics has the potential to help enrich learning experiences for all students. Regardless of institution-level efforts to harness learning analytics, the classroom (online or otherwise) is where the learning process is enacted and ultimately how students are primarily guided towards successful outcomes. So it is important to understand what teachers think about learning analytics and how they currently make use of data in their everyday practice. Doing so can help to identify what needs to be in place to facilitate meaningful and sustained use of learning analytics.

2 RESEARCH GOALS

The purpose of this study is to examine how university teachers use learning analytics in their everyday teaching practice. Little is known about the broader practices of university teachers with a unit of work from start to finish and how learning analytics fits within this context. As pointed out by Scanlon et al. (2013), many different education technologies that have failed to understand the perspectives of teachers and the circumstances of their work do not end up being used for sustained periods of time in authentic settings.

The central research question for this study is:

How do university teachers use learning analytics in their teaching practice?

The following sub-questions guide the study:

1. How does a university teacher use learning analytics in a unit?

2. What influences a university teachers' use of learning analytics in a unit?

3 STATE OF CURRENT KNOWLEDGE

Other studies have highlighted the importance of context and the pedagogical complexities that arise in different education settings (Avramides et al., 2015; Kaliisa et al., 2021; Klein et al., 2019; Ma et al., 2018). So, when these complexities are taken into account, learning analytics has been shown to have more success integrating into teacher practice (Knobbout & Van Der Stappen, 2020; Shibani et al., 2020). These findings provide support for the premise that aligning learning analytics with teacher practices can help realise the potential benefits of learning analytics through sustained and meaning application in class.

Finding common ground with how learning analytics can be integrated into teacher practices can be difficult when nuance exists in different university settings. The increased investment put into learning analytics by universities has been criticised for being disconnected from the lived experience of educators as it is often approached from a top-down institution-wide perspective (Colvin et al., 2015; Kaliisa et al., 2021; Sclater & Mullan, 2017). There has been relatively little focus on the perspectives of university teachers and how learning analytics fits with their existing practice in the research to date (Heilala et al., 2022; Shibani et al., 2020). To understand more about the use of learning analytics by university teachers, contextual factors that influence practice need to be investigated. Despite the importance of combining learning analytics with pedagogical approaches, few empirical studies demonstrate how this actually takes place in practice.

4 SIGNIFICANCE

The unique investigative focus of this study provides an opportunity to advance understanding by examining the wider practices of university teachers as they undertake a unit of work from start to finish and how learning analytics is positioned within this setting. Doing so will help explain teacher motivations, implementation approaches, challenges and outcomes when using learning analytics readily available in everyday practice. Some of this work has already begun with Shibani et al. (2020) looking at teachers use of an automated writing feedback tool in their classrooms. The doctoral study outlined here takes a more holistic viewpoint by examining teacher practices as they prepare for, and teach a unit to explore how learning analytics tools, readily available across the institution, are used and/or not used. This provides an opportunity to advance the learning analytics research agenda through a richer understanding of the perspectives of teachers and the circumstances of their work.

5 METHODOLOGY OVERVIEW

This study takes a qualitative approach to explore the issues involved with university teachers' needs and practices concerning learning analytics. Five university teachers have been selected using a stratified purposive sampling strategy (Creswell, 2020), seeking a wide variation on a common dimension. Here the common dimension is university teachers using learning analytics in practice. The variation comes from different educational settings, such as a large first year undergraduate unit or a small postgraduate unit delivered fully online. Cases are bounded at the same site with similar infrastructure. The University of Wollongong, Australia has been selected for the study. I have been working at this site in the field of learning analytics for eight years, allowing me to be aware of a range of academics across the university that use and do not use learning analytics in their teaching. Because of this work I am also familiar with the technical infrastructure that enables learning analytics to be implemented. The selection of this site also takes advantage of circumstances that enable information to be collected from participants that are easily accessible.

The study will be anchored to specific experiences and accommodate situational complexity to investigate similarities and differences between each case. The Theory of Practice Architectures (TPA) (Kemmis & Grootenboer, 2008) will be used as a way of explaining what practices are made up of and how they are both influenced by and shape the environment in which they exist. TPA is a form of practice theory, which encompasses a variety of sociocultural theories concerned with practice (Nicolini, 2013). Practice architectures are seen as socially constructed activities that involve sayings, doings and relatings that come together in the "project" of a practice (Mahon et al., 2017). A project covers the intentions motivating a practice, the actions (interconnected sayings, doings, and relatings) carried out during the conduct of a practice, as well as the ultimate goals aimed for through the practice. There are three kinds of arrangements seen to exist at the same time at a site of practice: cultural-discursive arrangements, material-economic arrangements, and socio-political arrangements. These arrangements are interrelated and provide the conditions in which practice unfolds. Here the focus is on teaching as a practice rather than on the teacher themselves.

To understand how a teacher uses learning analytics in a unit they have been followed over an entire semester. Different types of qualitative data (interviews, observations, and artefacts) were collected along the way, with data interpretation occurring contemporaneously to inform subsequent data collection points. Six one-to-one semi-structured interviews were conducted, one before semester, four subsequent check-in interviews during semester, and a final interview after semester. Following negotiation with participants during the initial interview, data was also collected through face-to-face classes. Figure 1 summarizes the method and timing of data collection for each case.

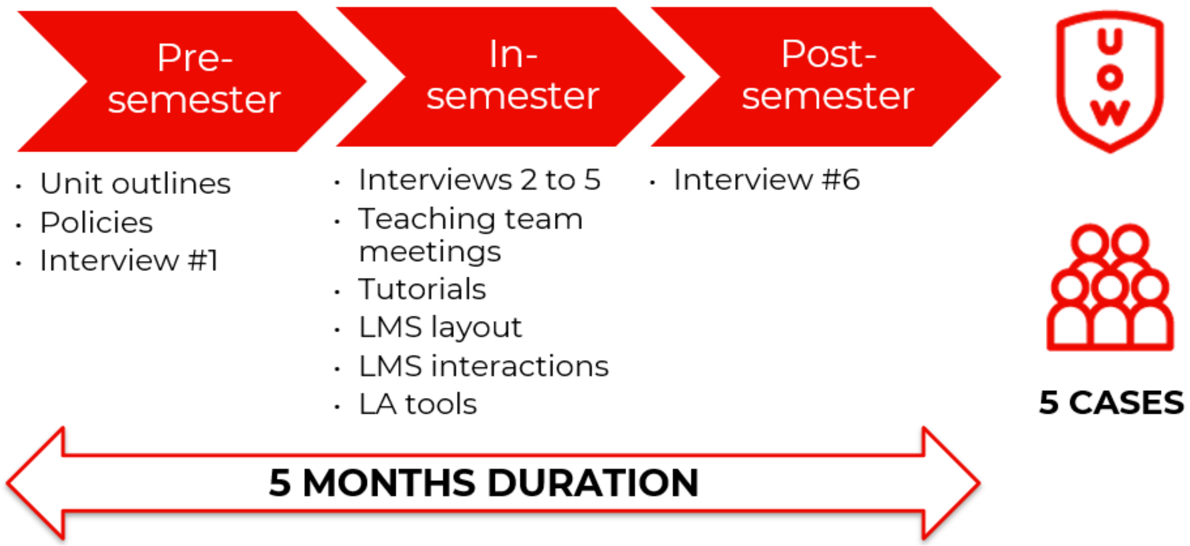


Figure 1: Study Design Overview

Ethical considerations for this study include the use of pseudonyms for participants and the use of a consent form to confirm participants understand what information is being collected and they consent to it being used in the intended manner. While some student data was noted during observations, this was to set context and potentially serve as stimulus for interview questions with teachers, not for in-depth data analysis. Obtaining informed consent from students was therefore not necessary, as I did not need to identify any of these people for the purposes of the study. Nevertheless, any student data collected has been de-identified via a concordance file in line with best practice for data management in case studies (Yin, 2018). Furthermore, any results arising from such data will be reported in aggregate form to further protect student anonymity.

This study operates within a qualitative research paradigm based on particular philosophical assumptions. The stance here is that 'reality' is constructed within historical and social contexts by individuals. Knowledge is socially constructed rather than standing alone objectively from everything else. Under this paradigm the researcher embodies the data collection mechanism, so it is important to acknowledge my own subjectivities to help minimize bias (Creswell, 2020). The motivation for this study stems from my professional work in a central university unit developing learning analytics tools for teachers and others to identify students in need of support. This study provides an opportunity to extend this work by providing an opportunity to better understand the perspectives of teachers and how learning analytics fits with their existing practice. I am approaching this study as somebody who is not a teacher in this context. Similarly, participants may view me through the lens of my operational role rather than as a researcher. While I cannot get rid of these subjectivities, different strategies will be adopted to reduce their impact on the quality of the study: member checking and triangulation. In addition to this a reflexive journal is being used throughout the study to continually interrogate my own practice as a researcher and to remember I am not a neutral force and that my own subjectivities shape what I see.

6 STATUS

I have developed and defended my research proposal. I am about to complete data collection for the study. Data collected for each case consists of interviews, observations, and artefacts. I have also started writing up the methodology and results sections of my thesis.

REFERENCES

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning*, *20*(2), 13–29.

Biggs, J. B., & Tang, C. S. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). McGraw-Hill/Society for Research into Higher Education/Open University Press.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., & Fisher, J. (2015). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*.

Corrin, L., De Barba, P., Lockyer, Gašević, D., Williams, D., Dawson, S., Mulder, R., Copeland, S., & Bakharia, A. (2016). *Completing the loop: Returning meaningful learning analytic data to teachers*.

Creswell, J. (2020). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (Global). Pearson Education, Limited.

Heilala, V., Jääskelä, P., Saarela, M., Kuula, A.-S., Eskola, A., & Kärkkäinen, T. (2022). “Sitting at the Stern and Holding the Rudder”: Teachers’ Reflections on Action in Higher Education Based on Student Agency Analytics. In L. Chechurin (Ed.), *Digital Teaching and Learning in Higher Education: Developing and Disseminating Skills for Blended Learning* (pp. 71–91). Springer International Publishing. https://doi.org/10.1007/978-3-031-00801-6_4

Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y.-S., Muñoz-Merino, P. J., Broos, T., Whitelock-Wainwright, A., & Pérez-Sanagustín, M. (2020). Identifying needs for learning analytics adoption in Latin American universities: A mixed-methods approach. *The Internet and Higher Education*, *45*, 100726. https://doi.org/10.1016/j.iheduc.2020.100726

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-020-09788-z

Kaliisa, R., Mørch, A. I., & Kluge, A. (2021). ‘My Point of Departure for Analytics is Extreme Skepticism’: Implications Derived from An Investigation of University Teachers’ Learning Analytics Perspectives and Design Practices. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-020-09488-w

Kemmis, S., & Grootenboer, P. (2008). Situating praxis in practice: Practice architectures and the cultural, social and material conditions for practice. In S. Kemmis & T. Smith (Eds.), *Enabling Praxis: Challenges for Education* (pp. 37–62). Sense.

Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019). Technological barriers and incentives to learning analytics adoption in higher education: Insights from users. *Journal of Computing in Higher Education*, *31*(3), 604–625. https://doi.org/10.1007/s12528-019-09210-5

Knobbout, J., & Van Der Stappen, E. (2020). Where is the Learning in Learning Analytics? A Systematic Literature Review on the Operationalization of Learning-Related Constructs in the Evaluation of Learning Analytics Interventions. *IEEE Transactions on Learning Technologies*, *13*(3), 631–645. https://doi.org/10.1109/TLT.2020.2999970

Kollom, K., Tammets, K., Scheffel, M., Tsai, Y.-S., Jivet, I., Muñoz-Merino, P. J., Moreno-Marcos, P. M., Whitelock-Wainwright, A., Calleja, A. R., Gasevic, D., Kloos, C. D., Drachsler, H., & Ley, T. (2021). A four-country cross-case analysis of academic staff expectations about learning analytics in higher education. *The Internet and Higher Education*, *49*, 100788. https://doi.org/10.1016/j.iheduc.2020.100788

Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies* (2nd ed.). RoutledgeFalmer.

Mahon, K., Kemmis, S., Francisco, S., & Lloyd, A. (2017). Introduction: Practice Theory and the Theory of Practice Architectures. In *Exploring Education and Professional Practice: Through the Lens of Practice Architectures*. Springer Singapore Pte. Limited. http://ebookcentral.proquest.com/lib/uow/detail.action?docID=4772456

Nicolini, D. (2013). *Practice Theory, Work, and Organization: An Introduction* (First Edition.). Oxford University Press.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systemic literature review of empirical evidence. *Educational Technology and Society*, *17*(4), 49–64.

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham [England] ; Philadelphia, PA : Society for Research into Higher Education & Open University Press, 1999.

Scanlon, E., Sharples, M., Fenton-O'Creevy, M., Fleck, J., Cooban, C., Ferguson, R., Cross, S., & Waterhouse, P. (2013). *Beyond prototypes: Enabling innovation in technology-enhanced learning*. Open University. http://oro.open.ac.uk/41119/1/

Sclater, N., & Mullan, J. (2017). *Jisc briefing: Learning analytics and student success – assessing the evidence*. Jisc.

Shibani, A., Knight, S., & Buckingham Shum, S. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education*, *46*, 100730. https://doi.org/10.1016/j.iheduc.2020.100730

Valle, L. D., Stander, J., Eales, J., Wei, Y., & Gresty, K. (2018). Stakeholder perspectives on graphical tools for visualising student assessment and feedback data. *Research in Learning Technology*, *26*. https://doi.org/10.25304/rlt.v26.1997

van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, *90*, 80–94. https://doi.org/10.1016/j.compedu.2015.09.006

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, *89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition.). SAGE Publications, Inc.

Measuring and Supporting Self-regulated Learning in Blended Learning: Developing an Analytical Framework

Esteban Villalobos
IRIT, Université de Toulouse, France
esteban.villalobos@irit.fr

ABSTRACT: Despite the advantages of Blended Learning (BL), several studies have shown that students require high levels of self-regulation to succeed in these practices. Still, there is little understanding of how students organize their learning in BL settings. This paper presents the objectives and current status of a thesis that seeks to understand how students' Self-regulated Learning (SRL) strategies manifest in BL contexts and how technological solutions can promote them. The contributions of this thesis are three-fold. First, we aim to develop novel analytical solutions to understand the dynamics of how SRL unveils in BL contexts. Second, we propose studying dashboard-based solutions to support students' SRL and teachers in promoting this ability. And third, we propose several case studies for evaluating both the analytical framework and technological solutions in authentic BL contexts. We expect that this thesis will contribute to Learning Analytics (LA) with new measures and tools to better understand the dynamics of SRL in BL.

Keywords: Self-regulated Learning, Blended Learning, Temporal Analysis, Feedback

1 INTRODUCTION

In recent years, and especially after the COVID-19 pandemic, Blended Learning has become more varied and commonly applied (Pelletier et al., 2022). Blended Learning (BL) combines traditional face-to-face and computer-mediated instruction (Graham, 2005). While BL has been shown to have positive pedagogical advantages, many students have problems regulating their learning (Broadbent, 2018; Graham, 2005), raising great interest in understanding and supporting students' self-regulation abilities in these scenarios.

Self-regulated Learning (SRL) is defined as a complex process that combines meta-cognitive, motivational, and emotional processes (Panadero, 2017). Prior research shows that students' SRL ability is a good predictor of their behavior and success in a course (Maldonado-Mahauad et al., 2018). However, most studies on SRL have been conducted in online contexts and cannot be directly extrapolated to BL settings. Only a few studies have been proposed to understand how these SRL processes manifest in BL (Araka et al., 2020; Broadbent, 2017).

To provide meaningful SRL support in BL, understanding how both external factors (e.g., the influence of the teacher or face-to-face classes) and internal factors (e.g., students' self-regulation abilities) influence learners' behavior in these contexts is key. This is one of the current challenges of the LA community since the analyses based on data traces are heavily influenced by the dynamics of the system in which the students operate (Dawson et al., 2019; Jovanović et al., 2021). Moreover, this is especially difficult in BL settings, where students are exposed to different learning modes (i. e., face-to-face or online) supported by various technologies. To better understand how students self-regulate in BL, there is a need to develop approaches that consider both external and internal factors across the different learning modes of BL scenarios in a holistic manner.

2 OBJECTIVES AND RESEARCH QUESTIONS

The general objective of this thesis is to (1) investigate how students' SRL strategies manifest in BL scenarios and to propose and (2) evaluate a Learning Analytics (LA) technological solution based on user-centered dashboards (for teachers and students) to support effective learning strategies. Three main objectives are derived from this general objective:

- **Objective 1:** To develop an analytical framework to study how students' SRL strategies manifest as a process across BL modes in a holistic manner.
- **Objective 2:** To offer novel dashboard-based technological solutions for supporting SRL in BL.
- Objective 3: To study the adoption and impact of the proposed technological solutions on students' SRL behavior and teachers' actions.

2.1 Measuring SRL in BL

Different methods have been proposed for studying how SRL manifests in different learning contexts, especially in online learning environments. These proposals go, from using self-reported data (Araka et al., 2020), to detecting tactics and strategies using the trace data collected from the course's Learning Management Systems (LMS) (Fan et al, 2021; Fincham et al., 2019). Some studies have also made the relationship between these techniques and the SRL theory to overcome the limitations of the context-specific nature of LA (Fan et al., 2021; Maldonado-Mahauad et al., 2018). Most of these methods have been applied online, and very few in Blended Learning settings. One of the problems is that the methods currently used fall short of capturing the influence of external factors, such as teacher interventions. In fact, analytical methods proposed in prior work encounter difficulties in providing indicators in real-time and in giving temporal meaning to the collected data (Jovanović et al., 2021), especially important in BL environments, in which face-to-face activities are combined online tasks. From this, we derive the following research question:

- **RQ1**: How can existing LA methods be adapted and combined with qualitative data to create an analytical framework for characterizing the dynamics of students' SRL strategies in BL?

2.2 Supporting SRL in BL

Researchers have proposed different approaches to support students' SRL (Pérez-Álvarez et al., 2022). According to Wong et al. (2019), the most common techniques consist of using educational prompts and integrated support systems based on dashboards. These solutions transform raw data into 'actionable insights' to produce students' behavioral changes (Jørnø and Gynther, 2018). So far, most of this prior work has been conducted in online settings, such as Massive Open Online Courses (MOOCs), where students have low interaction with the teacher (Wong et al., 2019).

Only a few studies have studied these solutions in BL contexts (e.g., Shyr et al., 2018; Michel et al., 2017; Pérez-Sanagustín, 2021). These studies suggest that dashboards could be appropriate for supporting SRL strategies and assisting students' goal setting, strategic planning, and time management, promoting their motivation and impacting their course performance (Pérez-Sanagustín et al., 2021; Matcha et al. 2020). And of these works, we identify two main limitations. First, current tools focus on supporting the students, misleading the role of the teacher, which is important in BL settings. Second, while some tools are based on theoretical models for SRL, it is still unclear if students understand them and incorporate them into their SRL strategies. This poses the following research questions for the project:

- **RQ2**: How useful (interpretable, actionable, and comprehensive) are the existing indicators provided dashboard-based solutions for supporting students' SRL and teachers in promoting this ability?

- **RQ3**: How do these dashboard-based solutions influence students' strategies and teachers' decision-making in BL scenarios?

3 PROJECT METHODOLOGY

Design Based Research (DBR) will be used as a methodological approach for guiding the whole thesis development. This approach combines experiments in real-world settings with theoretical models (Reimann, 2011). The interventions will be based on the NoteMyProgress (NMP) tool (Pérez-Sanagustín et al., 2022), a Moodle plug-in that provide both, teachers and students with dashboards based on indicators for supporting self-regulation. Three experimental cycles will be carried out to improve the tool and the analytical frameworks iteratively. After each cycle, the results will be published as part of the LASER project following an Open Science Framework[1].

All data collected during the thesis will consider the ethical considerations imposed by the laboratory in which this research is taking place. A Data Management Plan will be published for each experiment and revised by the ethical committee of the University using the DMP OPIDoR tool[2]. Also, a simplified data declaration treatment for the whole project will be defined and sent to the ethical committee.

4 CURRENT RESULTS

So far, we have finished the first design cycle of the project and started the second one. The main results of this first cycle are (1) two research conference papers published at the European Conference on Technology Enhanced Learning (EC-TEL) 2022 about the design and development of the NMP Moodle Plugin (Pérez-Sanagustín et al., 2022) and its evaluation in a BL authentic context (Villalobos et al., 2022); and (2) the first version of the NMP Moodle Plugin[3]. These results contributed to the advancement of the different objectives of the project.

4.1 Objective 1 – An analytical framework for measuring SRL in BL

The proposed analytical framework combines (1) self-reported data, (2) trace data and (3) course metadata (e.g., prior grades and face-to-face class schedules). The framework uses self-reported

[1] See https://osf.io/s86au/

[2] See https://dmp.opidor.fr/plans/16597/export.pdf?export%5Bquestion_headings%5D=true

[3] Available at: https://gitlab.com/laser-anr/notemyprogress-plug-in

instruments to measure the SRL ability profile of the students. This is combined with an analytical method that analyzes the trace data to extract tactics and strategies used by the students.

In Villalobos et al. (2022), we extended an analytical method proposed by Fincham et al. (2019) to incorporate the information from face-to-face classes, a crucial factor of BL. The method divides the data analysis into four steps: defining actions, detecting tactics, detecting strategies, and running statistical comparisons between students based on their employed strategies. Tactics are the underlying process that explains the students' actions during a learning session. These tactics are detected using a Hidden Markov Model (HMM). Strategies are defined as sequences of tactics applied by the student. To detect the strategy used by one student, we make a sequence of their applied tactics and when the tactic took place with respect to the face-to-face session. These sequences are then clustered using hierarchical clustering methods. Finally, we analyzed statistical comparisons between the students that applied different strategies.

4.2 Objective 2 - Novel technological solutions for supporting SRL in BL

The first version of NMP consisted of a dashboard-based solution for the students. These dashboards provide information about students' activities in the course to support time management and strategic planning. The design processes, detailed in Pérez-Sanagustín et al. (2022), consisted of workshops with experts and teachers. The tool's usability was tested in a local evaluation with teachers. Finally, we assessed the tool's impact on two university courses during a class semester. The results show that, although teachers found the tool useful to monitor students' advances, it lacks of support for students, who need actionable information in real-time.

4.3 Objective 3 - Adoption and impact of SRL support tools

In Villalobos et al. (2022), we observed that the use of the NMP tool was relatively sparse. The tool was adopted mainly by students with a high level of self-reported self-regulation ability. The students that used the tool mostly incorporated it into existing tactics, meaning they probably added NMP into prior learning strategies. Finally, no relationship was found between the student's grades and their use of NMP. Further research will explore how to improve the support provided to students.

5 FUTURE WORK

One of the main limitations of our advances so far is that the proposed analytical framework can only be used *a posteriori* (once the course is finished). That is, we can only analyze what happened in the course but not influence the students' actions as the course moves forward, in real time. Furthermore, the results found using the sequential analysis methods, such as Hidden Markov Models and clusters, are challenging to interpret. All of this limits our ability to give insights to the students and teachers to promote action in real time.

Recent works by Jovanović et al. (2021) and Saqr et al. (2021) propose context-dependent and context-independent indicators that could be used as real-time feedback for students. Short-term, and as part of the second research cycle, we propose building upon these indicators to propose temporal-based indicators that capture students' dynamics in the course and how they affect their behavior. Specifically, we propose conducting a qualitative study using user-centered design and anthropological methods to evaluate the students' understanding of these temporal-based indicators and whether this supports action. This study will be done in collaboration with the Millennium Nucleus Student Experience in Higher Education in Chile (NMEDSUP)[4] to see how this work can be extended to different institutions and contexts. Long-term, and as part of the third and final cycle of the thesis, we propose to integrate these temporal indicators into NMP or other technological solutions and evaluate their impact in actual learning contexts to update the analytical framework of the thesis.

The contributions of this thesis will have implications at the theoretical, analytical, and teaching practice levels about how to understand the different factors that affect the dynamics of SRL in BL in a holistic manner.

REFERENCES

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—Systematic review (2008–2018). Research and Practice in Technology Enhanced Learning, 15(1), 6.

Broadbent, J. (2017). Comparing online and blended learners' self-regulated learning strategies and academic performance. The Internet and Higher Education, 33, 24–32.

Broadbent, J., & Fuller-Tyszkiewicz, M. (2018). Profiles in self-regulated learning and their correlates for online and blended learning students. Educational Technology Research and Development, 66(6), 1435–1455.

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the Impact of Learning Analytics. Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 446–455.

Fan, Y., Matcha, W., Uzir, N. A., Wang, Q., & Gašević, D. (2021). Learning Analytics to Reveal Links Between Learning Design and Self-Regulated Learning. International Journal of Artificial Intelligence in Education, 31(4), 980–1021.

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

Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2019). From Study Tactics to Learning Strategies: An Analytical Method for Extracting Interpretable Representations. IEEE Transactions on Learning Technologies, 12(1), 59–72.

Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting Learning Strategies with Analytics: Links with Self-reported Measures and Academic Performance. Journal of Learning Analytics, 4(2).

Gašević, D., Mirriahi, N., Dawson, S., & Joksimović, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. Computers in Human Behavior, 67, 207–220.

Graham, C. R., Allen, S., & Ure, D. (2005). Benefits and Challenges of Blended Learning Environments: In M. Khosrow-Pour, D.B.A. (Ed.), Encyclopedia of Information Science and Technology, First Edition (pp. 253–259). IGI Global.

Jørnø, R. L., & Gynther, K. (2018). What Constitutes an 'Actionable Insight' in Learning Analytics? Journal of Learning Analytics, 5(3).

[4] https://nmedsup.cl/

Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. Computers & Education, 172, 104251.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Delgado-Kloos, C. (2018, September). Predicting learners' success in a self-paced MOOC through sequence patterns of self-regulated learning. In European conference on technology enhanced learning (pp. 355-369). Springer, Cham.

Matcha, W., Ahmad Uzir, N., 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.

Michel, C., Lavoué, E., George, S., & Ji, M. (2017). Supporting Awareness and Self-Regulation In Project-Based Learning through Personalized Dashboards. International Journal of Technology Enhanced Learning, 9(2/3), 204–226.

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. Frontiers in Psychology, 8, 422.

Pelletier, K., McCormack, M., Reeves, J., Robert, J., & Arbino, N. (2022). 2022 EDUCAUSE Horizon Report, Teaching and Learning Edition. 58.

Pérez-Álvarez, R. P., Jivet, I., Perez-Sanagustin, M., Scheffel, M., & Verbert, K. (2022). Tools Designed to Support Self-Regulated Learning in Online Learning Environments: A Systematic Review. IEEE Transactions on Learning Technologies, 15(4), 508–522.

Pérez-Álvarez, R., Maldonado, J., & Pérez-Sanagustín, M. (2018). Tools to Support Self-Regulated Learning in Online Environments: Literature Review: 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Leeds, UK, September 3-5, 2018.

Pérez-Sanagustín, M., Pérez-Álvarez, R., Maldonado-Mahauad, J., Villalobos, E., & Sanza, C. (2022). Designing a moodle plugin for promoting learners' self-regulated learning in blended learning. In European Conference on Technology Enhanced Learning (pp. 324-339). Springer, Cham

Pérez-Sanagustín, M., Sapunar-Opazo, D., Pérez-Álvarez, R., Hilliger, I., Bey, A., Maldonado-Mahauad, J., & Baier, J. (2021). A MOOC-based flipped experience: Scaffolding SRL strategies improves learners' time management and engagement. Computer Applications in Engineering Education, 29(4), 750–768.

Reimann, P. (2011). Design-Based Research. In L. Markauskaite, P. Freebody, & J. Irwin (Eds.), Methodological Choice and Design: Scholarship, Policy and Practice in Social and Educational Research (pp. 37–50). Springer Netherlands. https://doi.org/10.1007/978-90-481-8933-5_3

Saqr, M., Jovanovic, J., Viberg, O., & Gašević, D. (2022). Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics. Studies in Higher Education, 1–22.

Shyr, W.-J., & Chen, C.-H. (2018). Designing a technology-enhanced flipped learning system to facilitate students' self-regulation and performance. Journal of Computer Assisted Learning, 34(1), 53–62.

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G.-J., & Paas, F. (2019). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. International Journal of Human–Computer Interaction, 35(4–5), 356–373.

Yoon, M., Hill, J., & Kim, D. (2021). Designing supports for promoting self-regulated learning in the flipped classroom. Journal of Computing in Higher Education, 33(2), 398–418.

Predicting Drop-out in Blended Professional Development

Selina Griffin
The Open University, UK
selina.griffin@open.ac.uk

SUPERVISORS: Bart Rienties (bart.rienties@open.ac.uk)
Luce Withington (luce.withington@open.ac.uk)

ABSTRACT: The use of learning analytics to build predictive models identifying students at risk of not finishing their studies is becoming more common in Higher Education. In professional education, focused on skills rather than knowledge-based education, the practice is still relatively unheard of. My research is based on Toastmasters International (TMI); a global public speaking organization of over 350,000 Members. My Doctorate aims to see whether predictive models can apply in the area of Professional Development; providing data-driven insights into drop-out (retention in 2019/20 was around 54%) to support these volunteer-run Clubs. My first step has been to obtain anonymized data on the UK population of Toastmasters (8,000+) and their interaction with the Learning Management System (LMS). The next step is to combine this with data about their Membership (such as tenure and Club) to evaluate if an effective predictive model can be built. My full study will then attempt to create a model for the entire population.

Keywords: Learning analytics, retention, communications, professional development, retention, training, professional education, predictive analytics, predictive modelling

1 SUMMARY OF MY RESEARCH

1.1 Background and Problems

Whilst there is significant research on the use of predictive models in Higher Education (HE) and to a lesser degree in Massive Open Online Courses (MOOCs) (Phillips and Ozogul, 2020), there is little outside of these fields, including the realm of Professional Development. Like many educational and development organisations outside of formalised education, TMI struggles with retention. The Toastmasters educational programme (Pathways) focuses on learning by doing. Members attend regular Club meetings where they perform roles and give speeches to improve their abilities in leadership and public speaking. The educational material needed to carry out these projects is on Pathways (this is the name of the LMS as well as the program name) which Members must therefore access for the material and to indicate progress. When a particular level (akin to a badge/ certificate in MOOCs) is achieved, this attainment is authorised by a relevant Club Officer before being bestowed.

People join such organisations because they want to improve in the skill that the organisation purports to help its Members develop; not because they wish to achieve a particular educational milestone that exists in the mind of the program authors (Handoko et al., 2019). This aligns the area of Professional Development with perhaps more similarities to MOOC research rather than in formal HE.

An additional complication of TMI is the nature of its decentralised, hierarchical construction. At the top Head Quarters level, paid employees design and deploy educational material, oversee policies, collect Membership fees and provide instruction. Most individual Toastmasters will only ever have a vague knowledge of this ivory HQ tower based in Colorado and instead will know Toastmasters by the volunteer Officers who run their local Club. This dynamic means that, similar to distance learning (Herodootou et al., 2020), the individuals who interact with the learners are removed from the centre where the concept and program originated. Spotting or acting on signs of disengagement or disenfranchisement is down to individuals who may not be trained, engaged, comfortable or incentivised in doing so. In such a context, a centralised data-based predictive model could provide visibility to HQ of what is happening with Members and their engagement with the program on the ground, as well as support and early warning to Officers and Members themselves.

Although my research is looking at this potential application of predictive modelling through the lens of TMI, the problem of retention itself is applicable to many other contexts (other decentralized programs of development or education, more informal learning contexts, even membership). Constructing a predictive model in a less formal educational environment (to those seen in HE such as Herodotou et al., 2019) will also necessitate the use of less typical fields associated with more traditional predictive model approaches; the findings of which could broaden the potential applications of such modelling techniques.

1.2 Research Goals and Questions

My research goal is to see whether we can predict Member drop-out of Toastmasters. In order to progress this goal I have two lines of enquiry open to me depending on the lens that I wish to use for my enquiry. The first is a micro lens -if I wanted to make predictions about an individual, my research would be qualitative and the prediction not based on the presence/absence of certain variables as an ordered and observable event but on the subjective nature of an individual’s interaction and their personal circumstances.

In order to contribute new knowledge of significance I will have to move beyond a Club or local level and adopt the second line of enquiry, applicable to the wider level of TMI. This is where my framework comes in as a critical realist – combining the theory of science (the macro lens) with the complex nature of the social world that we inhabit (Williams et al., 2017). With a micro lens an individual is complicated but with enough people (and a macro lens) trends and patterns in groups can be observed that will indicate if that sub-group is less or more likely than others to perform a certain action. The prediction will not be true for all individuals but it is likely to hold for a large enough proportion that the exercise holds value. Therefore, my aim is to investigate whether predictive models can be used in a Toastmasters context to meaningfully predict Members at risk of not renewing. Specifically:

RQ1: How can we make use of predictive learning analytics in the Toastmasters LMS to identify Members at risk of non-renewal in the next period.

RQ2: To what degree do Member demographics such as tenure, qualifications achieved and strength of Club affect the predictive model – potentially necessitating different models for different clusters, or rendering certain clusters unpredictable.

1.3 Current Knowledge

In formal education, retention is normally the completion of a qualification or programme of study. Outside of this arena, matters are less clear. The arena of MOOCs is a common example; MOOC authors may see retention as the completion of the course (and dependent on the provider model, obtaining a certificate) but it is possible that learners signed up for curiosity or to learn more about a subject with no intention of completing a qualification. In Toastmasters despite the qualification element, arguably this Professional Development experience is more akin to a membership. Members sign up to achieve a goal (such as improve public speaking) and judge their success for themselves. A recent TMI survey of non-renewing Members (594 responded, a response rate of 6.1%) said 89.9% of them felt that Toastmasters had helped them reach their goals and on average 7.7 out of 10 would recommend Toastmasters (Toastmasters, 2020). These statistics appear positive (though of course do not include the views of non-responders which may differ); most Toastmasters who had now left, felt they had been helped to achieve their goals.

The goals that a learner sets for themselves (or had in mind when signing up, as discussed in Jansen et al., (2020) study on self-regulated learning in MOOCS where learners detailed their intentions for the course on sign-up) will therefore have an effect on what MOOC authors would call “completion” rates or retention for TMI (see also Handoko et al., 2019 and Mrhar et al., 2020). TMI themselves define retention as the payment of two consecutive Membership payments (due twice a year). Whilst Pathways completion is encouraged, there is no obligation or enforcement to complete Pathways Levels within time-frames (or in fact at all).

The creation of predictive models has become one of the most common aims of Learning Analytics (LA) and has been expanding from 37 studies in 2011 and 450 in 2018 (Phillips and Ozogul, 2020). Other examinations of case studies (Ferguson and Clow, 2017 and Viberg et al., 2018) may have differed on their definitions and findings regarding positive outcomes from LA but the main conclusions are similar; there was not enough robust evidence in the field or systematic research in this area. With the large numbers involved in TMI, research into predictive modelling here could provide widespread benefits not only for Toastmasters and skills development, but arguably also contribute something of value to this wider evidence base for HE.

1.4 My Contribution

As a Professional Doctorate student I bring my own practice and context to my research. I am a Member and Club Officer (a voluntary position needed to run Clubs at a local level) of Toastmasters International (TMI), an international public speaking and leadership development organisation of over 350,000 Members. My experience and insider status places me in a privileged position to not just access the necessary data for predictive model construction, but also the understanding and sensitivity of the context that has generated this data; both within the Learning Management System (LMS), Pathways and also the Membership tenure data that surrounds it.

There is limited research in the public domain to be found on Toastmasters. The research that does exist (such as Plourde et al., 2018) is mostly descriptive and anecdotal with effectiveness “proven” by opinions from those who enjoyed the programme (led/set up by an enthusiastic individual). For an educational programme to stay relevant and purport to be of high quality, there should be robust and empirical research on it. This is the first obvious gap in the literature review.

Pathways, the Learning Management System, is a new addition to TMI and provides a wealth of data points for analysis and potentially construction of a predictive model to look at Member retention by flagging those Members whose behaviour in the system shows a similar trajectory to others who then dropped out. This illustrates the gap that as the LMS is new, there is no prior research on it. Application of Learning Analytics is also rarely found outside the field of HE (or to a lesser extent, beyond MOOCs) marking another gap my research could contribute to.

These gaps both lead to the first research question: How can we make use of predictive learning analytics in the Toastmasters Learning Management System (Pathways) to identify Members at risk of non-renewal? This research is completely new to Toastmasters and I have the support and a Memorandum of Understanding from TMI themselves. As illustrated previously, this research is also likely to have wider implications for Professional Development and more informal educational spheres in general.

Alongside this, data will allow us to see how Members progress through the program itself (such as total time spent in the system, duration between and order of project completion), which could indicate if there are issues with particular projects or tasks and if Members are behaving as TMI intended. This marks another gap in current knowledge; we do not know how Members interact with Pathways and what this behavior might tell us about their tenure.

Learners are not all the same and this is demonstrably true of Toastmasters Members. Analysis of the data should identify different clusters of Members which could aid our understanding of their behaviour as different Member clusters are likely to act in different ways.

1.5 Research Methodology

My study is going to be a non-experimental quantitative design. Pathways is a new system for Toastmasters which brings with it more data-points than were previously held by TMI. The data for the modelling/quantitative part of my research falls into the category of secondary data. TMI has agreed to give me access to their Pathways data in an anonymized format. For the pilot study, I have received the UK dataset (c.8,000 active Members) and a number of specific data points summarized in Table 1. My experience as a Toastmaster leads me to some hypotheses, combined with my understanding of the available LMS and TMI Membership data has focused my request on data points relating to: membership tenure and qualifications achieved (someone who has completed one Pathway (levels 1-5) may decide there is no more to learn, someone who has been a Member for fifteen years is perhaps less likely to leave as it is likely to have become part of their social life); Club joined (Clubs have various levels of quality); and time between various educational achievements (someone progressing quickly may have a specific aim in mind and leave when they achieve it, someone taking months between activities lacks momentum and commitment and may drop-out).

Table 1: summary of data fields obtained from TMI

Field	Data type
Id	Unique id field

Current Member	Y/N
Member since	Date
Last Membership renewal	Date
Enrolled in Pathways	Y/N
Highest Pathways level obtained	1-5
Date Level 1 Completed	Date

The full list of fields requested is given in the Appendix as Table 2 and will give measures of a Member's Toastmaster history. I will also need to construct a number of additional variables for my data based on calculations between supplied fields for my models (such as time between dates, total Pathways levels completed and years of Membership).

I aim to initially use SPSS (IBM statistics software) for my modelling. It is a program that I have access to via the university and I have previously (successfully) used it to produce predictive models (via binominal logistic regression) to find charitable donors in a large dataset (250,000+ individuals).

Once my pilot study is complete and I have a preliminary model, I aim to conduct a small number of interviews (study 2) of participants whose data in the model is of particular interest – due to being an outlier or an average example. TMI will have kept an index translating the id they have supplied me in my anonymized set so that they can go back to the individual it relates to and then pass on a message from me asking if they would be willing to volunteer for an interview. If they consent and respond, I will then be able to explore the behaviour I saw in the data to see if this insight helps me to better understand and ultimately improve my model.

For my main study I will construct a model for the whole TMI database (350,000+ active Members) and ultimately hand this over to TMI themselves for their use. Whether this requires the exact same fields to be collected or if more/alternative fields are needed will depend on the findings of my pilot model and subsequent participant interviews.

1.5.1 Ethical Considerations

The pilot data I received is anonymized. The interview portion (study 2) of my research will change the ethical dynamics as these individuals will make themselves, their opinions and behaviors known to me; I will have to ensure proper informed consent is obtained and that my interviewer position as a Toastmaster Member working on behalf of TMI does not influence responses as much as I am able to.

My position as a practitioner in my own research space brings with it advantages. I have an understanding of the context Toastmasters operate in and some of the nuances behind data-points due to my position as an insider. However, this may bring a particular bias to my study – I am a practitioner therefore I think that my area of practice (Toastmasters) is inherently worthy and positive and I have my own hypotheses formed by my experience. Making use of assistance from my supervisors and the LAK23 Doctoral consortium will help to review my model with objectivity.

1.6 Current Status

I am working through a Professional Doctorate Programme at the Open University. I have completed the two year taught phase of the program in which I have written up draft literature review, methodology and research methods sections, upgraded to the research phase, obtained ethical approval for my pilot and received my initial set of data from TMI. By the time of LAK23 I aim to have developed a predictive model, be writing-up my pilot and considering the full study. I am seeking support from others interested in predictive analytics in professional development, considerations when moving from a small pilot to a larger dataset, critique and feedback on my pilot model and the experiences of those who have used qualitative interviews to supplement their models.

REFERENCES

Ferguson, R. and Clow, D. (2017) ‘Where is the evidence? A call to action for learning analytics’, *ACM International Conference Proceeding Series*, p. 56–65, doi: 10.1145/3027385.3027396.

Handoko, E., Gronseth, S. L., Mcneil, S. G., Bonk, C. J. and Robin, B. R. (2019) ‘Goal Setting and MOOC Completion : A Study on the Role of Self-Regulated Learning in Student Performance in Massive Open Online Courses’, *International Review of Research in Open and Distance Learning*, 20(3).

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z. and Hlosta, M. (2019) ‘A large-scale implementation of predictive learning analytics in higher education: the teachers’ role and perspective’, *Educational Technology Research and Development,* doi*:* 10.1007/s11423-019-09685-0

Herodotou, C., Naydenova, G., Boroowa, A., Gilmour, A. and Rienties, B. (2020) ‘How can predictive learning analytics and motivational interventions increase student retention and enhance administrative support in distance education?’, *Journal of Learning Analytics*, 7(2), doi: 10.18608/jla.2020.72.4

Mhrar, K., Douimi, O. and Abik, M. (2020) ‘A Dropout Predictor System in MOOCs Based on Neural Networks’, *Journal of Automation, Mobile Robotics and Intelligent Systems*, 14(4), doi: 10.14313/JAMRIS/4-2020/48

Jansen, R., van Leeuwen, A., Janssen, J., Conijn., R and Kester, L. (2020) ‘Supporting learners' self-regulated learning in Massive Open Online Courses’, *Computers and Education,* 146, doi: 10.1016/j.compedu.2019.103771

Phillips, T. and Ozogul, G. (2020) ‘Learning Analytics Research in Relation to Educational Technology: Capturing Learning Analytics Contributions with Bibliometric Analysis’, *TechTrends. TechTrends*, 64(6), pp. 878–886, doi: 10.1007/s11528-020-00519-y.

Plourde, J. M. H., Purdy, S. C., Moore, C., Friary, P., Brown, R. and McCann, C. M. (2018) ‘Gavel Club for people with aphasia: communication confidence and quality of communication life’, *Aphasiology,* doi: 10.1080/02687038.2018.1453043.

Toastmasters (2020) ‘Non-renewing member survey July 2020’. Available at: http://www.toastmasters.org/leadership-central/statistics-and-data-hub (Accessed: 14.11.20)

Viberg, O., Hatakka, M., Bälter, O. and Mavroudi, A. (2018) ‘The current landscape of learning analytics in higher education’, Computers in Human Behavior, 89. doi: 10.1016/j.chb.2018.07.027

Williams, L., Rycroft-Malone, J. and Burton, C. (2017) ‘Bringing critical realism to nursing practice: Roy Bhaskar's contribution’, *Nursing Philosophy*, 18(2), doi: 10.1111/nup.12130.

APPENDIX

Table 2: Data points for study 1

Field	Subfield
Some sort of id (that can be backtracked by TMI)	
Member since (date)	(Calculated field on membership length)
Current member (Y/N)	Last renewal date
Club(s)	Age of club Corporate sponsorship Y/N
Pathways initiated (Y/N)	Date
Pathway(s) selected	(name of Pathway)
Level 1 completed (Date)	(Calculated field pathways date and level 1 date)
Level 2 completed (Date)	"
Level 3 completed (Date)	"
Level 4 completed (Date)	"
Level 5 completed (Date)	"
Highest Pathways level achieved	"
Date of most recent project completed	Calculated field most recent project completion to now)

Competency-based summative assessment through online simulation leveraging analytical methods and techniques

Author: Ágnes Meleg
Corvinus University of Budapest
agnes.meleg@uni-corvinus.hu

ABSTRACT: Competence assessment is challenging, especially in a form of summative assessment, measuring multiple competencies simultaneously. Due to the imbalanced student-teacher ratio and the variety of programs, competence assessment can be highly demanding in higher education. The goal of this research is to develop a scalable, simulation-based approach and assess students' abilities and 21st century skills, relying on learning analytics and educational data mining techniques, to get insight into student competencies. After having the particular method developed – a 3-hour long online simulation with embedded student surveys –, 160 students completed the assessment in two waves. Leveraging the trace data generated during the simulations and the survey data (self- and peer assessment), several different analyses are being conducted, building on descriptive statistics, clustering, and structural equation modeling (SEM), in combination with qualitative research (focus group, interviews). Based on the preliminary findings the solution gives valuable self-reflection and learning experience for students, and beneficial insight for teachers about student competencies.

Keywords: Doctoral Consortium, Competence Assessment, Summative Assessment, Online Simulation, Learning Analytics

1 BACKGROUND

Competence assessment is resource-intensive and challenging, especially in higher education (HE) due to different programs, learning paths, goals, and student profiles. On the other hand, considering the competency-based approach both in the educational and working environment, there is an increasing need for competence assessment, with extreme attention to 21st century skills, such as communication and teamwork.

The current study seeks to close an existing gap in competence assessment in higher education presenting a scalable, semi-automated, simulation-based approach, combined with learning analytics and educational data mining methods and techniques, using trace data and self-reported data.

2 GOAL OF THE RESEARCH

There is a lot of uncertainty and confusion around the definition of competence and competency. In this research – building upon the work of Moore et al. (2002) – the author refers to *competence* as an area of work (macro level), and to *competencies* as attributes (knowledge, abilities, and skills) that underlie successful professional performance (micro level).

Even though various methods have been introduced to assess student competencies – such as survey-based self-assessment (for example, Lachmann & Nilsson, 2021), peer assessment, supervisor assessment (Liang et al., 2020); multi-participant assessment through exercises (González-Marcos et al., 2016; Brilingaitė et al., 2020); assessment through sensors (Dominguez et al., 2021) –, there is not one single well-accepted way, as all approaches have both strengths and limitations, motivating the researchers to look for new possibilities.

With this current work through embedded case studies, the researchers present a competence assessment method addressing certain limitations of previous works and suggesting novelty in terms of the simulation itself and the use of learning analytics methods and techniques to analyze the (trace) data generated during the simulations for assessment purposes (in combination with other data sources – e.g., peer review).

3 RESEARCH QUESTIONS AND METHODOLOGY

In the research, both qualitative and quantitative methods are applied, following the guidance of Viberg et al. (2018). In the first phase surveys were used for data; while in the second and third phases experimental simulation with embedded surveys, interviews and focus group were applied, combined with different analytical methods.

As part of the research a particular simulation has been developed for Management Information Systems Bachelor Students, mapping the competencies set in the program (based on centrally defined regulation: Hungarian Ministerial Decree: 18/2016. VIII. 5. – See extract in Appendix), and adding 4C (communication, collaboration, critical thinking, creativity) components to the exercises, as generic skills are essential based on employers' feedback (e.g., Burns et al., 2018; Cummings et al., 2020).

In terms of analytical solutions, descriptive statistics and clustering have been applied, and structural equation modeling (SEM) is under progress in the current work.

Table 1: Research methodology

Main research areas	Methods
Get insight into student competencies	
How do students see themselves and each other compared to the simulation results?	Descriptive statistics, Clustering, Interviews, Focus group
What competencies should the students improve?	Descriptive statistics
What student groups could be distinguished based on the assessment?	Clustering
Investigate how best to provide students with feedback about their result	Focus group
Learn about the relations that may exist among the different competencies	Structural equation modeling

4 CURRENT ACHIEVEMENTS

4.1.1 Competency-based summative assessment through online simulation

Following the previously published framework (Meleg & Vas, 2020), as a first step, we examined the possibilities of assessing student competencies in HE. Besides studying previous research works and interviewing experts in the field, the researchers launched a survey targeting the main stakeholders: students and teachers, focusing on the Management Information Systems Bachelor Program of the University. 132 graduating students and 22 professors participated in the survey covering several questions connected to competence assessment in HE. The main outcome was that there was a surprisingly high interest from both sides towards student competence assessment in the program: 92.4% of the students stated that competence assessment would be beneficial (55.7% of the students interested in this opportunity also provided their contact details for a future pilot assessment); similarly, 95.5% of the professors stated that competence assessment would be beneficial for students.

Based on the positive outcome, the researchers started the work to design a student competence assessment method considering the following: i) students and professors' opinions, ii) professional experience and opinion of experts of the field, iii) previous works related to competency-based student assessment.

Considering the fact that graduating students are very close to starting their careers, when talking about their summative assessment, it is suggested to move away from the traditional educational assessment methods (exams, tests), and get closer to practices of the working environment. In many cases assessment center (AC) is part of the recruitment process in the working environment to test candidates' behavior, skills, and abilities. Traditional assessment centers are not commonly used in educational settings due to their resource-intensive nature. In the current study, the researchers applied an online simulation approach, which is – similarly to the AC – suitable for simultaneous assessment of several people performing simulated work exercises, and at the same time, it is less resource-intensive.

The exercises and the platforms were designed and created in a 2-month period thanks to the iterative, collaborative work, and to the involvement of business professionals. During the assessment, generic 4Cs and program-specific competencies are measured with a mapping method (Figure 1).

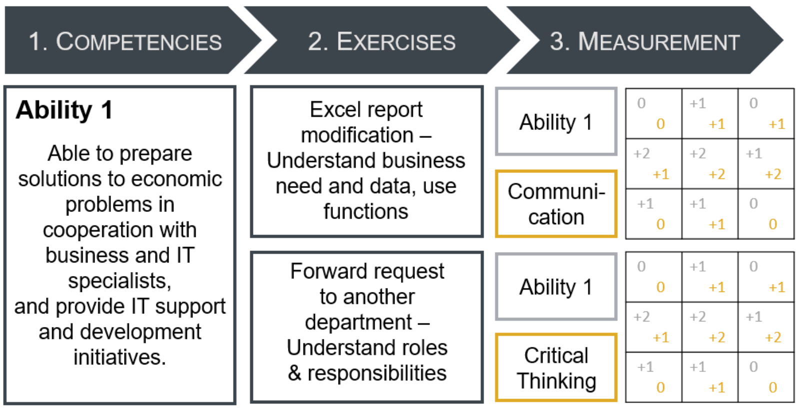


Figure 1: Schematic example for competency - exercise - score mapping

As a result, a 3-hour online simulation event was conducted in 2021 as a pilot assessment to evaluate student competence against the expectations set in the program, and it was followed by another in 2022. In total 160 students participated (80-80 in both case studies). All students gave their consent to analyze their data – considering ethical concerns.

The students received a guideline for the simulation, explaining to them the simulated work environment and what to expect during the event. According to the scenario, the students work for a fictional company, *Vision Consult*, as Business Informatics Consultants in the EMEA region. They work in small teams, in various locations, receiving requests on a daily basis from colleagues and from clients. Their task and goal are to give solution to all the client requests.

The requests arrived at the simulated CRM system, from four imaginary companies. In addition, one internal request was sent to their inbox from an "HR colleague". Some of the requests/exercises were connected to each other, therefore the participants had the chance to familiarize themselves with the personas, understand their business needs, and get an insight into the different industries. With one exercise several competencies could be assessed simultaneously, and one competency could be assessed through several exercises (many to many relations).

In the first stage, the students had 1 hour to work on 10 exercises, individually, after that they received access to an interactive platform where they had to assess what they would do or answer to the requestor based on their preliminary work. They could always choose from 9 options; every option meant different score combinations, considering the measured competencies (Figure 1). In total students could earn a maximum of 64 points (40 coming from ability scores and 24 from 4C skills).

In the second stage, the participants were asked to work in their teams to discuss the requests and decide together which option they would go with prior to a common agreement. Finally, in the third stage, the students were asked to evaluate themselves and their team peers (the questionnaire statements were created according to the Hungarian Ministerial Decree: 18/2016. VIII. 5., and based on previous works, e.g., Hinyard et al., 2019). At the end of the event, in the last survey, the participants had the chance to give their opinion about the event itself.

The characteristics of the assessment could be summarized as the points show below:

- Thanks to the design, it could be easily repeatable, only changing the request numbers, and/or the personas, and/or the order of the options in the matrices;
- It is feasible on a larger scale thanks to the online platforms and the ETL processes for automatic evaluation;
- Multiple competencies could be evaluated at the same time;
- It is enjoyable and beneficial for students considering the simulation set-up.

4.1.2 Results through data analytics

For the analysis the researchers had the following data sources: i) digital activity and interaction data (MS Teams), ii) data from the matrix choices (Qualtrics), iii) self-assessment questionnaires (Qualtrics), iv) peer assessment questionnaires (Qualtrics), v) mapping table of choice options, exercises, and competencies (MS Excel).

The researchers first conducted data analysis in MS Excel and R Studio to have a first insight into the results. Firstly, the overall scores were calculated on an individual level. Secondly, detailed, competency level scores were calculated on an individual level. Thirdly, overall scores and detailed scores were calculated at a team level. Finally, self- and peer assessment scores were calculated on an individual level, aggregating the peer values. The individual results show normal distribution in case of both assessment events (Shapiro-Wilk normality test: i) p-value = 0.6917, ii) p-value = 0.3764).

As a next step, clustering and multidimensional scaling (MDS) were applied to analyze the classes and to get an insight into students' profile. Three data sets were analyzed next to each other: i) data from the results of the simulation on an individual level, ii) data of the self-assessment, iii) data of the peer assessment. In all cases, the same steps were followed. Data set was loaded into R 4.0.2 from a CSV file, it was cleaned, selecting the necessary data fields, and finally, data were standardized. Since our variables are numeric performance scores, they were standardized to have a mean of 0 and a standard deviation of 1. Both the k-means algorithm and hierarchical clustering were applied. Clusters were identified from all three data sets, and results were compared to each other.

The main results are summarized as follows. Through the assessment, the researchers could get an insight into students' competencies and could compare the results of the two different classes (students who participated in the first event vs. students who participated in the second). Also, the main competency-specific improvement areas have been identified in both cases. We could see that even though we talk about students of the same program, their overall competency-set and profiles look slightly different. In addition, we learned that students could not assess themselves and each other as precisely as the simulation itself. Thanks to the clustering methods, four student groups were identified in both classes based on their assessment results, which could be leveraged in many ways – for example, by mixing the students across the clusters when creating teams for project works in their final semester.

Moreover, we could notice that with teamwork the students could achieve better results; at the same time, it should be highlighted that by the time of the team section, the students were more familiar with the concept and the exercises – as a result, the simulation also served as a learning experience for the participants. It is also notable that students appreciated the opportunity and university effort. They rated on a 0-100 analog scale the first event as follows: "The competence assessment through online simulation was..." difficult: 74.8, beneficial: 56.1, relevant: 61.5, and enjoyable: 53.7; whereas, the second event: difficult: 72.3, beneficial: 76.7, relevant, 77.9, and enjoyable: 72.2. This also means that as designers the researchers could improve the assessment event; and many positive comments were received as well from the participants, for example:

> „The problems were relevant, quite difficult and enjoyable. It felt like a real working environment with the teams and the individual work. All-in-all a fun and useful event. :)".

A paper has been submitted to *Educational Measurement: Issues and Practice*, describing in detail the assessment itself as well as the results of the descriptive analysis and the clustering. Since the method is generic, it could be applied in other programs of any other educational institution.

5 ONGOING AND FUTURE WORK

5.1.1 Focus group for student dashboard in the light of Human-Centered Learning Analytics

To provide students with valuable feedback in a form of a student dashboard, a student focus group was organized after the second assessment event. The students participating in the focus group received their assessment results in the current MS Excel-based printed PDF format as a starting point.

Through the moderated conversation, we could learn about student preferences and understand what would be valuable for them as feedback (e.g., results compared to their peers in an anonymous way). Based on the information, the researchers started to work on an enhanced dashboard in Tableau – with students, for students – in the light of Human-Centered Learning Analytics (HCLA).

5.1.2 Structural equation modeling for competency relation evaluation

The aim of the third research area within the whole research is to explore the (potential) relations among the competencies, e.g., how communication/collaboration/critical thinking/creativity relate to project management. The intention is to conduct SEM analysis in ADANCO, and at the same time, we are investigating further analytical methods for additional insights.

REFERENCES

Brilingaitė, A., Bukauskas, L., & Juozapavičius, A. (2020). A framework for competence development and assessment in hybrid cybersecurity exercises. Computers & Security, 88, 101607. https://doi.org/10.1016/j.cose.2019.101607

Burns, T., Gao, Y., Sherman, C., & Klein, S. (2018). Do the Knowledge and Skills Required by Employers of Recent Graduates of Undergraduate Information Systems Programs Match the Current ACM/AIS Information Systems Curriculum Guidelines? Information Systems Education Journal, 16(5), 56–65.

Cummings, J., & Janicki, T. N. (2020). What Skills Do Students Need? A Multi-Year Study of IT/IS Knowledge and Skills in Demand by Employers. Journal of Information Systems Education, 31(3), 208.

Dominguez, F., Ochoa, X., Zambrano, D., Camacho, K., & Castells, J. (2021). Scaling and Adopting a Multimodal Learning Analytics Application in an Institution-Wide Setting. IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, 14(3), 400–414. https://doi.org/10.1109/TLT.2021.3100778

González-Marcos, A., Alba-Elías, F., & Ordieres-Meré, J. (2016). An analytical method for measuring competence in project management. British Journal of Educational Technology, 47(6), 1324–1339. Scopus.

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 and the Health Professions, 42(4), 450–472.

Lachmann, H., & Nilsson, J. (2021). Dual use of instruments for assessing nursing students professional- and clinical competence. Nurse Education Today, 96, 104616. https://doi.org/10.1016/j.nedt.2020.104616

Liang, H. Y., Tang, F. I., Wang, T. F., & Yu, S. (2020). Evaluation of Nurse Practitioners' Professional Competence and Comparison of Assessments Using Multiple Methods: Self-Assessment, Peer Assessment, and Supervisor Assessment. Asian Nursing Research. https://doi.org/10.1016/j.anr.2020.10.004

Meleg, Á., & Vas, R. (2020). Towards Evidence-Based, Data-Driven Thinking in Higher Education. In A. Kő, E. Francesconi, G. Kotsis, A. M. Tjoa, & I. Khalil (Eds.), Electronic Government and the Information Systems Perspective (pp. 135–144). Springer International Publishing. https://doi.org/10.1007/978-3-030-58957-8_10

Moore, D. R., Cheng, M., & Dainty, A. R. J. (2002). Competence, competency and competencies: Performance assessment in organisations. Work Study, 51(6), 314–319. https://doi.org/10.1108/00438020210441876

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

APPENDIX

Hungarian Ministerial Decree: 18/2016. VIII. 5. – Extract (translated by the Author)

Abilities as an outcome for students graduating from the Management Information Systems Bachelor Program:

- Able to prepare solutions to economic problems in cooperation with business and IT specialists, and provide IT support and development initiatives;
- Able to understand and analyze business processes, create requirement specifications for software applications, and perform basic programming tasks;
- Able to help adaption of economic applications, initiate organizational changes required for the introduction of IT applications, and cooperate during implementation;
- Able to explore the operating conditions of applications, weigh and communicate the benefits, threats and risks in real business, organizational environment;
- Able to perform database management related tasks, and execute basic data migration tasks;
- Able to apply system development principles and methods, and use development tools (business modeling and computer-aided development tools);
- Have the ability to explore and research problems specific to business informatics, and identify and collect the necessary resources to address them;
- Capable of operating economic applications and providing user services;
- Able to plan and manage smaller development projects;
- Able to resolve IT conflict situations in economic environment.

Promoting Effective Scaffolding for Self-Regulated Learning Strategies

Tongguang (Tony) Li
Monash University
tongguang.li@monash.edu

ABSTRACT: Self-regulated learning (SRL) is a concept that describes how a learner controls, monitors, and regulates their learning process. Being able to effectively choose and use learning strategies is crucial to becoming a skilled, self-regulated learner. Researchers aim to help learners improve their self-regulated learning and corresponding learning strategies by implementing learning analytics. This research aims to address the current issue of scaffolding on learning strategies by implementing state-of-the-art temporal and sequential analysis to identify learning signatures for each learner, which will reflect how learners use, regulate, and monitor their learning and their use of learning strategies. Learning signatures will subsequently be used for identifying the learner's zone of proximal development (ZPD) in terms of learning strategies. Therefore, it is hypothesized that when the design and implementation of scaffolding is informed by the creation of learning signatures and by the identified ZPD, the adaptivity of scaffolding will be improved.

Keywords: SRL, Scaffolding, Zone of proximal development, Learning signature

1 INTRODUCTION

Self-regulated learning (SRL) is a concept that describes learners' cognitive, motivational, and emotional aspects of learning and has been widely considered and applied in educational studies (Panadero, 2017). Whether learners can successfully and effectively self-regulate their learning is dependent on their ability to use, modify, and maintain appropriate learning tactics and strategies (Lim et al., 2021) and learners need to be properly trained to be able to effectively use, organize, and regulate learning strategies (Matcha et al., 2019). Learning analytics (LA) is a promising research field which aims at improving learning experiences through the analysis of learning data, which could be considered as one of the technologies to promote learners SRL skills. However, limited attention has been emphasized on how LA-based scaffolding could facilitate improvements on learners' SRL, especially on learners' effective use of SRL strategies and tactics. As such, to promote the practical meaning of LA, more studies are expected to promote effective instructional practices and interventions (Gasevic et al., 2015).

Given that the SRL is an informative concept that explains the learning process, and that learning strategies and tactics are interconnected and embedded in the SRL model, the overall purpose of this research is to consider how to design and implement actionable and effective interventions to scaffold learners' use of learning strategies and tactics, as well as to improve learners' SRL skills so that learners can effectively regulate their use of learning strategies and tactics.

2 BACKGROUND

2.1 Learning tactics and learning strategies

Learning strategies and learning tactics are two concepts that are constantly used in both general educational psychology and the field of learning analytics, though researchers are yet to reach consensus on their conceptualizations. To date, there are many terms for describing and explaining how learners learn and how to learn effectively, including but not limited to, learning strategies (Weinstein et al., 2011), learning approaches (Biggs, 1993), learning tactics (Winne, 2013), learning skills (Winne, 2013), and learning techniques (Dunlosky et al., 2013). However, each direction is conducted separately, and has resulted in divergent views among each scholar. Some scholars use the same term but explain it in different ways, while others may use different terms but are in fact describing the same understanding.

When examining the various empirical studies that focus on the identification, interpretation, or instruction of learning strategies, there are two main conceptualization ambiguities. First, it is noted that different researchers tend to adopt different definitions in different empirical studies (Gasevic et al., 2017; Fincham et al., 2018). This could result in weakened comparability of findings among different empirical studies as different conceptualisations present distinct interpretations. Second, most learning strategy-related empirical studies did not provide clear rationale on why certain a conceptualisation was chosen instead of others. For this reason, researchers from many empirical studies tend to adopt a broader definition so that to avoid potential ambiguities and inconsistencies (Fincham et al., 2018). This tendency of avoiding specific definitions or to escape from providing clear rationales could result in reducing interpretability of their research findings. Therefore, it is imperative to provide a clear review on how learning strategies and tactics are conceptualised in the literature so as to improve comparability and interpretability among different empirical studies. As such, the first research question (RQ1) is: **How are learning strategies and learning tactics conceptualised and measured in the literature?**

2.2 Learning signature

The term "learning signature" is used to describe how a learner learns by reflecting their use of learning strategies and tactics that are captured by trace data (Winne, 2022). The word "signature" is used here to imply that each learner should demonstrate a unique pattern of learning processes, just as individual's signature is different from others. The main purpose of creating learning signatures is to present an all-round snapshot of a learner's SRL skills from a specific learning task so that researchers are better informed in understanding what and how learning strategies can be improved. It is expected that a learner's learning signature could provide a holistic and accurate picture of the learner's cognitive and metacognitive processes during a learning task, which would potentially allow researchers to discover fine-grained-level 'spots' where intervention and feedback should be designed for. Meanwhile, the representation of learning signature can be framed based on the SRL model developed by Winne and Hadwin (1998). There are three constructs that explain Winne and Hadwin's SRL model, and are summarized as three acronyms – COPES (Winne & Hadwin, 1998), AEIOU (Winne & Marx, 1989), and SMART (Winne, 2017) (see the Appendix for the relationship among the three models). It is anticipated that learning signatures will represent how learners choose, use, and modify learning strategies and tactics in specific learning tasks. Specifically, being able to productively employ

SRL means being able to accurately evaluate the learning context in which cognitive and metacognitive operations are applied, being able to strategically choose effective learning strategies and tactics, being able to be reflective in judging the extent to which applied strategies and tactics are effective, and being able to be adaptive to calibrate or improve their cognitive and metacognitive operations. It is therefore expected that learning signature could model 1) how learners enact learning strategies and tactics on the information (i.e., learning content) and 2) how learners metacognitively choose, adopt, and adjust their learning strategies and tactics. As such, the second research question (RQ2) is: **How to create learning signatures based on the strategies and tactics evident in learner data.**

2.3 Zone of proximal development

Once learning signatures are created for each learner, the next step involves clarification of what and to what extent the scaffolds should be provided. To address this question, the zone of proximal development (ZPD) is a promising theoretical construct which could guide the direction of scaffolding. ZPD is defined as "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with a more capable other (Vygotsky & Cole, 1978; p. 86)." The key idea derived from the concept of ZPD is that skills such as learning strategies or learning tactics that are taught outside the learner's ZPD will be less likely to be accepted or be encouraged to use by learners. For example, if trace data indicate that the learner has predominantly used a low level of cognitive learning such as rehearsing or rereading, immediate suggestions of using metacognitive level strategies such as self-evaluation or organization might be too difficult for the learner to learn.

Some learning analytics researchers have recently begun using process mining techniques to identify and cluster learners into different performance groups, and compare learning patterns between the highest performing group and the lowest performing group (Bannert et al., 2014; Saint et al., 2022). However, this research direction is limited to making practical inferences such as directly suggesting detected learning strategies (i.e. the "ideal" activities) from more successful learners to less successful learners (Bannert et al., 2014). It is therefore expected that more studies will address this limitation and propose and design more adaptive and practical scaffolding for interventions. This proposal will therefore attempt to address this concern by introducing the theory of ZPD to create and evaluate an algorithm that will produce more adaptive and actionable scaffolding and feedback. Specifically, ZPD refers to the zone of a learner's cognitive ability within which corresponding learning strategies or tactics are learn-able and usable to the learner. As such, the third research question (RQ2) is proposed as: **To what extent can we identify ZPD reliably for each individual learner with respect to their use of learning strategies and tactics?**

2.4 Just-in-time and just-in-case scaffoldings

Implementing personalised scaffolding while learners are engaged in the learning task is a promising research direction, as Pardo et al. (2019) concluded that LA-based personalised feedback positively correlates with academic performance and because of the improved quality of LA-based feedback, learners' satisfaction with LA-related support has improved. However, it is still unclear whether and to what extent personalised scaffoldings would trigger learners' cognitive or metacognitive adjustments to their use of learning strategies. Creating learning signatures provides opportunities for researchers to investigate how and to what extent learners apply ZPD-informed personalized instruction. This doctoral research is therefore designed to evaluate the effectiveness of ZPD-informed personalised scaffolding and to investigate to what extent learners will adopt the instructions. As such, the last research question (RQ4) is proposed as: **To what extent could the LA-based scaffolding effectively support learners in the adoption and use of learning strategies and tactics?**

3 METHODOLOGY AND PROCEDURES

3.1 Current projects and data collection

Two major projects are involved for this doctoral proposal – "Facilitating Self-Regulated Learning with Personalised Scaffolds on Student's own Regulation Activities (FLoRA)" project funded by UK's Economic and Research Council, Dutch Research Council (NWO), and German Research Foundation (DFG) and "Data analytics-based tools and methods to enhance self-regulated learning" funded by the Australian Research Council as a Discovery Project 2022 (ARC DP22).

All data collections are conducted in a technology-enhanced learning (TEL) environment (see the Appendix for the interface of the TEL) where tools such as annotation, highlighting, time-planning, and feedback in the mode of just-in-time auto-generated scaffolding are provided. The ethical approval for both projects (the FLoRA project and the ARC DP22 project) were granted by Monash University. The data collection process includes three parts – a pre-test session, a writing session, and a post-test session. The pre-test session involves a survey that asks participants to complete, and a questionnaire which collects information about participants' pre-knowledge, background relevancy, self-efficacy, motivation, and English proficiency. The writing session consists of two hours where participants are expected to complete a 300 – 400 words essay. Learning trace data (mouse click, navigation log, and keyboard stroke) will be collected while participants are engaged in the task. Lastly, the post-test session includes a transferring test, a post-test, and an interview were conducted in the post-test session. Types of data that will be used to answer each research question is summarized in Table 1.

3.2 Methodology

The proposed methodologies for each research question are concluded in Table 1. In detail, the study 1 will be conducted as a narrative literature review. The Study 2 involves utilizing process mining to model the learning signature. It is expected that the process mining analysis will visualize each learner's sequence of learning tactics/strategies used. In addition, the meta-data includes survey data that represents learners' internal factors, including efficacy expectation, outcome expectation, incentives, etc. Lastly, the ideographical approach will be used to model learning signatures for each learner. For Study 3, same data will be used as in the Study 2, and clustering analysis will be conducted to identify groups of learners who demonstrate a similar learning signature. Finally, Study 4 involves collecting new data (scaffolding included) and modeling the learning process of the learner after the scaffold has been received. Whether and to what extent learners will calibrate their use of learning strategies and tactics will be examined so that to inform the effectiveness of scaffolding is informed by learners' ZPD.

Table 1： The data collection and methodology overview

Study 1	
Data type	Subjective selection of four strands of conceptualisation.
Methodology	Narrative literature review
Study 2	
Data type	Trace data collected from log file (no scaffoldings)
	Meta-data collected from surveys
Methodology	Process mining analysis based on the Trace-SRL measurement protocol
Study 3	
Data type	Task score
	Trace data collected from log file
	Meta-data collected from surveys
Methodology	Clustering analysis
Study 4	
Data type	Trace data collected from log file (scaffoldings included)
	Meta-data collected from surveys
Methodology	Process mining analysis based on the Trace-SRL measurement protocol

4 PRELIMINARY FINDINGS AND PUBLICATIONS

Preliminary findings for answering RQ1 are summarized in Figure 1. The preliminary finding indicates that there are both convergence and divergence among the four strands.

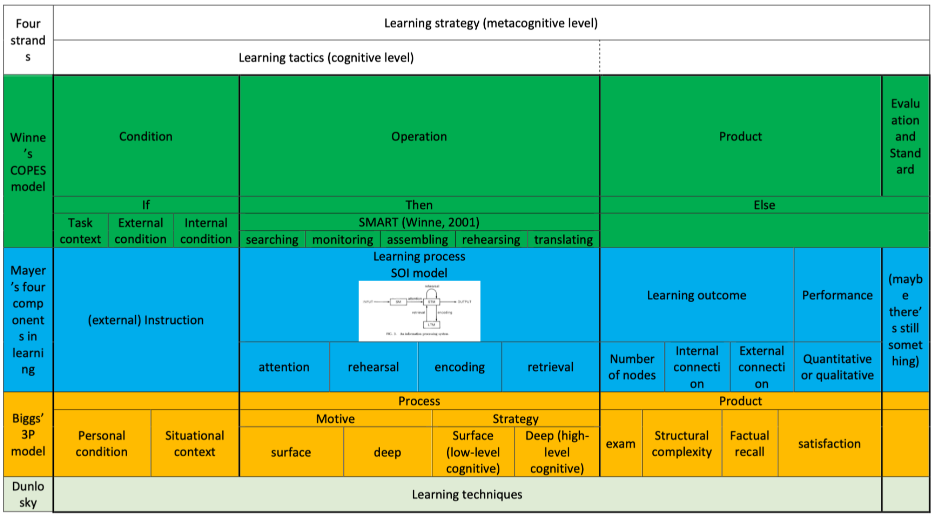


Figure 1: Four strands of conceptualization -- preliminary comparison

To answer RQ2 and RQ3, the data collection is currently in progress through the TEL tailored for the ARC DP22 project and is expected to be finished by the end of October, 2022. In detail, trace data will be analyzed to reveal how learners operate on the platform, how they self-regulate their learning, and what learning strategies and tactics will they use without scaffoldings. To address the RQ4, a qualitative study is conducted which aims at analyzing interview data to obtain insights about learners' perception and experiences from scaffolding (see the second publication below).

All publications related to this doctoral proposal are summarized below:

- One conference paper is submitted to the 13th International Conference on Learning Analytics and Knowledge (LAK'23) – first author. Title: **Analytics of Personalized Scaffoldings for Self-regulated Learning: Effects on Learning Processes**.

- One journal paper is in progress and is related to the RQ4 (the proposal has been accepted as a special issue in the Journal of Computer-Assisted Learning) – second author. Title: **When and why learners benefit from personalised scaffoldings for self-regulated learning: a qualitative study**.

- One journal paper submitted to the Journal of Computer Assisted Learning -- joint second author. Title: **Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-source writing task**.

- One paper is in progress and is related to the RQ1– first author. Title: **What do we mean by learning strategies: a narrative review.**

REFERENCES

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning*, *9*(2), 161-185.

Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British journal of educational psychology*, *63*(1), 3-19.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the public interest*, *14*(1), 4-58.

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.

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

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71.

Gašević, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113-128.

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of chiropractic medicine*, *5*(3), 101-117.

Lim, L., Bannert, M., Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., ... & Gasevic, D. (2021). Temporal assessment of self-regulated learning by mining students' think-aloud protocols.

Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 461-470).

Mayer, R. E. (1988). Learning strategies: An overview. *Learning and study strategies*, 11-22

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, 422.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalized feedback. *British Journal of Educational Technology, 50*(1), 128-138.

Saint, J., Whitelock-Wainwright, A., Gašević, 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.

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers and Education: Artificial Intelligence*, 100060

Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-regulation and learning strategies. *New directions for teaching and learning*, *2011*(126), 45-53.

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

Winne, P. H., & Marx, R. W. (1989). A cognitive-processing analysis of motivation within classroom tasks. *Research on motivation in education*, *3*, 223-257.

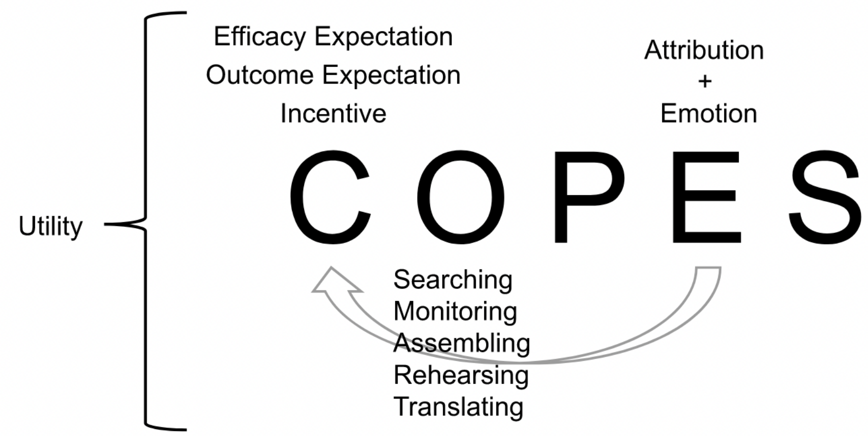
Winne, P. H. (2013). Learning strategies, study skills, and self-regulated learning in postsecondary education. In *Higher education: Handbook of theory and research* (pp. 377-403). Springer, Dordrecht.

Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In *Handbook of self-regulation of learning and performance* (pp. 36-48). Routledge.

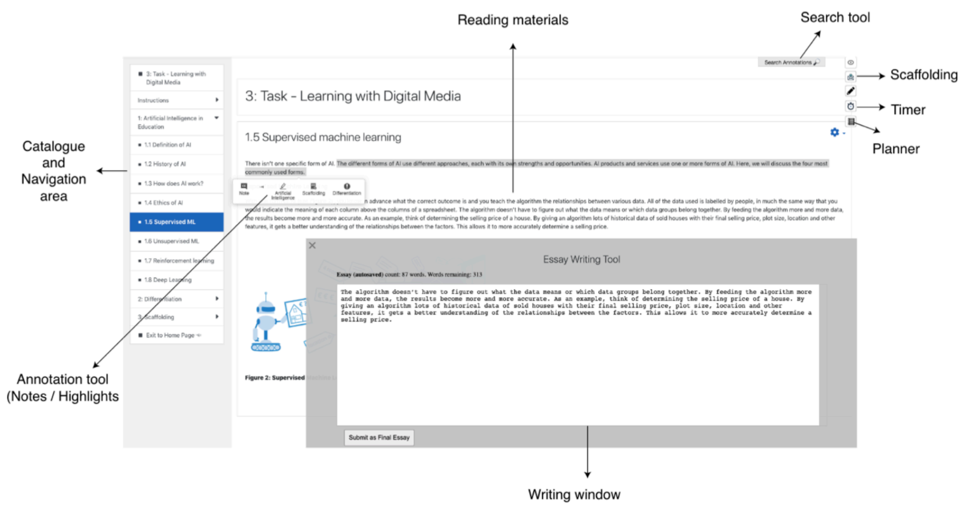
Winne, P. H. (2022). Modeling self-regulated learning as learners doing learning science: How trace data and learning analytics help develop skills for self-regulated learning. *Metacognition and Learning*, 1-19.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

APPENDIX



Appendix 1: The SMART, COPES and AEIOU models. Adopted from Winne (2022)



Appendix 2: The technology-enhanced learning (TEL) environment. Adopted from Fan et al. (2022).

Towards Educating Competent Data Science Problem Solvers in the 21st Century: an Exploration in Cognition and Metacognition

Maryam Alomair
[1] University of Maryland Baltimore County
[2] King Faisal University
[1] maryama4@umbc.edu
[2] mmalomair@kfu.edu.sa

ABSTRACT: Data Science is an interdisciplinary field that requires solving problems. Even though there are many Data Science education courses, programs, and certificates, the design of these courses does not concentrate on problem-solving. In addition, mastering data science problem solving (DSPS) requires careful design of educational interventions to provide learners with ample practice opportunities on those skills. The rich problem-solving literature underscores the importance of cognition and metacognition; however, they have yet to be explored in the DSPS context. Using various learning analytics and machine learning techniques, this dissertation research attempts to address this gap by (1) building models of cognitive and metacognitive processes while learners are engaged in DSPS activities and (2) exploring the role of cognition and metacognition in predicting learning outcomes. This work may inform the design of DSPS learning environments that can encourage the education of competent data science problem solvers.

Keywords: Data Science Problem Solving, Data Science Education, Cognition, Metacognition, Log Analysis, Learning Log, Behavior Modeling, Learning Gain

1 BACKGROUND

Data Science (DS) is an interdisciplinary field with a strong emphasis on problem-solving. In recent years, many new Data Science programs, curricula, certifications, and training programs have been created in response to the rising demand for data science and analytics professionals. While many of these programs focus on teaching knowledge in methods, algorithms, and skills in programming with data, they often lack dedicated support to promote learners' skills in Data Science Problem Solving (DSPS). For data scientists to become proficient in problem-solving, they need to develop adaptive expertise (Carbonell et al., 2014) by being exposed to a variety of problem contexts and engaged in deliberate practice of DSPS to encourage a kind of knowledge organization that can facilitate problem-solving, which is often a quite different from the one they use while they acquire the method-focused component skills. However, those DSPS focused learning opportunities are limited in current data science education.

Despite its importance, DSPS is challenging to master and difficult to teach. DSPS, just like any problem-solving, is a higher-order cognitive process that involves an iterative process of problem understanding, problem formulation, data understanding, experiment design, model diagnosis, interpretation, and communication. Those cognitive steps need to be supported by an equally complex metacognitive process that requires learners to monitor their problem-solving processes closely, identify gaps toward goals, and enact self-regulation to identify and implement relevant problem-solving strategies and tactics. Meanwhile, data science is a fast-developing field with many new emerging methods, techniques, and approaches; data scientist in the 21st century needs to solve related and challenging learning problems of their own to monitor their knowledge state continuously and identify knowledge gaps and resort to learning strategies and tactics to fill the gaps.

In parallel to the young field of data science, there is a limited amount of research in Data Science education, specifically focusing on data science problem solving and related cognition and metacognition. This sparse literature could be attributed to the lack of data collected from DSPS processes and learning activities designed to train data science problem solvers. In this research, we explore data collected from data science students in a higher education context while they work with a DSPS learning intervention called caselet (Chen & Dubrawski, 2018) - a case-based learning intervention specifically designed for training learners in DSPS. We collect various kinds of data from students that can be used to characterize the cognitive and metacognitive processes during DSPS and while they are learning various component skills in a Data Science course. In the following section, I will describe the road map of my dissertation grounded in this specific learning context, learning intervention, and data collection. I will describe the preliminary analysis done with a dataset and future work in addition to data collection and analysis.

2 OVERARCHING GOALS AND RESEARCH QUESTIONS

The overarching goal of this study is to (1) explore learning analytics and machine learning techniques to gain a deep understanding of DS learners' cognitive and metacognitive processes while they are engaged in case-based DSPS practices, and other related learning activities; (2) understand the role of cognition and metacognition on learning outcomes. The study has three broad lines of inquiry, each with specific research questions. Table 1 summarizes the research questions, related data, and analytical methods to answer each research question.

2.1 Cognition

A. What is the role of prior knowledge in learning gain?
B. What is the relational structure of knowledge components based on learners' performance?
C. What is the knowledge growth of DS knowledge components based on learners' performance?

2.2 Metacognition

A. How to measure metacognitive processes objectively and non-intrusively at a fine-grained level?
B. How do metacognitive behaviors relate to each other?

2.3 Interconnection Between Cognition and Metacognition

A. What is the role of metacognition in learning gain?
B. What is the relationship between metacognition processes and learning processes?

3 THE CONTRIBUTION OF THE STUDY

With the rapid increase of data, current Data Science education will not be able to meet the surging need for highly skilled data scientists capable of solving real-world problems. Therefore, there is a high-priority need to equip learners with a well-designed rationale tool that facilitates solving DS real-world problems with various contexts, scalable questions, feedback, and more. Those features will allow learners to explore real-world problems and solutions. Data Science, as stated in Section I, requires solving problems. DSPS is a new concept in Data Science education, with no previous studies examining the role of cognitive and metacognitive processes and their effectiveness in DSPS with a lens on Data Science knowledge components (Koedinger et al., 2012). Our ambition is that these learners, who practice DSPS by struggling, self-regulating their learning, reflecting on their understanding, and using their cumulative knowledge by reattempting different learning opportunities, become capable of moving from the educational environment to the workplace to solve real-world problems successfully. A DSPS intervention should be built upon the suggestions from the proposed study.

Table 1: Overview of Research Questions, Data Sets, and Methodology

	Research Questions	First Dataset	Second Dataset	Methodology
Cognition	● 2.1, A ● 2.1, B ● 2.1, C	● LAK23- Poster - under review ● AAAI23 - Student Abstract - accepted ● In progress		● Section 4.2.1 ● Section 4.2.2 ● Section 4.2.2
Metacognition	● 2.2, A ● 2.2, B	● Not applicable ● Not applicable	● To-do ● To-do	● Section 4.2.3 ● Section 4.2.3
Interconnection between cognition and metacognition	● 2.3, A ● 2.3, B	● Partially solved, LAK23 Poster - under review ● Not applicable	● To-do ● To-do	● Section 4.2.2 ● Section 4.2.3 ● Section 4.2.3

4 RESEARCH METHODOLOGY

4.1 Data collection

This research will use the data from data science students in a real-world teaching context in a higher education setting. We plan to explore two datasets.

The **first dataset** is collected in graduate-level Data Science courses. The first wave was collected in the Fall of 2021, the second wave of data collection was in the Fall of 2022, and the third wave will be in the Spring of 2023. The dataset includes:

- Self-reported Metacognition Awareness Inventory (MAI): consists of Knowledge about cognition and regulation of cognition (Schraw & Dennison, 1994)
- Cognitive assessment: pre/post
- Caselet practices: Students are task(ed) with seven caselets per semester. For each caselet, we collect the following data from Google form
  - learners' responses and reflections on their confidence level
  - perceived difficulty level, and interest level in the caselets
  - help-seeking strategy: the used search keywords and the visited websites
  - self-reported time-on-task for answering questions, reading feedback and the reflection

- Learning logs: learners weekly learning activities and processes throughout the semester. Learners were encouraged to reflect on their understanding of the materials and state their sentiments.

The **second dataset** will be collected through a web application designed to serve this study. The application hosts the caselets. The application-based data is an improvement from the previous Google form-based data collection as it helps to capture the fine-grained log data that has the potential to be used to model the metacognitive behaviors non-intrusively and objectively. We also plan to collect think-aloud data while learners solve the problem to probe into their cognitive and metacognitive processes.

4.2 Data Analytics Techniques

4.2.1 Exploratory Data Analysis

We performed various data analytic techniques to discover relations among metacognition, pre-assessment, and DSPS practices. For example, we used descriptive analysis to explore the distribution of the variables. We also utilized cosine similarity to gain insight into the similarity between DS knowledge components in DSPS.

4.2.2 Machine Learning Techniques

We plan to use machine learning techniques to predict learning gain based on predictors such as metacognition: i.e., MAI and caselets' metacognitive processes, and cognition: i.e., prior knowledge and caselets performance. In addition, we want to model the knowledge growth of learners while practicing DSPS by applying one of the learning trajectory methods. We also plan to apply Linguistic Inquiry and Word Count (LIWC), a Natural Language Processing (NLP) technique, to capture social, cognitive, and effective processes used in the learning log and the reflection on understanding caselets' questions and feedback to measure metacognitive processes.

4.2.3 Log Analysis and Behavior Modeling

Log file plays a significant role in examining problem-solving learning (Zoanetti & Griffin, 2017). A study (VanLehn, 1988) articulated the role of modeling learners' behavior to understand their learning processes. Another study has applied supervised learning (Beck & Woolf, 2000) to predict how likely learners are to answer a question correctly, relying on learners' behavior. The log file in our study captures learning activities that act as non-intrusive metacognition processes, such as help-seeking behaviors, moving back and forth between questions, and time spent on answering and reading the question's feedback besides the cognitive processes. After deploying the designed web application tool, we want to go further in our study by performing a log analysis to understand metacognition in the learning process and caselets context in depth and model learning behavior to predict learning gain and characterize learners' cognitive and metacognitive processes based on their performance.

5 CURRENT STATUS OF THE STUDY

As summarized in Table 1, we have finished the first analysis wave on Dataset 1. The main objective of this analysis is to understand the relationship between metacognition, cognition, and learning gain. We noted variations in correlations between and within prior knowledge, caselets, and self-reported metacognition. The direct connection between some caselets and MAI components is a sign of the plausible role of metacognition in explaining DSPS. This analysis shows that prior knowledge and MAI are the most predictive variables. Random Forest model can predict learning gain using MAI, and prior knowledge with about 9% mean absolute percentage error (MAPE). This study was accepted as a student abstract paper to AAAI 2023 titled "Modeling Metacognitive and Cognitive Processes in Data Science Problem Solving." In addition, a poster that reports the result of this study was submitted to LAK 2023 titled "The Role of Cognition and Metacognition in Data Science Problem Solving: Insights from a Field Study." In addition, we are working on an analysis to characterize the learning growth at the DS knowledge component level. Besides that, we aim to understand how detailed knowledge component level data may reflect their underlying cognitive and metacognitive process using the first dataset already collected. For example, we aim to study how self-confidence, help-seeking behavior, time spent on a task, and self-reflection can affect learning gain at a granular level. Moreover, since the self-reporting metacognition (MAI) collected in the initial data is subjective and cannot capture the dynamics of metacognition, we designed a DSPS web application tool that collects dynamic temporal data in a non-intrusive way, which we plan to utilize in the upcoming data collection.

REFERENCES

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive science, 36(5), 757-798.

Chen, L., & Dubrawski, A. (2018). Accelerated apprenticeship: teaching data science problem solving skills at scale. In Proceedings of the Fifth Annual ACM Conference on Learning at Scale, 1-4.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. Contemporary educational psychology, 19(4), 460-475.

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. Metacognition and learning, 3(2), 123-146.

Zoanetti, N. & P. Griffin (2017). Log-file data as indicators for problem-solving processes, in Csapó, B. and J. Funke (eds.), The Nature of Problem Solving: Using Research to Inspire 21st Century Learning

VanLehn, K. (1988). Student modeling. Foundations of intelligent tutoring systems, 55, 78.

Beck, J. E., & Woolf, B. P. (2000). High-level student modeling with machine learning. In International Conference on Intelligent Tutoring Systems 584-593.

Carbonell, K. B., Stalmeijer, R. E., Könings, K. D., Segers, M., & van Merriënboer, J. J. (2014). How experts deal with novel situations: A review of adaptive expertise. Educational Research Review, 12, 14-29.

Supporting Student Collaboration using Learning Analytics Dashboard

Arslan Azad
University of South Australia
azaay005@mymail.unisa.edu.au

ABSTRACT: The ability to collaborate and work in high-performing teams is a critical process and skill. There is a growing demand to ensure students have opportunities to learn and develop skills in undertaking collaboration. With the advancement of learning technology, the research shifted fom observations of student learning and group work towards using Learning Analytics to support collaborative learning. Most recently, Multimodal Learning Analytics (MMLA) – use of multiple types of data – has shown potential to expand understanding. Using MMLA, a broader array of data sets can bring novel insights about student collaboration. This doctoral research aims to leverage MMLA to understand group dynamics and communication patterns among student groups working in collaboration by designing, implementing, and evaluating a real-time dashboard for teachers. The dashboard will be created from the verbal and non-verbal indicators of speech obtained from the audio of student conversations. The study also will seek to understand how MMLA dashboards can support teachers in assessing and supporting collaboration. Interviews with teachers and focus groups with students regarding the use of dashboard will complement the analyses. It is anticipated that the user-centred approach will provide a deeper, and more accurate real-time analytics tool for use in classroom environments.

Keywords: Collaboration, Classroom analytics, MMLA, Dashboards

1 INTRODUCTION

Communication and collaboration are important traits of an effective team/group member. The increasing complexity of work environments requires employees to collaborate and solve complex problems Stover et al. (2018). It is essential that students entering the workforce can also work in teams effectively. Partnership for 21st-Century skills report Greenhill (2010) suggests that a student requires four fundamental skills to succeed in college - communication, collaboration, creativity, and innovation. In a collaborative learning group, the members must actively communicate, engage, and interact with each other, working towards a common goal (Akkerman et al., 2007).

The opportunity to work in group settings opens avenues for collaboration and, therefore, the potential to develop and build teamwork and collaborative skills. However, group work also presents challenges that can adversely affect student learning. Interviewing 23 students and 19 teachers, Le et al. (2018) identified four common challenges in collaboration, including lack of collaborative skills in students, competence, friendship, and free-riding – also defined as social loafing (Shimazoe et al., 2010). Free riding, where one or more group members do most of the work and the rest depend on them, has often been discussed in the literature and seen mostly in unskilled group members. Competence of members of collaborative groups is shown to hinder collective learning, where more competent members often ignore less competent members, and sometimes do not actively participate themselves as a result (Bunderson et al., 2011).

With the commercialisation of education, the growing classroom sizes and a limited number of teaching staff pose a great challenge to student learning and collaboration. Especially in practical learning environments, students working in groups need constant support from the instructor. In such settings, some students or groups can easily be overlooked in providing support in time. Groups can have members relying entirely on other members (Bunderson et al., 2011). Misalignment of group members with the group goal is also a potential challenge that students face during group work. For optimal group performance, the participation of members, alignment with the goal, and reflection on the artefacts produced by the group are essential (van Boxtel et al., 2000). In classrooms with several groups, it can be challenging for educators to identify and give timely support to students and groups to ensure optimal collaboration.

Learning analytics can support the teacher in providing more scaffolds about student activities by using visualisations and dashboards. The advent of digital technology and access to trace data has shown significant promise in supporting teaching and learning in various settings. The use of learning analytics has enabled teachers to monitor student progress and support students to self-identify areas of misconception or learning needs. While there is little debate regarding the importance of collaboration and teamwork skills, there remain challenges of how to assess developmental skills and provide students with ongoing actionable feedback. To date, there has been limited work investigating the use of real-time classroom support to better enable the development of individual collaboration skills. This doctoral research aims to address this deficit by incorporating learning analytics in the R-12 context.

2 BACKGROUND

2.1 Collaborative learning

In this doctoral research, collaborative learning is adopted and defined as two or more people working towards a shared learning goal (Jeong et al., 2016). Collaborative learning involves groups of learners working together towards solving a problem, completing a task or creating a product (Laal et al., 2012). Previous research has shown that merely placing members in a group and assigning them a task does not necessarily ensure that effective collaboration will take place (de Jong, 2019). De Jong further states that collaboration requires learning activities that motivate and guide students. Collaborative learning is a complex process that does not always produce the desired results (Kirschner et al., 2009). For instance, not all members of a working group contribute to completing the task and outcomes. Individual student contributions aligned with the common goal result in effective student collaboration. Individuals make parts of the group, and therefore, to improve the group outcomes, individuals require support in learning and development. To have a well-performing collaborative team, it is important to make a deliberate effort and take measures to find ways to increase collaboration among group members in a learning environment.

2.2 Learning Analytics for Collaborative Learning

Learning analytics dashboards and feedback tools have gained attention over the past few years as catalysts to Computer Supported Collaborative Learning (CSCL) classroom orchestration (Prieto et al., 2019). Providing effective teacher support to students requires the teacher to be aware of activities in the classroom, monitor student groups, and identify the groups that need attention (Kaendler et al., 2015). Therefore, identifying the needs of teachers, the research community is directing its focus on real-time augmentation of classrooms to support the teachers and students (An et al., 2020). Investigating teachers' use of a dashboard to support classroom collaboration, Amarasinghe et al. (2020) designed, implemented, and evaluated a dashboard across 16 class sessions, positing that teachers found the dashboard helpful in facilitating student groups. The results indicated that as the teacher saw a decline in student participation in a group, they read the dashboard more often, leading to greater support given to students. This research also identified that, in a relatively smaller group work setting, contrary to the concern with the dashboard increasing orchestration load for the teacher, the use of the dashboard helped teachers provide more scaffolds.

2.3 Multimodal LA

Much work has been carried out with log data to capture traces of learner behaviours on learning systems (Siemens, 2013). The resulting analyses from studies based on log data are limited and rely on course designs and learning activities. The use of multimodal data, especially verbal and non-verbal audio of student conversations, can add more layers of understanding, providing a richer understanding of student learning. However, the use of multimodal data in learning environments is still in its early stages. To allow for a better understanding of learning than possible with just log data, some of the recent research (Ciordas-Hertel et al., 2021); Yan et al., 2022) uses cameras to capture student gestures, sensors to measure student stress, as well as audio data to capture student speech patterns. A useful approach to capture multimodal traces of behaviours from learner activities is via the combination of sensors with log data, analyse them and build feedback mechanisms (Niemi et al., 2018).

In another study, Martinez-Maldonado (2019) implemented a real-time analytics tool to investigate teachers' perspectives on its use in supporting CSCL classroom orchestration. The study involved postgraduate students working on small group tasks or their capstone projects, with multimodal data captured from Microsoft Kinect sensors and interactive tabletops that enabled brainstorming. The dashboard included visualisations about group participation and the progress of tasks obtained from the size of students' solutions. It then sent notifications to teachers about the correctness of the content of artefacts produced by students. Maldonado mentioned the challenges identified, including accurate representation of classroom activity due to the incomplete nature of classroom data, the trade-off between early intervention and delayed intervention (which can allow the development of ideas), possible disruption caused by analytics resulting in undesired actions due to inaccurate judgement from the visualisation, and a possible increase in orchestration load for the teacher. While research is being conducted in multimodal data fusion and the use of multiple sensors with the challenges it offers, there is an opportunity to use data generated from sources other than log data to study and support collaboration.

3 RESEARCH GOALS AND QUESTIONS

This doctoral research will support teacher scaffolding to ensure better support in enhancing student collaboration during group work. To measure levels of collaboration, group composition, group dynamics and communication patterns exhibited by students can be observed. In the context of this study, group dynamics are defined as an outcome of differing actions and behaviours of various group members, their activities, and communication patterns when working towards a common goal. Some examples of the observable group dynamics can include contribution of group members, conflicts, turn-taking, and involvement. The project involves the use of multimodal audio data and observations generated by learners in a classroom setting, working in a group, to understand group dynamics as well as to discover patterns in the way group members collaborate. Specifically, this thesis focuses on answering the following four research questions:

- **RQ1:** What group dynamics can we observe emerging in multimodal classroom learning environments?
- **RQ2:** What kind of information and visuals do teachers find actionable and useful in a dashboard to facilitate them in supporting student collaboration?

- **RQ3:** How do teachers perceive dashboards' usefulness? How can we use these insights to improve the design of dashboards to better support student collaboration?
- **RQ4:** What are students' perceptions regarding the use of dashboards by the teachers?

The term multimodal audio data refers to verbal audio data (i.e., speech content) and non-verbal audio data (e.g., pauses and length of speech). With substantial existing research on log data, we aim to present student conversations to the teacher in the form of visualisations. Previous similar research Praharaj, Scheffel, Drachsler, et al. (2021); Scherr et al. (2009) has been primarily on post-analysis of conversations, and there is currently no automated, real-time dashboard for use in the classrooms. This thesis project will present a near real-time, automated dashboard and study its use for teachers and students in the classroom to support collaboration.

4 METHODOLOGY

The project will use classroom analytics to gain insights and understanding of students' collaborative learning processes. This will be achieved by developing a real-time teacher dashboard, presenting visualisations from discussions among group members in a collaborative classroom environment. The study will involve setting up an experiment in a classroom of R-12 students working on a practical lab task delivered by UniSA Connect, which is UniSA's school outreach initiative. Student conversations in small groups will be recorded as audio files using microphone arrays. Data science techniques explained further in the proposal will be used to create visualisations in the form of a dashboard. These visualisations will give instructors insights into student collaboration and support their decision-making and instruction.

The study will involve setting up an audio/speech analytics dashboard in a classroom of R-12 students working in groups of 4 or 5 on a STEM task involving practical lab projects. The methodology includes understanding the content of student conversations using Topic Analysis techniques Blei et al. (2003); Papadimitriou et al. (1998), statistical analysis techniques, and visualisations like social network analysis graphs, among others. Consequently, the generated insights from the speech will be visualised in a dashboard to be used by the teacher to understand the group dynamics exhibited by students. We will determine the group collaboration stage by using the ICAP Framework (Chi et al., 2014), which terms the stages of collaboration as Interactive, Constructive, Active, and Passive. Using this framework, our experiment will involve identifying which stage of collaboration each group falls into and observing their communication patterns to find any correlations with collaboration or with their success in the project. This will help the educators to focus their attention on the group dynamics in action and identify those individuals and groups that are likely to need support. By conducting post-experiment interviews with the teachers, the study will delve deep into the teacher's feedback about their expectations and experience with dashboard-supported classroom instruction. By conducting focus groups with the students, we will also obtain student feedback and perceptions about the impact of the use of dashboard on student learning.

RQ1 aims to understand group dynamics better and add to the literature on group dynamics in a lab environment for R-12 students. To achieve this, we will use Re-speaker Mic Array to record the audio, AWS transcription service in AWS cloud to transcribe raw audio recordings, and Topic Modelling techniques to analyse speech and identify key themes in student discussions. The results of these analyses will then be presented to teachers using several different visualisations. In conjunction with analysing the verbal content of conversations, we will also use nonverbal audio indicators like pitch, overlapping of members' speech, and time taken by each member to understand the group dynamics in a collaborative team environment, which have been studied in the past (Kim et al., 2015; Praharaj et al., 2021b).

RQ2 will utilise teachers' knowledge of teaching and pedagogical practices to guide the initial choice of visuals in the dashboard. For this purpose, we will conduct teacher interviews before the design of the dashboard. It is anticipated that the teacher's involvement will not only inform the content of the dashboard but it will also contribute to the technology adoption.

While RQ1 will contribute to the research about group dynamics in a lab/classroom environment and RQ2 will utilise existing teaching practices to inform the visuals, RQ3 will help understand and present more efficient use cases to help classroom learning. After the initial dashboard design, interviews will be conducted with the teacher to get their feedback. Teacher feedback will enable actionable insights, thus making the tool a useful addition to classroom instruction. Understanding the efficiency of the speech analytics dashboard will potentially enable the students to learn and collaborate better. The study for RQ3 will look at ways in which the speech analytics of collaboration in groups can facilitate the teacher with classroom instruction.

RQ4 expands the study further to arguably the major stakeholder in the project, the students. We will analyse the effect of teacher's use of the dashboard on students and investigate any ethical concerns among students. The response of students to teacher interventions, together with their responses in focus groups, will allow us to determine the efficacy of a real-time collaboration dashboard from both the teacher and the students' perspectives.

5 CURRENT PROGRESS

Before the data collection could be started, we experimented with various audio recording hardware setups and decided to use 4-mic HAT with Raspberry Pi. As part of the pilot stage of the project, we started preliminary data collection at Pulteney Grammar School in Adelaide in their STEM Robotics class. Following the user-centred approach, this pilot phase focuses on integrating the teacher's needs into the design process. At the preliminary stage, the data obtained after running the recording through the transcription service is of good quality, without significant loss, despite the noisy lab environment. We are also conducting preliminary analyses on the data obtained from the classroom to improve the design of the dashboard and data-collection methods.

6 CONTRIBUTION

The primary motivation behind this doctoral project is to facilitate students' collaboration in a classroom and to help teachers in classroom instruction by utilising a combination of data science methodologies and research about collaboration. Although research has been conducted previously on understanding collaboration, there is an opportunity to understand the group dynamics that emerge in a collaborative classroom environment using audio/speech data. Our research specifically focuses on setting up the speech analytics experiment in lab projects, with R-12 students collaborating in groups on a usual lab task. Most similar research involved a game environment, a meeting, or in cases where experimentation was done in a classroom environment, was merely done for collecting data for post-class processing. The completion of this project will equip classrooms with a near real-time analytics dashboard visualising student activities in a collaborative learning environment.

REFERENCES

Akkerman, S., Van den Bossche, P., Admiraal, W., Gijselaers, W., Segers, M., Simons, R.-J., & Kirschner, P. (2007). Reconsidering group cognition: From conceptual confusion to a boundary area between cognitive and socio-cultural perspectives? *Educational Research Review*, *2*(1), 39–63.

Amarasinghe, I., Hernández-Leo, D., Michos, K., & Vujovic, M. (2020). An Actionable Orchestration Dashboard to Enhance Collaboration in the Classroom. *IEEE Transactions on Learning Technologies*, *13*(4), 662–675. https://doi.org/10.1109/TLT.2020.3028597

An, P., Holstein, K., d'Anjou, B., Eggen, B., & Bakker, S. (2020). The TA Framework: Designing Real-time Teaching Augmentation for K-12 Classrooms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bunderson, J. S., & Reagans, R. E. (2011). Power, Status, and Learning in Organizations. *Organization Science*, *22*(5), 1182–1194. https://doi.org/10.1287/orsc.1100.0590

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243.

Ciordas-Hertel, G.-P., Rödling, S., Schneider, J., Di Mitri, D., Weidlich, J., & Drachsler, H. (2021). Mobile Sensing with Smart Wearables of the Physical Context of Distance Learning Students to Consider Its Effects on Learning. *Sensors*, *21*(19), Article 19.

de Jong, T. (2019). Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning*, *35*(2), 153–167.

Greenhill, V. (2010). 21st Century Knowledge and Skills in Educator Preparation. In *Partnership for 21st Century Skills*. Partnership for 21st Century Skills. https://eric.ed.gov/?id=ED519336

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher Competencies for the Implementation of Collaborative Learning in the Classroom: A Framework and Research Review. *Educational Psychology Review*, *27*(3), 505–536.

Kim, J., Truong, K. P., Charisi, V., Zaga, C., Lohse, M., Heylen, D., & Evers, V. (2015). Vocal turn-taking patterns in groups of children performing collaborative tasks: An exploratory study. *INTERSPEECH 2015: 16th Annual Conference of the International Speech Communication Association*, 1645–1649.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A Cognitive Load Approach to Collaborative Learning: United Brains for Complex Tasks. *Educational Psychology Review*, *21*(1), 31–42.

Laal, M., & Laal, M. (2012). Collaborative learning: What is it? *Procedia - Social and Behavioral Sciences*, *31*, 491–495. https://doi.org/10.1016/j.sbspro.2011.12.092

Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: Teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education*, *48*(1), 103–122.

Martinez-Maldonado, R. (2019). A handheld classroom dashboard: Teachers' perspectives on the use of real-time collaborative learning analytics. *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 383–411. https://doi.org/10.1007/s11412-019-09308-z

Niemi, D., Pea, R. D., & Saxberg, B. (2018). *Learning Analytics in Education*. Information Age Publishing, Incorporated. http://ebookcentral.proquest.com/lib/unisa/detail.action?docID=5516043

Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. (1998). Latent semantic indexing: A probabilistic analysis. *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 159–168.

Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2021). Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics—Can We Go the Whole Nine Yards? *IEEE Transactions on Learning Technologies*, *14*(3), 367–385.

Prieto, L. P., Rodríguez-Triana, M. J., Martínez-Maldonado, R., Dimitriadis, Y., & Gašević, D. (2019). Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology*, *35*(4), Article 4. https://doi.org/10.14742/ajet.4314

Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction*, *27*(2), 147–174.

Shimazoe, J., & Aldrich, H. (2010). Group Work Can Be Gratifying: Understanding & Overcoming Resistance to Cooperative Learning. *College Teaching*, *58*(2), 52–57.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Stover, S., & Holland, C. (2018). Student Resistance to Collaborative Learning. *International Journal for the Scholarship of Teaching and Learning*, *12*(2). https://doi.org/10.20429/ijsotl.2018.120208

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330.

Yan, L., Martinez-Maldonado, R., Zhao, L., Dix, S., Jaggard, H., Wotherspoon, R., Li, X., & Gašević, D. (2022). The role of indoor positioning analytics in assessment of simulation-based learning. *British Journal of Educational Technology*, *n/a*(n/a). https://doi.org/10.1111/bjet.13262

Supporting Teamwork Reflection in Healthcare Simulation through Analytics of Communication

Author: Linxuan Zhao
Monash University
linxuan.zhao@monash.edu

ABSTRACT: Teamwork skills in healthcare are significant as inadequate teamwork skills threaten patients' safety. To learn teamwork skills, healthcare students usually need to participate in embodied simulations. Learning through reflection is a critical part of such simulation-based learning. Reflection evidence, such as video footage, can be used to improve the effectiveness of reflection. However, it is hard to collect such evidence in practice due to the difficulty of manual data processing. This PhD project aims to explore how to capture students' communication, a critical part of teamwork, through audio trace data to provide evidence to support students' reflection on teamwork. This project would analyse the feasibility of mapping audio trace data through communication to teamwork construct. Then, the finding of the analysis would be used to develop reflection evidence to support students' reflection and learning on teamwork.

Keywords: Teamwork, Simulation-Based Learning, Learning Analytics, Audio Trace Data

1 INTRODUCTION

Teamwork is one of the most in-demand skills across many professional sectors (Suarta et al., 2017). To meet this demand, higher education providers have increased attention to developing effective teamwork skills (Riebe et al., 2016). Healthcare is one of the fields where teamwork plays a significant role (Rosen et al., 2018). Healthcare teams who demonstrate poor teamwork skills have been associated with an increased potential threat to patients' safety in healthcare activities (Rosen et al., 2018). This situation has raised the attention of healthcare teamwork education.

One education strategy to improve healthcare students' teamwork skills is simulation-based learning. Students would participate in a series of high-fidelity healthcare simulations to practice their skills, such as teamwork and teamwork-related skills (e.g. communication) (Fanning & Gaba, 2007). In such simulations, students are commonly presented with a challenging situation and limited information to explore and interact with various resources and actors in the physical space while enacting a team role (Parker & Myrick, 2009). To solve the challenge posed to them, students need to communicate effectively with other team members or human actors enacting various roles (e.g., medical staff, patients, or family members) (Foster et al., 2019), collect information from various physical resources and digital devices (e.g., computers) and act according to professional guidelines. A debrief session would be held after the simulation to guide the students' reflection to learn from what they did in the simulation to improve their skills (Fanning & Gaba, 2007). Such reflection activities had been found can be supported by providing evidence for reflection (e.g. video footage). However, it is not easy to provide such evidence in practice due to the difficulty of manually processing data (Wise et al., 2021).

There is an established body of research in Learning Analytics (LA) that focuses on automatically analysing audio trace data captured from sensors to provide information about higher-order constructs (e.g. teamwork) (Mu et al., 2020; Praharaj et al., 2021). The application of sensors for capturing data and automated data processing techniques provides an opportunity to generate evidence of reflection automatically. This PhD project aims to explore the potential to capture students' communication through audio trace data to map to higher-order teamwork constructs and provide evidence to support students' reflection in a healthcare simulation setting.

2 RELATED WORK AND EXISTING SOLUTIONS

2.1 Small-group communication analysis

The learning analytics studies about communication using audio trace data can be divided into two types: 1) using the summary properties of communication (e.g., overlaps in communication) and 2) using the dialogue content of communication (e.g., what students said).

Regarding summary properties, Bachour et al. (2010) developed a digital device embedded in a table to detect the total speaking time of students and showed these metrics to help them regulate their learning or contribute to the conversation equally. Scherer et al. (2012) proposed a more sophisticated modelling approach by extracting acoustic features (physical characteristics of speech sounds, such as intensity) from speech signals to assess students' expertise and leadership in groups of students working collaboratively to solve math problems. In a similar setting, Oviatt et al. (2015) demonstrated that the number of communication interruptions in collaborative problem-solving activity could be a robust indicator of expertise level. Along the same lines, Bassiou et al. (2016) showed that overlaps in speech can be a key behaviour to detecting collaboration quality.

Regarding the dialogue content of communication, an emerging body of research was established to use automated speech recognition techniques to support learning activities. For example, Jensen et al. (2010) created a model to automatically detect teacher-student instances of communication for providing feedback to improve teachers' discourse skills in the classroom. Focusing also on teacher-student dialogue, Kelly et al. (2018) built a software system to detect the level of authority of teachers' questions posed to students. Pugh et al. (2021) designed a software system to provide analytics on students' collaborative problem- solving (CPS) skills. Praharaj et al. (2021) created data visualisations for a group collaboration task that portrayed the usage of words, such as frequency or connection between certain keywords, to demonstrate collaboration quality and provoke reflection. Similarly, Southwell et al. (2022) provided a dashboard for classrooms that automatically coded discourse segments among groups of students into constructs.

Overall, existing studies demonstrate that analysis of group communication is important for a wide range of educational scenarios and gaining traction.

2.2 Existing solutions to support team reflection

There are several studies applying sensor-captured trace data to generate evidence for reflection. Fernandez-Nieto, Echeverria, et al. (2021) designed a user interface to visualise the timeline of students' logged actions in a healthcare simulation. This visualisation was then provided to the students to support their reflection. Echeverria et al. (2019) developed a series of visualisation using physiological data, positioning data, and logged action to show the collaboration within healthcare simulation. They analysed each visualisation's visibility, awareness, and accountability by interviewing the students who participated in the simulation. Those studies focused on providing evidence to support reflection through spatial or log trace data. The way to apply audio trace data to provide evidence for reflection remains to be explored.

3 RESEARCH QUESTIONS

The main research question can be phrased as follows: How can digital traces of verbal and non-verbal communication be mapped to higher-order teamwork constructs to support reflection in embodied simulation-based learning? Throughout this project, we will address this main research question by addressing three sub-research questions: **(RQ1)** To what extent can summary properties of communication be mapped to higher-order teamwork constructs in embodied simulation-based learning? **(RQ2)** To what extent can dialogue content of communication be mapped to higher-order teamwork constructs in embodied simulation-based learning? **(RQ3)** To what extent can visual interfaces for demonstrating students' communication be designed to support reflection on teamwork constructs in embodied teamwork learning? RQ1 and RQ2 aim to analyse whether summary properties and dialogue content of communication can be mapped to teamwork construct and provide reflection evidence. RQ3 aims to design the visual interface (such as visualisation) of audio trace data according to findings from addressing RQ1 and RQ2 to support students' reflection.

4 METHODOLOGY

4.1 Data collection

The dataset of healthcare simulation will be collected in collaboration with the Faculty of Nursing in Monash University. This data collection will be held around July and September from 2021-2023. For each year, there will be around 70 healthcare simulation sessions and four students in each session, so we have the chance to collect data from around 280 students each year.

In this healthcare simulation setting, four students would do healthcare activities with four manikin patients separated on different beds. Those students needed to work collaboratively to complete healthcare tasks, respond to an emergent event, call for external help, and communicate with a doctor efficiently. The simulation teaching team would control the manikin patients and observe students' performance behind a glass mirror to prevent intervening in their activities. We collected the digital traces of audio, video, and spatial, where video would only be used for observation, while audio and spatial data would be processed and used for providing reflection. Besides the sensor-captured data, we also collected the evaluation results of students' performance regarding their teamwork, communication, and understanding of team role.

4.2 Resolving complexity in communication dynamics

In this simulation setting, students need to separate to different locations to complete different tasks in parallel, resulting in complexity in the communication that multiple dialogues can happen in parallel. In such healthcare settings with complex communication dynamics, embodied team dialogue usually happens in close proximity (such as less than 1.5 meters in a hospital ward (Sorokowska et al., 2017) to have effective dialogue without interference from other dialogues happening in parallel. Thus, we combined the audio and spatial data to detect whether two students were in the same dialogue. This step enables the analysis of communication to resolve RQ1 and RQ2.

4.3 Analysing the feasibility of applying summary property to support embodied teamwork reflection

To address RQ1, we will analyse the feasibility of using the summary property of communication, such as turn-talking behaviour (Kim et al., 2015), and speech overlaps (Bassiou et al., 2016), in the simulation-based learning settings. Pearson's correlation coefficient will be used to determine whether the summary properties were related to the three teamwork performance criteria (teamwork, communication ability, and role understanding). We assume the summary properties would be feasible to be used if they demonstrated statistically significant correlations to teamwork performance criteria. Finally, a semi-structured interview will be conducted with the healthcare educators (they provided the rating of students) to acquire their idea to determine whether those summary properties of communication can truly be used to support student reflection.

4.4 Analysing the feasibility of applying dialogue content of communication to support embodied teamwork reflection

To address RQ2, the same dataset in RQ1 will be used. The students' dialogue will be transcribed to obtain the dialogue content of communication. Then, an annotation scheme will be developed to extract communication behaviours in the transcriptions. This coding scheme will be developed by adapting teamwork and healthcare communication related studies. As suggested by theories of communication in healthcare teams (Haig et al., 2006), the co-occurrence of communication behaviours can serve to identify highly-effective teamwork. Thus, those communication behaviours will be analysed through methods that account for co-occurrences, such as Epistemic Network Analysis (ENA) (Shaffer et al., 2016).

4.5 Designing visual interfaces to support reflection of students

To resolve the RQ3, we will apply the digital traces of verbal and non-verbal communication with the mapping methods found in previous research questions to develop visualisations that healthcare students can make sense and reflect through it. The potential methods that can be applied to develop visualisation are epistemic network (Fernandez-Nieto, Martinez-Maldonado, et al., 2021), social network (Echeverria et al., 2019), and timeline (Echeverria et al., 2019), as previous studies demonstrated that those methods are effective for healthcare educators and students to make sense and help their reflection. We also plan to hold co-design sessions with the nursing educators to create visualisations that can make sense to them and support their students. At last, a platform will be built to process spatial and audio data and then generate the visualisations to support students' reflection. This platform will be tested in the last data collection in 2023.

5 ETHICAL CONSIDERATION

Ethical approval was obtained from Human Research Ethics Committee of Monash University. We only collected the data from students who consented through filling a formal written consent form. No personal information will be collected in the data collection. The collected data was saved on cloud and the access was strictly controlled to individuals within the ethical approval list to ensure the data will only be used for research purpose.

6 CONTRIBUTION OF SUGGESTED SOLUTION

The solution suggested in this project has two major differences from the existing solutions. First, this project explored a learning setting with complex communication dynamics where multiple dialogues can happen in parallel. This is a learning setting lacking exploration in learning analytics study, as previous studies in embodied learning setting focused on single dialogue, such as dialogue in a collaborative problem-solving group (Bachour et al., 2010). The second difference is applying audio trace data to generate the reflection evidence, where previous studies focus on spatial (Fernandez-Nieto, Martinez-Maldonado, et al., 2021) or log trace data (Echeverria et al., 2019).

7 ACHIEVED SO FAR

We published a paper at the international conference of Learning Analytics and Knowledge 2022 (LAK22) demonstrating the method to resolve the complexity in communication dynamics and analysis of the feasibility of applying summary property to support embodied teamwork reflection (Zhao et al., 2022). Another paper was submitted to LAK23 for the analysis of using dialogue content of communication to support embodied teamwork reflection, which is currently under revision.

REFERENCES

Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies*, *3*(3), 203-213. https://doi.org/10.1109/TLT.2010.18

Bassiou, N., Tsiartas, A., Smith, J., Bratt, H., Richey, C., Shriberg, E., D'Angelo, C., Alozie, N., & Int Speech Commun, A. (2016). Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration. *17th Annual Conference of the International Speech Communication Association*, 888-892. https://doi.org/10.21437/Interspeech.2016-1569

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,

Fanning, R. M., & Gaba, D. M. (2007). The Role of Debriefing in Simulation-Based Learning. *Simulation in healthcare*, *2*(2), 115-125. https://doi.org/10.1097/SIH.0b013e3180315539

Fernandez-Nieto, G. M., Echeverria, V., Shum, S. B., Mangaroska, K., Kitto, K., Palominos, E., Axisa, C., & 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

Fernandez-Nieto, G. M., Martinez-Maldonado, R., Kitto, K., & Buckingham Shum, S. (2021). Modelling spatial behaviours in clinical team simulations using epistemic network analysis: methodology and teacher evaluation. LAK21: 11th International Learning Analytics and Knowledge Conference,

Foster, M., Gilbert, M., Hanson, D., Whitcomb, K., & Graham, C. (2019). Use of simulation to develop teamwork skills in prelicensure nursing students: an integrative review. *Nurse educator*, *44*(5), E7-E11. https://doi.org/10.1097/NNE.0000000000000616

Haig, K. M., Sutton, S., & Whittington, J. (2006). SBAR: a shared mental model for improving communication between clinicians. *The joint commission journal on quality and patient safety*, *32*(3), 167-175.

Jensen, M. L., Meservy, T. O., Burgoon, J. K., & Nunamaker, J. F. (2010). Automatic, Multimodal Evaluation of Human Interaction. *Group Decision and Negotiation*, *19*(4), 367-389. https://doi.org/10.1007/s10726-009-9171-0

Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M., & D'Mello, S. K. (2018). Automatically measuring question authenticity in real-world classrooms. *Educational Researcher*, *47*(7), 451-464.

Kim, J., Truong, K. P., Charisi, V., Zaga, C., Lohse, M., Heylen, D., & Evers, V. (2015). Vocal turn-taking patterns in groups of children performing collaborative tasks: an exploratory study. Sixteenth Annual Conference of the International Speech Communication Association,

Mu, S., Cui, M., & Huang, X. (2020). Multimodal Data Fusion in Learning Analytics: A Systematic Review. *Sensors*, *20*(23), 6856.

Oviatt, S., Hang, K., Zhou, J., Chen, F., & Acm. (2015). Spoken Interruptions Signal Productive Problem Solving and Domain Expertise in Mathematics. *Icmi'15: Proceedings of the 2015 Acm International Conference on Multimodal Interaction*, 311-318. https://doi.org/10.1145/2818346.2820743

Parker, B. C., & Myrick, F. (2009). A critical examination of high-fidelity human patient simulation within the context of nursing pedagogy. *Nurse education today*, *29*(3), 322-329.

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). https://doi.org/10.3390/s21093156

Pugh, S. L., Subburaj, S. K., Rao, A. R., Stewart, A. E., Andrews-Todd, J., & D'Mello, S. K. (2021). Say What? Automatic Modeling of Collaborative Problem Solving Skills from Student Speech in the Wild. *International Educational Data Mining Society*.

Riebe, L., Girardi, A., & Whitsed, C. (2016). A Systematic Literature Review of Teamwork Pedagogy in Higher Education. *Small Group Research*, *47*(6), 619-664. https://doi.org/10.1177/1046496416665221

Rosen, M. A., DiazGranados, D., Dietz, A. S., Benishek, L. E., Thompson, D., Pronovost, P. J., & Weaver, S. J. (2018). Teamwork in healthcare: Key discoveries enabling safer, high-quality care. *American Psychologist*, *73*(4), 433.

Scherer, S., Weibel, N., Morency, L.-P., & Oviatt, S. (2012). Multimodal prediction of expertise and leadership in learning groups. Proceedings of the 1st International Workshop on Multimodal Learning Analytics,

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.

Sorokowska, A., Sorokowski, P., Hilpert, P., Cantarero, K., Frackowiak, T., Ahmadi, K., Alghraibeh, A. M., Aryeetey, R., Bertoni, A., Bettache, K., Blumen, S., Błażejewska, M., Bortolini, T., Butovskaya, M., Castro, F. N., Cetinkaya, H., Cunha, D., David, D., David, O. A., . . . Pierce, J. D. (2017). Preferred Interpersonal Distances: A Global Comparison. *Journal of Cross-Cultural Psychology*, *48*(4), 577-592. https://doi.org/10.1177/0022022117698039

Southwell, R., Pugh, S., Perkoff, E. M., Clevenger, C., Bush, J. B., Lieber, R., Ward, W., Foltz, P., & D'Mello, S. (2022). Challenges and Feasibility of Automatic Speech Recognition for Modeling Student Collaborative Discourse in Classrooms. *thinking*, *27*, 29.

Suarta, I. M., Suwintana, I. K., Sudhana, I. F. P., & Hariyanti, N. K. D. (2017, 2017/09). Employability Skills Required by the 21st Century Workplace: A Literature Review of Labor Market Demand.

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In *International handbook of computer-supported collaborative learning* (pp. 425-443). Springer.

Zhao, L., Yan, L., Gasevic, D., Dix, S., Jaggard, H., Wotherspoon, R., Alfredo, R., Li, X., & Martinez-Maldonado, R. (2022). Modelling Co-located Team Communication from Voice Detection and Positioning Data in Healthcare Simulation. LAK22: 12th International Learning Analytics and Knowledge Conference,

Representing Assimilation Patterns for Learning

Praseeda
International Institute of Information Technology Bangalore, India
Praseeda@iiitb.ac.in

ABSTRACT: Assimilation, as stated by Piaget (1954), is the process of fitting new information into existing cognitive schemas and is integral to how people learn. The way in which learners assimilate knowledge depends on their past experiences, prior knowledge, self-efficacy, epistemological beliefs, mindsets, beliefs about self-worth, interests, innate proclivities, curiosity, emotions, anxiety, and many other latent factors. This diversity of assimilation processes among learners is often lost when pedagogy focuses solely on measurable outputs from uniform assessments. Assimilation patterns for a learner represent their transformation as an individual, after going through a learning journey. Understanding patterns of assimilation are hard when done at scale, but are important as they provide feedback to the learner about their learning process. In this research, we propose an approach to automatically capture, represent and reason about assimilation of knowledge by students in an online learning platform. This approach can complement the single dimensional feedback like grades and marks that reflect on educational outcomes.

Keywords: Learning Outcomes, Assimilation Patterns, Online Learning Spaces, Learner's Polyline

1 INTRODUCTION

Learning is a process of acquiring conceptual knowledge in various domains, procedural knowledge of various skills and higher order reasoning skills (Mukunda, 2009). Learners participate in various activities as a part of the learning journey. Two different individuals have different memories of their experience, as they interpret everything based on the knowledge they already possess. Learners construct knowledge when they make interpretation of new experiences based on the knowledge they already possess. Assimilation is a process of linking the new piece of information to the pre-existing cognitive structures to understand and create new cognitive structures (Piaget, 1954; Ausubel, 1968).

Educational practices have focused on observing the learning experience through the lens of learning outcomes and outputs. Outcomes refers to the abilities that the learner achieves after completing a course, while outputs like exams, viva, essays give us a way to assess if we have achieved the outcomes. These outcomes are expected to give us a sense of how the learners have attained their educational goals after going through the learning journey in a course, by collecting observable outputs in terms of essays, quizzes, exams, seminars, creative outputs etc. The need to manage large classrooms have led to the problem of measuring learning outcomes at scale, further led to standardization of educational practices and assessment models. This standardization has led to uniform processes for managing classroom, conducting exams, measuring outcomes etc. (Brady, 1996; Tam, 2014).

Even though learning outcomes and outputs were used to observe the assimilation of the learners, the process of observing the outcomes has been standardized with the help of exams. This has led to all stake holders of the education system, including the teacher and the students focusing and investing energy in maximizing marks or grades in the exams. The focus on learning outcomes, rather than improvement of learning process has resulted in has many issues in the education system including rote learning, academic dishonesty, etc. (Morgan, 2016; Dendir & Maxwell, 2020)

In this research, we would like to focus on providing feedback to the learner on the process of knowledge construction, specifically assimilation. To reflect assimilation, we suggest to look at learning as a process of understanding and integrating different topics. Hence, topics becomes the central piece of how assimilation can be reflected. These topics are mapped by the course instructor based on the objectives of the course. Not all learners learn and assimilate all topics of the course with the same interest and ability. Some learners prefer to master some topics in depth and have little interest in others. Understanding this topical interest of the learner could give an insight to the learner about their learning patterns and interests and this feedback would help the learner in planning their next steps. This also provides an intuition of their strengths and weakness in the course. Learners in a constructivist environment provide evidence of how well they have integrated the knowledge in the form of essays and seminars. In this study, we use data that can serve as evidence of assimilation like essays and seminars to the design a two-dimensional representation mechanism that can capture and represent the assimilation patterns of the learners after experiencing a learning journey automatically with the help of *Discourse Learning Map*.

2 BACKGROUND

Learners receive knowledge in disconnected chunks and they organize these chunks in the brain in the form of organized structures called schemas (Piaget, 1954). Experiences are continuously fitted into these schemas through assimilation and schemas change to allow new information in accommodation. Ausubel's assimilation theory (1968) states that meaningful learning occurs as a result of the interaction between new information that the individual acquires and a particular cognitive structure that the learner already possesses that serves as an anchor for integrating the new content into prior knowledge. Since each individual has different prior experiences and prior knowledge, the way they assimilate and construct knowledge will be diverse (Woolfolk & Kapur, 2019).

Learning in constructivist learning environments, where learner constructs new knowledge based on their prior knowledge is understood through assessments tools like concept maps, portfolios, rubrics, etc. Rubrics are designed to provide feedback for an assignment and characterizes the learner's work into various dimensions mapped specific to the task given to them (Jonassen, 1991; Stevens & Levi, 2013). Each of these assessment tools requires effort from teachers and students to understand the knowledge acquisition. These tools provide feedback only with respect to an activity done by the student and would only help the student in improving that activity if its repeated. We need models that empower the learners by providing feedback on the topics understood and assimilated by the learners that would give a sense of their interests, and if they are achieving their learning goals in the course as a whole.

Recently, learning analytics researchers and practitioners have created Learning Analytics Dashboards, that help the learners to analyze, visualize, organize and track their learning. These tools provide descriptive feedback to the learning journey in terms of the learner's performance against class average, their engagement metrics, participation metrics like resources consumed weekly, with the help of graphs and visualization elements (Susnjak et al., 2022; Jivet, 2018). Though these dashboards are ways for the learner to observe their learning patterns, it does not give any insights into what topics of the course are favored over the other, how are other learners assimilating the course through their activities. A learner sees statistics shown in the dashboard as the behavioral features like usage patterns may or may not help in taking next steps and neither help in navigating through the course to reach their learning goals. We suggest an alternative technique to facilitate the process of understanding the learner's assimilation that can empower both the teachers and the learners in understanding the topical assimilation that the learners have after experiencing a learning journey.

3 AIMS OF THE RESEARCH

Our broad research is to focus on the learning process as a whole and reflect on the assimilation patterns using computational analysis of learner's data that they construct during their learning journey. During my initial stages of work, I have examined if learning activity signals, such as the time spent on using course resources and the number of course resources, examined by the learner, can predict the outcome. Through that work, I have observed that we can create models that could predict the outcome with 82% accuracy (Praseeda et al., 2019). We also observed that there was no one average model that could be used to predict the outcomes of the learner and hence we could validate "The Myth of Average", which states that we cannot use averaging models when the subjects of the models are individuals (Rose, 2016).

To understand the learner assimilation patterns, it was necessary to understand how learners learnt various topics in a course. To understand the topical assimilation pattern, a single value in terms of marks was not sufficient to represent the assimilation done by the learner. The questions we were interested to pursue were:

- RQ1: What are some representation techniques that capture assimilation patterns in a learning process?
- RQ2: How do we observe assimilation patterns of learning in such representation spaces?
- RQ3: How can these representation techniques empower learner and other stakeholders after the learner's undergo a learning journey?

4 CURRENT STATUS AND RESULTS

The online learning process involves learners perform various activities like attending lectures, solving quizzes, and writing exams. Irrespective of the type of assessment, the learner's outcomes are reduced to Aggregate percentage, Cumulative Grade Point Average (CGPA) and similar summative metrics. A single value is limited in reflecting the learner's assimilation in a course. Instead of a single value, there are ways in which learner's diversity is represented as jagged profile that indicates the student's proficiency and student's profiles, and they consist of multiple dimensions. Averaging those values to a single entity does not consider the individuality of the learners (Rose, 2016). Storing, reasoning, and comparing multidimensional learner's proficiency is important for teachers to understand how the students are assimilating in the course. It is hard for individuals to compare and analyze multidimensional entities intuitively (Li & Giudice, 2013). We wanted a representation technique that would help us compare, contrast, and analyze different learners intuitively and we chose two-dimensional maps as a representation technique to show the assimilation patterns as we are capable to observing and analyzing maps easily (Niedomysl et al., 2013; Li & Giudice, 2013).

The two-dimensional (2D) map, is called a *Discourse Learning Map (DLM)* for a course (c), contains m topics (T). It is formally defined as:

$$DLM_c = (T, L, \gamma, \mu)$$

In the above equation, T refers to the topics of the course. L refers to the learning entities. *Learning Entities* are any object of pedagogic interest in the course. Learning entities include: learning resources, assessments, learner proficiency, topics in a course. The function γ : L x T → [0,1] represents a mapping function, which maps the relevance or similarity of learning entity to each topic in the course. This mapping makes each learning entity to be interpreted as m-dimensional vector called *Polylines.* Polyline is representative data structure for all learning entities and is defined as a vector that represents the topical distribution of each topic in a course. The value in the i^{th} cell of each vector represents the topical similarity of the learning entity to the corresponding topic. Polylines for each learning entity is created with the help of Word2Vec embeddings generated from pre trained BERT model (Devlin et al., 2019). Learners Polyline are computed based on the knowledge they construct in the form of essays, seminars, and group discussion, as they give us evidence of their assimilation and are referred as learner's contribution. The contributions are evaluated by Teaching Assistant (TA) or teacher and the Polylines of the learners are adjusted based on the grades received. Learners create various contributions during the course and all the learner's contributions polyline vector's maximum value represent the learner's current polyline. The BERT model is trained on representative resource corpus to learn the course. These large word embeddings are created for topics of the course, resources in the course, learner's contributions in the form of essays, seminars, etc. These embeddings are converted to polylines by computing cosine similarity between embeddings and each topic embeddings to get a polyline for each learning entity.

The function μ maps polyline of all Learning Entities to a two-dimensional progression space. Dimensionality reduction has been achieved through various techniques like Multidimensional Scaling, Principal Component Analysis, Linear Discriminant Analysis, etc. (Shephard et al., 1972; Jolliffe, 2002). These methods only preserve the relative distance between high dimensional elements, when mapped to low dimensional space. We must optimize and preserve progression and semantic proximity. A two-dimensional space has *Progression* if the learning entities hold the covering property. The covering property is defined as if e and d are two learning entities and PL indicates the polyline of these entities, then e is said to cover d if e's value in all dimension is no less than d and in atleast one dimension its strictly greater than d as shown in the figure below.

$$PL(e) \sqsupseteq PL(d) \Leftrightarrow \forall i, p_i(e) \geq p_i(d) \wedge \exists j, p_j(e) > p_j(d)$$

Figure 1. Equation that shows the covering property among two polylines.

The idea is analogous to Progression in learning science, which talks about progress in a course. To achieve progression, we chose a constrained Multidimension Scaling Algorithm (cMDS) algorithm. MDS algorithm arranges the high dimensional vectors in low dimension by stress function which optimizes the distance between points in high dimension in low dimension. We modified the constraints by using a gradient descent where we maximize progression over stress between the points. It is difficult to get a mapping where all points hold the covering property. So, we choose a mapping where we get a local optimum based on least violations in terms of progression.

Once we obtained the High Dimension to Low Dimension Mappings, we visualized it in a two-dimensional space to observe the assimilation patterns. The map was designed such that, the left bottom corner or (0,0) position was called the START of the course and top right position (1,1) was referred as the logical END of the course which may not coincide with the actual end of the course. Learning Entities, that are mapped closer in this space have topical similarity. We mapped the learning entities – resources and learners of a course called "Network Science for Web" into the discourse map; we observed that entities that had similar topical distribution in the high dimensional space were mapped closer in the Map. NSW (Network Science for the Web) discourse map is shown in Figure 2. The learners who are placed closer to the START position indicate that they have just started learning the course and are yet to construct knowledge. As they assimilate more topics move their position will move away from the START.

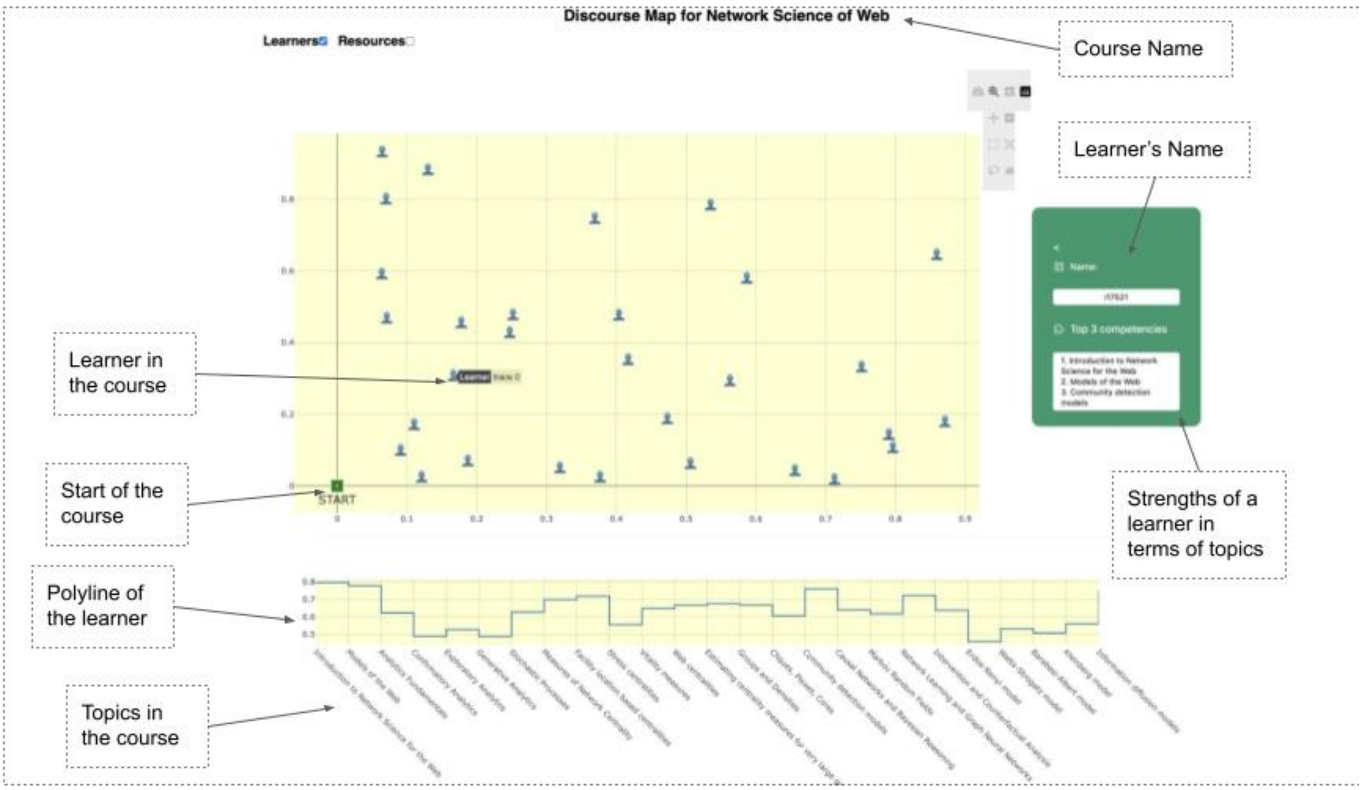


Figure 2: Design of the discourse map with learners and resources mapped to 2D space

The discourse learning map is not limited to observing the learning patterns, it also serves as a learning space where the topics of the course, resources in the course, learner's contributions, learners and many other visual elements like learning pathways and regions can be created. This representation helps in seeing how a class has assimilated in their topical interest, learner's contributions act as a resource that other learners can consume during their learning journey, learning pathways can help the learners in planning their next activity based on their location in the map.

5 NEXT STEPS

We have succeeded in identifying 2D map as a good representation technique, and were able to map learning entities to the discourse map with partial progression and semantic proximity (RQ1). We need to find a dimensionality reduction method that optimizes progression. We must also discover visualization techniques and elements that are used to design maps, that will help us visualize and observe the topical assimilation by learners. We must also implement ways to map other elements like pathways, regions that would help us observe these patterns intuitively (RQ2). Once we have designed the map with all the learners and other learning entities mapped to this space, we must evaluate the effectiveness of discourse map with all the stakeholders of the course: students, teachers, teaching assistants (TA) etc. to understand how the map would empower the students in the learning journey, help the teacher and TA in facilitating the learning process (RQ3).

REFERENCES

Mukunda,K.V. (2009). *What did you ask at school today?* Harper Collins.

Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.

Ausubel, D.P. (1968). *Educational Psychology: A cognitive view.* New York: Holt, Rinehart and Winston.

Brady, L. (1996). Outcome based education: a critique. *The Curriculum Journal*, 7(1):5–16.

Tam, M. (2014). Outcomes-based approach to quality assessment and curriculum improvement in higher education. *Quality Assurance in Education*, 22(2):158–168.

Morgan, H. (2016). Relying on High-Stakes Standardized Tests to Evaluate Schools and Teachers: A Bad Idea. The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 89(2), 67-72, DOI: 10.1080/00098655.2016.1156628.

Dendir, S. R. and Maxwell, S. (2020). Cheating in online courses: Evidence from online proctoring. *Computers in Human Behavior Reports.*

Jonassen, D. H. (1991). Evaluating Constructivistic Learning. *Educational Technology*, 31(9), 28–33. JSTOR

Stevens, D. D. and Levi, A. (2013). *Introduction to rubrics: an assessment tool to save grading time, convey effective feedback, and promote student learning*. 2nd ed. Sterling, Va., Stylus.

Rose, T. (2016). *The End of Average: How to succeed in a world that values sameness*. San Francisco, CA, HarperOne.

Woolfolk, A. and Kapur, P. (2019). *Educational Psychology*. Fourteenth Edition. Pearson.

Carwile, J. (2007). A constructivist approach to online Teaching and Learning. *Inquiry*, 12(1), Spring, 68 – 73.

Susnjak, T., Ramaswami, G.S. and Mathrani A. (2022). Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*.

Jivet, I. (2018). Adaptive learning analytics dashboard for self-regulated learning. In: *Proceedings of LAK'18*, 173-179. ACM

Praseeda, Srinivasa, S. and Ram, P. (2019). Validating the Myth of Average through Evidences. In: *Proceedings of The 12th International Conference on Educational Data Mining*, 631 – 634.

Niedomysl, T., Elldér, E., Larsson, A., Thelin, M., & Jansund, B. (2013). Learning Benefits of Using 2D Versus 3D Maps: Evidence from a Randomized Controlled Experiment. *Journal of Geography.*

Li, H. & Giudice, N. A. (2013). The effects of 2D and 3D maps on learning virtual multi-level indoor environments. *MapInteract '13: Proceedings of the 1st ACM SIGSPATIAL,* 7–12.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv, abs/1810.04805.

Shepard R. N. Romney A. K. Nerlove S. B. & Mathematical Social Science Board. (1972). *Multidimensional scaling; theory and applications in the behavioral sciences*. Seminar Press.

Jolliffe, I.T. (2002). *Principal Component Analysis*, Springer New York, NY.

The 5th Workshop on Predicting Performance Based on the Analysis of Reading Behavior

Brendan Flanagan[1*], Atsushi Shimada[2], Fumiya Okubo[2], Owen H.T. Lu[3], Stephen J.H. Yang[4], Hiroaki Ogata[1]
[1] Kyoto University, Japan
[2] Kyushu University, Japan
[3] National Chengchi University, Taiwan
[4] 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 and LAK22 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, and a multi-source dataset for higher education programming classes consisting of reading behavior and coded programming logs. As with previous years, additional information on lecture schedules and syllabus will also enable the analysis of learning context for further insights into the preview, in-class, and review reading strategies that learners employ. Participant contributions will be collected as evidence in a repository provided by the workshop and will be shared with the wider research community to promote the development of research into reading analysis systems.

Keywords: Student Performance Prediction, Data Challenge, Reading Behavior, Programming education

1 WORKSHOP BACKGROUND

Digital learning materials especially digital textbooks are a core part of modern education, and the adoption of digital textbooks in education is increasing. Digital textbooks and e-books are being introduced into education at the government level in a number of countries in Asia (Ogata et al., 2015). This has prompted research into not only the use of such materials within the classroom, but also the collection and analysis of event data collected from the systems that are used for support and distribution (Flanagan et al., 2018; Ogata et al., 2017; Ogata et al., 2015). In addition to its advantages on students' learning, digital text readers are capable of recording interactions regarding students' reading behaviors. As the materials are read by students using the system, the action events are recorded, such as: flipping to the next or previous page, jumping to different pages, memos, comments, bookmarks, and drawing markers to indicate parts of the learning materials that learners think are important or find difficult.

Despite the increase in use, research analyzing students’ interaction with digital textbooks is still limited. Recent review study (Peña-Ayala et al., 2014) revealed that almost half of the papers in Learning Analytics (LA) and Educational Data Mining (EDM) fields are using data from Intelligent Tutoring Systems (ITS) or Learning Management Systems (LMS). Previous research into the reading behavior of students has been used in review patterns, visualizing class preparation, behavior change detection, and investigating the self-regulation of learners (Yin et al., 2015; Ogata et al., 2017; Shimada et al., 2018; Yamada et al., 2017). The analysis of reading behavior can be used to inform the revision of learning materials based on previous use, predict at-risk students that may require intervention from a teacher, and identify learning strategies that are less effective and provide scaffolding to inform and encourage more effective strategies. The digital learning material reader can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

The main motivation of this workshop is to foster research into the analysis of students’ interaction with digital textbooks, and find new ways in which it can be used to inform and provide meaningful feedback to stakeholders, such as: teachers, students and researchers. This proposal builds upon previous workshops that have focused on student performance prediction based on reading behavior. In previous years at LAK and other international conferences, there have been workshops that have offered open ended data challenges to analyze e-book reading logs and predict the final grade score of learners (Flanagan, 2018; Flanagan, 2019; Flanagan, 2020; Flanagan, 2021; Flanagan, 2022), with 16, 14, 17, 12, and 23 participants respectively. However, to-date the data challenges have targeted onsite classes in higher education and secondary school settings.

This year we offered participants a new challenge to predict academic performance in higher education programming classes (Lu, 2022). The code name of this dataset is LBLS160 (Learning Behavior and Learning Strategy), which contains students' learning behaviors collected from e-book and programming environments, and questionnaire measurements from two learning strategies: Self-regulated Learning (SRL) and Strategy Inventory for Language Learning (SILL). The number 160 indicates that one hundred and sixty students were involved in this project. 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 implementing 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.

Participants can upload their scores to the workshop website to check the results of the evaluation periodically. A leaderboard will be provided with the best evaluation score that each participant has achieved to encourage competition between teams. Final data challenge results of prediction models will be confirmed by submission of prediction models for formal evaluation.

2 OBJECTIVES

While we welcome research questions from all participants, and we expect to emphasize the following topic which the organizers feel attention should be paid. Low retention and high failure rates are important problems in education (Villagrá-Arnedo et al., 2017). However, studies have shown that timely interventions for at-risk students can be effective in helping change their behaviors (Arnold et al., 2012; Tanes et al., 2011). Therefore, focusing on the early detection of at-risk students is an essential step to changing student's behavior for greater success. This broader task may be approached from the following perspectives:

- Student reading behavior self-regulation profiles spanning the entire course
- Preview, in-class, and review reading patterns
- Student engagement analysis; and behavior change detection
- Visualization methods to inform and provide meaningful feedback to stakeholders

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. Baloian et al., proposed a classification method based on the Dempster-Shafer plausibility theory with the aim of increasing both explainability and accuracy of student performance prediction. It was found to perform better than other explainable models, with accuracy being close to that achieved by high performing non-explainable classifiers. Leelaluk et al., proposed using LSTM with attention mechanism to help the model focus on important reading behaviors, and as a result it achieved higher accuracy and earlier prediction compared to previous models. Specific reading behavior at particular times in the course were identified as features that may affect a student's final score. Li et al. focused on quantitative analysis and used SHAP values over a trained SVM and a decision tree to offer the machine learning model's explainability on students' prediction outcomes. Their results demonstrated how to explain individuals who have been predicted as a risk or an outstanding student, and how to design a personalized tutoring plan accordingly. Bobea et al. investigated applying domain adoption to the task of cross-semester performance prediction, which is an important subtask in academic prediction performance. By using domain adoption PCA, it was found to improve model performance for cross-semester application, in some cases with large increases in accuracy compared to normal cross-semester models. The proceedings of the workshop can be found on the following website: https://sites.google.com/view/lak23datachallenge.

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 267-270). ACM.

Flanagan, B., & Ogata, H. (2018). Learning Analytics Platform in Higher Education in Japan, Knowledge Management & E-Learning (KM&EL), 10(4), 469-484.

Flanagan, B., Chen, W., & Ogata, H. (2018). Joint Activity on Learner Performance Prediction using the BookRoll Dataset. In *International Conference on Computers in Education (ICCE2018)* (pp. 480-485).

Flanagan, B., Shimada, A., Yang, S., Chen, B.L., Shih, Y.C., & Ogata, H. (2019). Predicting Performance Based on the Analysis of Reading Behavior: A Data Challenge, *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2020). The 2nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 10th International Conference on Learning Analytics and Knowledge*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2021). The 3nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2022). The 4nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 12th International Conference on Learning Analytics and Knowledge*.

Lu, O.H., Huang, A.Y., Flanagan, B., Ogata, H., & Yang, S.J. (2022). A Quality Data Set for Data Challenge: Featuring 160 Students' Learning Behaviors and Learning Strategies in a Programming Course. *30th International Conference on Computers in Education Conference Proceedings*, (pp. 64-73).

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.

LAK23 Trace-SRL: The Workshop on Measuring and Facilitating Self-regulated Learning based on Trace data

Yizhou Fan
Peking University
fyz@pku.edu.cn

Mladen Raković
Monash University
mladen.rakovic@monash.edu

Shaveen Singh
Monash University
shaveen.singh@monash.edu

Megan Wiedbusch
University of Central Florida
meganwiedbusch@knights.ucf.edu

Daryn Dever
University of Central Florida
daryn.dever@ucf.edu

Xinyu Li
Monash University
xinyu.li@monash.edu

Joep van der Graaf
Radboud University
joep.vandergraaf@ru.nl

Lyn Lim
The Technical University of Munich
lyn.lim@tum.de

Inge Molenaar
Radboud University
inge.molenaar@ru.nl

Maria Bannert
The Technical University of Munich
maria.bannert@tum.de

Roger Azevedo
University of Central Florida

roger.azevedo@ucf.edu

DraganGašević
Monash University
dragan.gasevic@monash.edu

ABSTRACT: Trace-SRL is an interactive workshop on measuring and facilitating self-regulated learning (SRL). Prior research has shown that SRL skills are essential for successful life-long learning. Measuring SRL based on unobtrusive trace data and facilitating SRL based on real-time analysis. Such trace data have been pointed out as very valuable research directions. However, major challenges and significant gaps in this area are still many, such as i) the detection, measurement, and validation of SRL processes with trace data is still a much-debated issue within the SRL community; and ii) the design principles for effective interventions and the complex conditions and contexts when these interventions facilitated learning are still not known. Therefore, we aim to improve the measurement and facilitation of SRL by hosting this full-day workshop, which will i) initiate a project-level dialogue to foster cross-team collaborations and ii) provide the participants with hands-on opportunities to experience the measurement and facilitation of SRL. Expected outcomes are forming a community of practice, potential collaborative projects, and possible follow-up joint publications and grant submissions.

Keywords: Learning analytics, Self-regulated learning, Trace data, Measurement protocols, Learning interventions, Scaffoldings and Dashboards

1 BACKGROUND

1.1 Challenges

One of the central focuses of education is to foster the competency of Self-Regulated Learning (SRL) amongst learners. Self-regulation can improve learning outcomes as revealed by the positive relation between SRL processes and measures of learning (Azevedo et al., 2022). **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) to think-aloud protocols (Bannert, 2007) and **trace-based measurement** (Siadaty et al., 2016; Fan et al., 2022). Trace-based methods are becoming a popular approach to measuring SRL (Saint et al., 2022), since trace data can unobtrusively record dynamic instances of cognition and metacognition in authentic learning environments and thus operationalize "what learners do as they do it" (Winne, 2010) and has been utilised in a number of studies (Siadaty et al., 2016; Fan et al., 2022). However, the detection, measurement, and validation of SRL processes with trace data is still a much-debated issue within the SRL community (Winne, 2020). Therefore, we would like to propose this **interactive workshop (aim 1)** for interested interdisciplinary researchers from different learning analytic projects focused on SRL to examine current 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 recognised, numerous previous studies have also shown that learners by themselves often experience difficulties in adequately and effectively self-regulating their learning across tasks, domains, and contexts (Winne, 2010). Despite the opportunities learners have to practice and hone them, SRL skills remain underdeveloped (Bjork et., 2013). Therefore, learners need to be supported to successfully regulate their learning and achieve their learning goals. Different types of interventions, such as **scaffolding, dashboards, or personalized feedback**, have been designed in learning analytics to effectively support learners' SRL 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 (Devolder et al., 2012). Importantly, the complex conditions and contexts when these interventions facilitate and enhance learning are emerging (Guo, 2022). Therefore, this **interactive workshop (aim 2)** will address these challenges by sharing how different interventions (e.g., artificial agents) 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 metacognition and self-regulation during learning.

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 detecting, measuring, and analysing SRL; iii) understand how combining student data and artificial intelligence can be used to create actionable insights into students' learning; iv) design new representations/forms of communicating SRL scaffolding, dashboards, or feedback to facilitate learning. From the participants' perspective, we expect to: i) improve the knowledge and skills of participants about SRL measurement, learning processes, and SRL support; ii) produce a repository of new requirements, considerations and approaches of instruments for SRL; iii) build a research community, foster partnerships, and facilitate deployment of similar tools and analytics platforms as collaborative projects; and iv) explore opportunities for joint publications (e.g., a journal special issue), grants, and future workshops resulting from the collaborations. The outcomes of the workshop will be housed on a Google Site. More specifically, we have two objectives different from other workshops or research tracks in LAK23:

- **Initiate a project-to-project level dialogue to foster deep, cross-team collaboration**. By bringing together SRL-related significant projects that measure and facilitate SRL following different analytic approaches, we aim to further promote research and practice that looks at SRL from a more comprehensive dimension of context, compared to single studies.
- **Provide hands-on opportunity to experience the measuring and facilitating of SRL**. Participants will also explore the FLoRA project and platform integrated with various instrumentation tools and personalised scaffoldings, and they will be able to explore their own multi-channel data and co-design possible SRL-related scaffoldings and feedback representations for learners and instructors.

2 ORGANISATIONAL DETAILS (FULL-DAY WORKSHOP SCHEDULE)

Table 1: Proposed Full-day Workshop Schedule (3.5 hours + 3.5 hours)

Timing	Descriptions	Hosts
Part 1: Morning Section		
10 minutes	Welcome & Introduction (Morning Section: focus more on Measuring SRL)	Megan Wiedbusch
90 minutes	Presentations about measuring and facilitating SRL using trace data	Daryn Dever

	• Multimodal modeling to validate digital traces of SRL and examine the robustness of their predictive validity when scaled to naturalistic settings (by Matthew L Bernacki and Linyu Yu) • Measuring the impact of instructional design on students' planning process via multi-level trace-data clustering (by Zhongzhou Chen) • Multicomponential analysis of self-regulated learning during diagnostic reasoning (by Alejandra Ruiz Segura) • Exploring the value of trace data for self-regulated learning in game-based learning environments (by Saerok Park) • Capturing, Modeling, and Transferring Trace Data between Simulated and Real-World Skill Development (by Megan Wiedbusch)	
30 minutes	Coffee Break and Socialization	All
40 minutes	Roundtable Discussion (Previous presenters + Audience) Guided by structured questions	Daryn Dever
30 minutes	Presentation of FLoRA Analytics Platform and Hands-on Task, then brainstorming about new direction of measuring SRL using trace data	Xinyu Li
10 minutes	Summarizing the morning section & Next Steps	Roger Azevedo
Part 2: Afternoon Section		
10 minutes	Welcome & Introduction (Afternoon Section: focus more on Facilitating SRL)	Mladen Raković
90 minutes	Presentations about measuring and facilitating SRL using trace data • Explainable, theory-guided prediction modeling to inform design and delivery of digital skill trainings that improve facility for SRL (by Matthew L Bernacki & Robert D. Plumley) • An ordered network analysis on Personalised Scaffolding for Self-regulated Learning (by Tongguang Li) • Simulations as Platforms for Capturing, Measuring, and Facilitating Self-regulated Learning (by Daryn Dever) • Human Digital Twins as a Research Platform to Study, Model, and Simulate Self-Regulated Learning in STEM (by Roger Azevedo) • Reducing Procrastination on Introductory Physics Online Homework for College Students Using a Planning Prompt Intervention (by Zachary Felker)	Mladen Raković
30 minutes	Coffee Break and Socialization	All
40 minutes	Roundtable Discussion (Previous presenters + Audience) Guided by structured questions	Mladen Raković
30 minutes	Presentation of FLoRA Scaffolding system and Hands-on Task, then brainstorming about new direction of facilitating SRL using trace data	Xinyu Li
10 minutes	Summarizing the afternoon section & Next Steps	Roger Azevedo

The event will be an open and hands-on workshop. The organization of the workshop will revolve 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, and each submission should contain both the measuring and facilitating of SRL and describe a complete picture of the SRL-related study. The main purpose of this arrangement is to make the two sections in the morning and afternoon to echo each other and provide workshop participants with a global understanding and in-depth discussion about SRL. The submission timeline will follow the timeline suggested by the conference organizers. All attendees will have the opportunity to discuss with the presenters in the roundtable and brainstorming sessions, and will also have hands-on experiences with SRL measurement and scaffolding design activities guided by organizers. We anticipate a registration of about 30 participants. We will use #LAKTRACESRL when referencing this event on social media.

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 at previous conferences). At the same time, we will send invitations to specific research teams who are working on measuring and facilitating SRL. 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 to maximise the impact of workshop.

REFERENCES

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., ... & Cerezo, R. (2022). Lessons learned and future directions of metatutor: leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. Frontiers in Psychology, 13.

Bannert, M. (2007). Metakognition beim lernen mit hypermedien. Waxmann Verlag.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. Annual review of psychology, 64, 417-444.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self‐regulated learning in computer‐based learning environments: systematic review of effects of scaffolding in the domain of science education. Journal of Computer Assisted Learning, 28(6), 557-573.

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., ... & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. Metacognition and Learning, 1-39.

Guo, L. (2022). Using metacognitive prompts to enhance self‐regulated learning and learning outcomes: A meta‐analysis of experimental studies in computer‐based learning environments. Journal of Computer Assisted Learning, 38(3), 811-832.

Pintrich, P. R., & De Groot, E. V. (1991). Motivated strategies for learning questionnaire. Journal of Educational Psychology.

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. Computers & Education: Artificial Intelligence, 3, 100060.

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

Winne, P. H. (2010). Bootstrapping learner's self-regulated learning. Psychological test and assessment modeling, 52(4), 472.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. Computers in Human Behavior, 112, 106457.

How to Open Science: Promoting Principles and Reproducibility Practices within the Learning Analytics Community

Aaron Haim, Stacy T. Shaw, Neil T. Heffernan
Worcester Polytechnic Institute
ahaim@wpi.edu, sshaw@wpi.edu, nth@wpi.edu

ABSTRACT: Across the past decade, open science has increased in momentum, making research more openly available and reproducible. In parallel, learning analytics, as a subfield of education technology, has been increasing as well, providing more accurate statistical models and integrations to improve learning. However, open science and learning analytics rarely tend to intersect, causing a bit of difficulty when trying to reuse methodologies, datasets, analyses for replication, reproduction, or an entirely separate end goal. In this tutorial, we will provide an overview of open science principles and their benefits and mitigation within research. In the second part of this tutorial, we will provide an example on using the Open Science Framework to make, collaborate, and share projects. The final part of this tutorial will go over some mitigation strategies when releasing datasets and materials such that other researchers may easily reproduce them. Participants in this tutorial will gain a better understanding of open science, how it is used, and how to apply it themselves.

Keywords: Open Science, Learning Analytics, Reproducibility

1 INTRODUCTION

Open Science is a term used to encompass making methodologies, datasets, analyses, and results of research publicly accessible for anyone to use freely (Kraker, 2011; Vicente-Saez, 2018). This term started to frequently occur in the early 2010s when researchers began noticing that they were unable to replicate or reproduce prior work done within a discipline (Spellman, 2015). There also tended to be a large amount of ambiguity when trying to understand what process was followed to conduct a study or whether a specific material was used but not clearly defined. Open science, as a result, started to gain more traction to provide greater context, robustness, and reproducibility metrics with each subtopic encompassed under the term receiving their own formal definition and usage. The widespread adoption of open science began to explode exponentially when large scale studies conducted in the mid-2010s found that numerous works were difficult or impossible to reproduce and replicate in psychology (Open Science Collaboration, 2015) and other disciplines (Baker, 2016).

Some principles commonly referred to as part of open science and its processes: open data, open materials, open methodology, and preregistration. **Open Data** specifically targets datasets and their documentation for public use without restriction, typically under a permissive license or in the public domain (Murray-Rust, 2008). Not all data can be openly released (such as with personally identifiable information); but there are specifications for protected access that allow anonymized datasets to be released or a method to obtain the raw dataset itself. **Open Materials** is similar in regard except for targeting tools, source code, and their documentation (Johnson-Eilola, 2002). This tends to be synonymous with **Open Source** in the context of software development, but materials are used to encompass the source in addition to available, free-to-use technologies. **Open Methodology** defines the full workflow and processes used to conduct the research, including how the participants were gathered, what was told to them, how the collected data was analyzed, and what the final results were (Kraker, 2011). The methodologies typically expand upon the original paper, such as technicalities that would not fit in the paper format. Finally, **Preregistration** acts as an initial methodology before the start of an experiment, defining the process of research without knowledge of the outcomes (Nosek, 2018; Nosek, 2019). Preregistrations can additionally be updated or created anew to preserve the initial experiment conducted and the development as more context is generated.

Open science principles and reproducibility metrics are becoming more commonplace within numerous scientific disciplines. Within many subfields of educational technology, such as learning analytics, however, the adoption and review of these principles and metrics are neglected or sparsely considered (Nosek, 2022). There are some subfields of education technology that have taken the initiative to introduce open science principles (special education, (Cook, 2018); gamification, (García-Holgado, 2020); education research, (Makel, 2019); however, other subfields have seen little to no adoption. Concerns and inexperience in what can be made publicly available to how to reproduce another's work are some of the few reasons why researchers may choose to avoid or postpone discussion on open science and reproducibility. On the other hand, lack of discussion can lead to tediousness and repetitive communication for datasets and materials or cause a reproducibility crisis (Baker, 2016) within the field of study. As such, there is a need for accessible resources and understanding on open science, how it can be used, and how to mitigate any potential issues that may arise within one's work at a later date.

Admitting our own initial lack of proper adoption and reproducibility first, in this tutorial, we will cover some of the basic principles of open science and some of the challenges and mitigation strategies associated with education technology specifically. Next, we will provide a step-by-step explanation on using the Open Science Framework (OSF) to create a project, collaborate with other researchers, post content, and preregister a study. Using examples from the field of educational technology, we will showcase how to incorporate open science principles, in addition to practices that, when implemented, would improve reproducibility.

2 FORMAT AND TIMELINE OF TUTORIAL

The tutorial will occur over three hours and focuses on introducing some common open science principles and their usage within education technology, providing an example on using OSF to create a project, post content, and preregister studies, and using previous papers to apply the learned principles and any additional reproduction mitigations. Based on prior workshops (Shaw, 2022), we assume around five to ten participants will attend. Attendee and tutorial information will be provided through social media. An outline of this tutorial can be found below:

- First, we will provide a presentation on an overview of a few problems when conducting research. Using this as a baseline, we will introduce open science and its principles and how they can be used to nullify some of these issues and mitigate others. In addition, we will attempt to dispel some of the misconceptions of these principles.

- Second, we will provide a live example of using OSF to make an account, create a project, add contributors, add content and licensing, and publicize the project for all to see. Afterwards, we will provide a guide to creating a preregistration, explaining best practices, and identifying how to create an embargo. Additional features and concerns, such as anonymizing projects for review and steps required to properly do so, will be shown.

- Third, we will discuss reproducibility metrics within work when providing datasets and materials. This will review commonly used software and languages (e.g., Python, RStudio) and how, without any steps taken, most work tends to be extremely tedious to reproduce or are not reproducible in general. Afterwards, we will provide some mitigation strategies needed to remove these concerns.

- Finally, we will take some existing papers either from the author's own research or from prior education technology conferences that do not meet some open science principles or cannot easily be reproduced and apply what has been learned across the entire tutorial. We will use a few papers, each containing different issues, and apply the necessary steps needed to reproduce the results within the paper.

3 DISSEMINATION OF INFORMATION

The dissemination of information for this tutorial will be provided before and after the conference. Before the conference, information about the tutorial itself will be stored on an OSF project, containing references to the papers used within the final part of the tutorial, any slides to be used within the conference, and additional resources that could provide better understanding of the issues and nuances of avoiding open science and reproducibility metrics. A website separate to the OSF project will also be set up containing the following information for ease of consumption; however, this will only be used as an alternative to the project in case the website disappears at some point in the future.

After the conference, any resources created or recordings taken will be uploaded to the project for preservation. Alternative links will be provided to separate sites for more formal hosting (e.g., videos on YouTube). As this tutorial wants to repeat and expand upon open science and reproducibility at prior workshops across conferences (e.g., *Using the Open Science Framework to promote Open Science in Education Research in the 15th International Conference on Educational Data Mining* (Shaw, 2022)) an additional project will be created on the OSF website containing components pointing to all previous conferences and resources discussed.

REFERENCES

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature 533, 7604, 452–454. https://doi.org/10.1038/533452a

Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. Exceptional Children, 85(1), 104-118. https://doi.org/10.1177/0741932516637198

García-Holgado, A., García-Peñalvo, F. J., de la Higuera, C., Teixeira, A., Ehlers, U. D., Bruton, J., ... & Burgos, D. (2020, October). Promoting Open Education Through Gamification in Higher Education: the OpenGame project. In Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality. 399-404. https://doi.org/10.1145/3434780.3436688

Johnson-Eilola, J. (2002). Open Source Basics: Definitions, Models, and Questions. In Proceedings of the 20th Annual International Conference on Computer Documentation (Toronto, Ontario, Canada) (SIGDOC ’02). Association for Computing Machinery, New York, NY, USA, 79–83. https://doi.org/10.1145/584955.584967

Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. International Journal of Technology Enhanced Learning 3(6), 643–654. https://doi.org/10.1504/IJTEL.2011.045454

Makel, M. C., Smith, K. N., McBee, M. T., Peters, S. J., & Miller, E. M. (2019). A path to greater credibility: Large-scale collaborative education research. AERA Open, 5(4). https://doi.org/10.1177/2332858419891963

Murray-Rust, P. (2008). Open data in science. Nature Precedings 1(1), 1. https://doi.org/10.1038/npre.2008.1526.1

Nosek, B., Ebersole, C., DeHaven, A., Mellor, D. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences 115(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B., Beck, E., Campbell, L., Flake, J., Hardwicke, T., Mellor, D., Veer, A., and Vazire, S. (2019). Preregistration Is Hard, And Worthwhile. Trends in Cognitive Sciences 23(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nosek, B. (2022). Making the Most of the Unconference. Presented at the Unconference on Open Scholarship Practices in Education Research. Available at https://osf.io/9k6pd/

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Shaw, S., & Sales, A. (2022). Using the Open Science Framework to promote Open Science in Education Research. In Proceedings of the 15th International Conference on Educational Data Mining. International Educational Data Mining Society. 853-853. https://doi.org/10.5281/zenodo.6852996

Spellman, B. (2015). A Short (Personal) Future History of Revolution 2.0. Perspectives on Psychological Science 10(6), 886–899. https://doi.org/10.1177/1745691615609918

Vicente-Saez, R. & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. Journal of Business Research 88, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

4th Workshop on Culturally Aware Learning Analytics: Value-sensitive design

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

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

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

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

ABSTRACT: Learning analytics (LA) have been implemented to improve teaching and learning practices in different countries using a variety of approaches and with different levels of success. To effectively transfer an LA system from one country to another, we need to carefully consider contextual, technical, and cultural factors. In this interactive workshop, we will explore the role of the cultural values that are important for the successful adoption and use of LA systems around the world: 1. the acceptance of LA services and related ethical concerns, 2. the design of LA systems, and 3. the evaluation of LA interventions. This one-day workshop will focus on the role of culture in LA from a value-sensitive perspective. In particular, we will: 1. discuss and identify possible cultural differences and similarities for the wider adoption of LA systems at scale, and 2. introduce and practice several culture- and value-sensitive design methods on selected LA tools.

Keywords: Cultural awareness, learning analytics, cultural values, scalability, value-sensitive design

1 INTRODUCTION

Over the last decade, we have seen many examples of learning analytics (LA) systems being implemented and used in various countries in different ways, however, often at a limited scale (Viberg et al. 2018). The scalability problem can be associated with, but not limited to, the differences in the expectations that stakeholders such as teachers and students have towards LA (e.g., Gray et al. 2022), concerns about ethical and privacy issues (e.g., Hoel & Chen 2019), and heterogeneity in the effects of LA interventions on learners across countries (e.g., Davis et al. 2017; Kizilcec & Cohen, 2017). All these differences can make the transfer of LA systems from one country to another challenging since there are varying contextual, technical, and also cultural factors that play an important role. Whereas technical and contextual aspects of LA systems' design and implementation have been addressed by

LA scholars and practitioners, important cultural factors have hitherto received limited attention (Viberg et al. *in press*). There have been efforts in the community to examine contextual and cultural factors, including at conferences and new journals on the topic (e.g., *Computer-based Learning in Context*). Considering the role that culture – both at the individual and national level – may have on the adoption of LA systems across countries and earlier research in related domains (see e.g., Leidner & Kayworth, 2006) on the role of culture in information systems design, or Baker et al. (2019) for computer-based learning systems), there is a critical need for more research and discussion of culture in LA.

Previous research in this area has hypothesized how culture might play a role in the implementation process of LA systems (e.g., Vatrapu, 2011). Although not many studies have looked at how students' learning patterns and learning strategies in higher education differ across cultures (Marambe et al., 2012), there is evidence that learners use the learning environment differently across countries and cultures (Liu et al. 2016; Rizvi et al. 2022). Cultural differences play a role in educational technology acceptance and use (e.g., Nistor et al. 2013) and students' collaborative learning practices (e.g., Vatrapu & Suthers, 2007). Cultural differences also influence the effectiveness of LA interventions that encourage self-regulated learning (Davis et al. 2017; Kizilcec & Cohen, 2017). This early body of work underlines the importance of not only considering the stakeholders' cultural values in the evaluation of LA systems, but also the importance of designing *culture*- and *value-sensitive* LA systems to increase their acceptance and use at a scale.

This workshop has two goals: **1**. discuss and identify possible cultural differences and similarities for the wider adoption of LA systems at scale, and **2.** introduce and practice several *culture*- and *value-sensitive* design methods on selected LA tools.

The LAK community would benefit from: (i) enabling a discussion and drafting a set of suggestions on how to create more inclusive tools that put stakeholders and their cultural values at the center of the design, and (ii) obtaining practical skills of working with the selected *value-sensitive design* methods.

We believe that the proposed workshop is of particular interest to the LAK community for several reasons, including a need to:

- scale up LA efforts across institutions worldwide
- provide inclusive and equitable quality education
- offer sustainable LA solutions that would facilitate and enhance the process of digital transformation of education
- enable and enhance stakeholders' agency in online learning settings.

2 WORKSHOP GOALS AND STRUCTURE

This workshop, building on the earlier three workshops conducted during the 12th International Conference on Learning Analytics and Knowledge (LAK22) and the European Conference on Technology Enhanced Learning (ECTEL2021 and ECTEL22), aims to:

1. further explore and raise awareness of possible influences of stakeholders' cultural values and preferences on the acceptance of LA systems and related privacy issues (e.g. their privacy concerns and trust in using LA tools), the design and evaluation of LA tools;
2. introduce the participants to culture- and value-sensitive design methods;
3. practice selected design methods (e.g., envisioning cards) that can be used to inform the inclusive and equitable human-centered design approach to LA.

To achieve these aims, we will use design approaches explored in the CHI community to facilitate a discussion on the possible cultural specificities to be considered for the design, adoption, and use of LA. As a proxy for 'culture' we will use Hofstede's model (Hofstede, 2001), defining what he claims to be national cultures, as a starting point for the analysis of chosen existing LA tools and the design of a selected LA tool. While these categories are contested as measures of national cultures, they contain elements that also in other contexts have been suggested to play a role in people's behavior and attitudes toward education (Mittelmeier et al., 2016) and technologies used in educational settings to improve students' learning (Baker et al., 2019). This means that even though they may not indicate the culture of entire nations, they might affect the adoption and implementation of LA worldwide. In other words, we do not rely on Hofstede's national cultural profiles to offer design recommendations for specific countries but explore cultural dimensions that may affect: i) students' expectations and attitudes towards LA services and related privacy concerns, and ii) the acceptance, design, and use of designed LA tools.

If there is interest, we would also be happy to support the formation of a SIG interested in the topic.

3 ORGANIZATIONAL DETAILS

- Interactive full-day workshop with maximum 30 participants.
- The workshop/tutorial activities that participants should expect: symposia elements, discussion groups, group-based demos, presentations.
- Proposed schedule in which the activities will take place:
  - The first part of the workshop will focus on the introduction to culturally aware LA, including related presentations and discussions.
  - The second part of the workshop will focus on the design activities, in which the workshop participants would be expected to work with some selected culture- and value-sensitive design methods with LA tools.
  - Finally, at the end of the workshop, we expect to draft a list of design considerations and implications based on the results of the aforementioned examinations and exercises as well as workshop participants' experience with respect to the integration (or lack thereof) of cultural aspects into the design, implementation, evaluation and use of LA tools.

REFERENCES

Baker, R.S., Ogan, A.E., Madaio, M., Walker, E. (2019). Culture in Computer-Based Learning Systems: Challenges and Opportunities. Computer-Based Learning in Context, 1(1), 1-13.

Davis, D., Jivet, I., Kizilcec, R., Chen, G., Hauff, C., & Houben, G. (2017). Follow the successful crowd: raising MOOC completion rates through social comparison at scale. In LAK17: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp.454-463. https://doi.org/10.1145/3027385.3027411

Gray, G., Schalk, A.E., Cooke, G., Murnion, P., Rooney, P., & O'Rourke, K. (2022). Stakeholders' insights on learning analytics: Perspectives of students and staff. Computers & Education, 187, 104550.

Hoel, T., & Chen, W. (2019). Privacy engineering for learning analytics in a global market – defining a point of reference. International Journal of Information and Learning Technology, 36(4), 288-298.

Hofstede, G. (2001). Cultural consequences: comparing values, behaviors, institutions, and organisations across nations. Sage Publications Ltd.

Kizilcec, R., & Cohen, G. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. Proceedings of the National Academy of Sciences, 114(17), 4348-4353. https://doi.org/10.1073/pnas.1611898114

Leidner, E. & Kayworth, T. (2006). A review of culture in information systems research: Toward a theory of information technology cultural conflict. MIS Quarterly 30(2), 357-399.

Liu, Z., Brown, R., Lynch, C. F., Barnes, T., Baker, R., Bergner, Y., & McNamara, D. (2016). MOOC Learner Behaviors by Country and Culture; an Exploratory Analysis. International Educational Data Mining Society.

Marambe, K., Vermunt, J., & Boshuizen, H. (2012). A cross-cultural comparison of student learning patterns in higher education. Higher Education, 64(3), 299-316.

Mittelmeier, J., Tempelaar, D., Rienties, B., & Nguyen, Q. (2016). Learning Analytics to Understand Cultural Impacts on Technology Enhanced Learning. International Association for Development of the Information Society. https://files.eric.ed.gov/fulltext/ED571378.pdf

Nistor, N., Gğüş, A., & Lerche, T. (2013). Educational technology acceptance across national and professional cultures: a European study. Educational Technology Research and Development, 61(4), 733-749. https://doi.org/10.1007/s11423-013-9292-7

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2022). Beyond one-size-fits-all in MOOCs: Variation in learning design and persistence of learners in different cultural and socioeconomic contexts. Computers in Human Behavior, 126, 106973.

Vatrapu, R. (2011). Cultural considerations in learning analytics. In LAK11 (pp.127-133). https://dl.acm.org/doi/pdf/10.1145/2090116.2090136

Vatrapu R., & Suthers D. (2007) Culture and Computers: A Review of the Concept of Culture and Implications for Intercultural Collaborative Online Learning. In: Ishida T., Fussell S.R., Vossen P.T.J.M. (eds) Intercultural Collaboration. IWIC 2007. Lecture Notes in Computer Science, vol 4568. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74000-1_20

Viberg, O., Jivet, I., & Scheffel, M. (in press). Designing Culturally Aware Learning Analytics: A Value Sensitive Perspective. Accepted for publication in Viberg, O., & Grönlund, Å. (Eds.). Practicable LA. Springer Nature.

3rd Workshop on Philosophy of Learning Analytics (POLA@LAK23)

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

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

ABSTRACT: This workshop aims to facilitate a dialogue related to philosophical stances and their impact on theory and practice in the field of Learning Analytics. LAK23 is the third year for this workshop which will build on the successes and reflections on POLA@LAK21 and POLA@LAK22. The workshop is designed to encourage conversation and collaborative ideation focused on the philosophical, conceptual, and theoretical foundations of LA. This year the workshop theme will be the inter-relationships between philosophies and theories in LA, and the extent to which they influence LA research and practice. The workshop is a half- day event. Participants will be invited to contribute a lightning talk abstract for review in advance of the workshop. Selected talks will provide a catalyst for dialogue throughout the workshop. The workshop will curate a collaborative, respectful environment to support robust, but intellectually stimulating and constructive conversations.

Keywords: Philosophy, learning analytics, theory.

1 BACKGROUND

As learning analytics (LA) continues to mature as a field, it also continues to draw expertise from multiple disciplines with practitioners from diverse backgrounds bringing varying skill sets. However, practitioners also bring their own disciplinary based philosophies to the field, whether they are conscious of them or not. Differences in philosophical assumptions underpinning varying approaches to LA can at times introduce confusion or even dissonance. These issues can be exacerbated when philosophical positions remain hidden and mitigated with increased transparency of diverse views.

We argue that there is a need to encourage conversations about the various ways that learning analytics research is (and can be) philosophically grounded, the extent to which our philosophies underpin our choice of theory, and the degree to which they influence our research and practice. To this end, this 3rd Workshop on The Philosophy of Learning Analytics (POLA@LAK23) will explore the theme:

The inter-relationships between philosophies and theories in learning analytics, and the extent to which they influence research and practice.

The workshop is particularly pertinent to the LAK23 conference theme, "Towards trustworthy learning analytics", as transparent accounting for philosophical positions makes visible to all the disciplinary assumptions that underpin different approaches to research and practice.

We can assume that all learning analytics practitioners share the same ultimate passion and goal, to find qualitative and quantitative ways to improve the learning experience for learners. However, the pathways we take are highly diverse, which can result in resistance when trying to bridge discourse between disciplines. We suggest that learning analytics as a transdisciplinary field would benefit from improving the visibility and understanding of the diverse philosophical positions that underpin learning analytics work.

An absence of philosophical discussions within the learning analytics community can have significant ramifications which may prevent the field from maturing. For example, confusion can result from trying to explain what is and is not in scope when different views inform those decisions. Further, the absence of philosophical understanding can slow the field down when attempting to test, experiment and scale up ideas and methods, as each of these areas can meet resistance from others that take different philosophical positions. Further, dominant philosophical positions can unwittingly silence alternative views, and blind peer review processes to the need for diverse approaches.

There is still considerable debate on ethics of learning analytics (e.g., Corrin et al., 2019, Ferguson 2019, Kitto and Knight 2019, West et al. 2020), which greater philosophical understandings would help resolve. Selwyn's (2020) provocations express the need to dig deep and assess whether the current direction of learning analytics is indeed what we want for the field. More importantly, Selwyn questions what is actually needed in society and what is missing from our background disciplines when moving into this transdisciplinary space. Finally, greater philosophical awareness would allow for the creation of momentum, as the field is reaching a critical turning point: it is needs to move beyond a few practitioners working in isolation or practicing in few classrooms to institutional or national plans to adopt and follow ethical use of learner data for pedagogical purposes. This last point is being made frequently (e.g., Ferguson 2012, Selwyn 2020 West et al. 2020), and while a recent survey showed that institutions are willing (Tsai and Gasevic 2017), when attempting to put in place these methods, we often fail (Ferguson 2012, Buerck 2014, Munguia et al. 2020).

2 ORGANISATIONAL DETAILS

WORKSHOP TYPE: Interactive Workshop – generative participatory style.

WORKSHOP SIZE: Optimum group of 12-15 Participants with 5-10 lightening talk abstract submissions.

DURATION: half a day.

TYPE OF PARTICIPATION: Presentation of short lightning talk, small group discussion, larger group discussion, robust and respectful critique and debate of salient ideas.

PROPOSED SCHEDULE: Table 1 shows the expected schedule of (1) Lightning talks, (2) Small group dialogue, and (3) Full group discussion.

Table 1: Workshop format summary.

Activity	Description	Time
Lightning talks	Contributors present key ideas lightning talk style (4-5 mins each). Other participants write short reactions. Talks and reactions are grouped to form the basis for small group dialogue.	1 hour
Small group dialogue	Small group break-out around key philosophical ideas from lightning talks - 'birds of a feather' style. Groups engage in dialogue around: (a) the potential role of the idea in LA, (b) the value to ALL LA stakeholders, (c) how the idea might advance or secure the field moving forward.	1 – 1.5 hours
Full group discussion	Reporting back of ideas from all groups. Discussion on synthesizing themes and ideas. Discussion on future possibilities – e.g. further discussion, paper writing, etc.	1.5 – 2 hours

2.1 Lightning Talks – abstract submission

Participants are invited to submit a 500 word abstract outlining a topic which they would like discussed at the workshop.

Abstracts will be reviewed by the workshop organizers in terms of: (1) relevance and significance to POLA theme, (2) potential for constructive discussion and debate, and (3) potential for ongoing influence within the LA community.

Participants whose abstracts are selected will be invited to present a 4-5 minute lightning talk during the workshop.

3 OBJECTIVES

The workshop aims to facilitate a sharing and collaborative workshop for two groups of people: (1) those that are committed to philosophical conversation with learning analytics, and which to engage with specific ideas; and (2) those with a more general interest in the topic, who would like to participate in the conversation centred on ideas proposed by others.

The workshop is designed to meet the following objectives:

(a) Develop an ongoing conversation on philosophical ideas significant to learning analytics
(b) Provide a forum of friendly critique for existing ideas
(c) Present the discussion of ideas in a form that can be disseminated to the wider community.

4 COMMUNICATION

The conceptual nature of the topic of this workshop has been a barrier for submitting publications to LA venues with a strong application requirement. While it may be feasible to take maturing ideas and develop them towards publication in other venues that accept conceptual contributions, at this point primary method of dissemination of workshop content will be via:

(a) Workshop website:
http://nlytx.io/pola/

(b) Email group:
https://groups.google.com/g/pola-chat
pola-chat@googlegroups.com

Other possibilities for workshop outputs will be discussed during the workshop itself and agree amongst participants.

REFERENCES

Buerck, J. P. (2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. Journal of Learning Analytics, 1(1), 129-139. https://doi.org/10.18608/jla.2014.11.7

Corrin, L., Kennedy, G., French, S., Buckingham Shum S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper.https://melbourne-cshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4, 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R. (2019). Ethical Challenges for Learning Analytics. *Journal of Learning Analytics*, *6*(3), 25–30. https://doi.org/10.18608/jla.2019.63.5

Kitto, K. and Knight, S. 2019. Practical ethics for building learning analytics. Br J Educ Technol, 50: 2855-2870. doi:10.1111/bjet.12868

Munguia, P., Brennan, A., Taylor, S., & Lee, D. (2020). A learning analytics journey: Bridging the gap between technology services and academic need. The Internet and Higher Education. 46: 100744.

Selwyn, Neil. "Re-Imagining 'Learning Analytics' … a Case for Starting Again?" *The Internet and Higher Education* 46 (July 2020): 100745. https://doi.org/10.1016/j.iheduc.2020.100745.

Tsai, Y. S., & Gašević, D. (2017). Learning analytics in higher education—challenges and policies: A review of eight learning analytics policies. Proceedings of the 7th International Learning Analytics & Knowledge Conference (pp. 233–242).

West, D., Luzeckyj, A., Toohey, D., Vanderlelie, J., & Searle, B. 2020. Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics. Australasian Journal of Educational Technology, 36(2), 60-70. https://doi.org/10.14742/ajet.4653

Situating affect in learning analytics: Addressing educational challenges

Elizabeth B. Cloude
University of Pennsylvania
ecloude@upenn.edu

Ryan S. Baker
University of Pennsylvania
ryanshaunbaker@gmail.com

Caitlin Mills
University of Minnesota
cmills@umn.edu

Vitomir Kovanovic
University of South Australia
Vitomir.Kovanovic@unisa.edu.au

Dragan Gasevic
Monash University
dragan.gasevic@monash.edu

ABSTRACT: This half-day workshop will focus on situating affect in learning analytics. Interdisciplinary researchers will present state-of-the-art research on techniques for measuring and modeling affect within the context of learning and education, emphasizing key conceptual, theoretical, methodological, and analytical challenges and opportunities for learning analytics (LA). Focus will be placed on interdisciplinary approaches that use contemporary theories of affect and learning sciences and discussing operational links with methods and analyses to other learning constructs. During the workshop, we will discuss implications of situating affect in learning analytics and highlight implications for building more inclusive, equitable, and quality educational experiences for diverse learning that contribute to trustworthy LA for a range of stakeholders including teachers and researchers.

Keywords: affect, learning processes, methods, analytics, education

1 WORKSHOP BACKGROUND

The objective of learning analytics is to increase the accessibility, inclusivity, and quality of education for all learners. However, current adaptive learning systems available at scale often adapt only to cognitive aspects of learning (Baker, 2016). Affect is at the core of learning experiences, and is described as a feeling, emotion, or mood. Represented by cognitive structures in the mind, affect transmits information about the world and compels us to act and make decisions. Several studies find that a learner's inability (or ability) to regulate affective states (e.g., confusion) greatly impacts their capacity to engage in learning processes and optimal performance with emerging technologies (e.g., game-based learning environments, MOOCs, intelligent tutoring systems). Yet only a handful of papers over the last few years have focused on the role of affect within the learning analytics community. Possible explanations for these gaps include the theoretical, methodological, and analytical challenges that arise when attempting to capture learners' affective processes and experiences with emerging technologies. This remote half-day workshop thus brings together interdisciplinary research groups to catalyze a more affect-sensitive education, situating affect in learning analytics, with a focus on assessment, theoretical implications, and other researcher-initiated topics of interest.

One of the most challenging aspects of studying affect is its assessment. Traditional methods define affect *before* or *after* a learning activity using self-report data. But these data are problematic because not all learners have the self-awareness to report their affective states (Pekrun & Marsh, 2022; Scherer & Moors, 2019). In contrast, state-of-the-art approaches for studying affect leverage rich streams of data generated during a learner's (or team of learners') interactions with an emerging technology while engaging in learning activities (Zamecnik, Kovanović, et al., 2022). Some studies leverage in-situ multichannel data using sensors like facial recognition software (Cloude et al., in press), eye tracking (Mills et al., 2021), and learner-system interactions (Hutt et al., 2019). Many of these methods rely upon ground truth from observational techniques (Baker et al., 2020) or video coding (Aslan et al., 2018). Yet, most of this work involves a single learner at a computer and fails to capture learning beyond the computer.

Another challenge for the field is in finding ways to pinpoint the explicit causal mechanisms between affect and learning processes. The classic approach for studying affect has been to collect data on affective labels of the end result or outcome of an affective state such as confusion, frustration, or engagement. But this approach misses if, when, and what precedes and precludes the emergence of affect, requiring researchers to make inferences about, not only the learner's label of an affective state, but also the root cause of *why* affect may have (or have not) emerged and its relation to learning processes (Scherer & Moors, 2019). Although many studies have made great strides in examining affective sequences or dynamics in relation to learning outcomes, it is correlational rather than causal. A possible contributing factor to this research challenge is that most education/learning theories discuss affect and its role in learning (e.g., information processing theory of self-regulated learning by Winne & Hadwin, 1998) but do not fully operationalize the links between affect and complex learning constructs. In our workshop, we will discuss theoretical and operational links to indices of affect in time series data to what precedes or precludes other learning constructs to gain a more holistic representation of affect, allowing for at least some degree of prediction within the context of learning.

Situating affect in learning analytics has the potential to supplement existing diagnostic data to more accurately pinpoint the root cause of a learner's struggle to engage in learning, and to understand the factors leading to student success where it occurs. Improving diagnostic abilities through the inclusion of affect could inform personalized instructional support, thus reducing barriers to learning and increasing the inclusivity and quality of education for all learners. The aim of our workshop will therefore be to bring together interdisciplinary groups of researchers interested in measuring and modeling affect by merging contemporary theories of affect and learning sciences, with a goal of galvanizing work in this area for learning analytics more broadly and for addressing the conference theme: **Toward trustworthy learning analytics.** In our workshop, pressing scientific methods and analytical challenges and opportunities will be highlighted in connecting contemporary theories of affect with learning sciences that require interdisciplinary teams to address that unlock the full potential of leveraging links between affect and learning analytics to augment education for all (Gasevic et al., 2022).

2 ORGANIZATIONAL DETAIL OF THE WORKSHOP

Type of event: Half-day remote workshop

Type of participation: mini-conference style including paper presentations, a keynote speaker, and panel discussion with leading experts. We will invite interdisciplinary research teams to submit 1) work-in-progress or 2) completed papers related to affect in learning and education and its role in LA and related themes with emerging technologies. The following highlights a list of possible themes:

1. Address the challenges associated with studying affect in the field of LA with emerging technologies, including a) State-of-the-art theoretical, methodological, analytical techniques for affect detection and b) Large- and small-scale approaches.

2. How can situating affect in LA contribute to increasing the accessibility, inclusivity, and quality of education for all learners, including a) Personalized and adaptive learning with emerging technologies, b) Promoting diversity in LA, and c) Trustworthy LA.

All papers will be peer-reviewed, and accepted papers will be published in the workshop companion proceedings (Scopus indexed). The workshop will have open participation, with everyone interested being able to register. We expect approximately 7 paper presentations (15 minute) and up to 40 participants that we will recruit through website and announcements to key academic and professional communities and networks (i.e., Society for Learning Analytics Research [SoLaR], Educational Data Mining Society [EDM], Affective Computing and Intelligent Interaction Society [ACII], Society for Artificial Intelligence in Education [AIED], Human Factors in Computing Systems Society [CHI], International Society of the Learning Sciences [ISLS], Cognitive Science, American Educational Research Association [AERA], and European Association for Research on Learning and Instruction [EARLI]), and special invitations to prominent researchers.

2.1 Workshop Format and Planned Activities

- 09:00am - 09:10am Workshop Opening
- 09:10am - 10:00am Keynote speaker, Dr. Jonathan Rowe, North Carolina State University
- 10:00am - 11:00am Paper Presentations + Discussion
- 11:00am - 11:15am Coffee Break
- 11:15am - 12:00pm Paper Presentations + Discussion
- 12:00pm - 12:20pm Panel discussion (keynoter + presenters)
- 12:20pm – 12:30pm Workshop closing

2.2 Structure and Contents of the Workshop Website

A website has been published to advertise our workshop that includes key information: https://affectla.wixsite.com/affectla2023

2.3 Workshop Objectives and Intended Outcomes

1. Unite interdisciplinary researchers to discuss key topics related to measuring and modeling affect using LA techniques with both sensor and sensor-free data.
2. Share advanced data mining, statistical, and/or machine-learning methods for affect detection and research and discuss implications for education and building trustworthy LA.
3. Lead a special issue at an interdisciplinary journal (e.g., Journal of Learning Analytics, IEEE Transactions on Affective Computing, IEEE Transactions on Learning Technologies) and summarize workshop findings and publish on workshop website.
4. Build new partnerships with industry and government institutions and private organizations. Establish a special interest working group to build bridges in the LA community with overall goals to increase knowledge and future methods and applications.

REFERENCES

Aslan, S., Okur, E., Alyuz, N., Esme, A.A., Baker, R.S. (2018) Human Expert Labeling Process: Valence-Arousal Labeling for Students' Affective States. Proceedings of the 8th International Conference in Methodologies and Intelligent Systems for Technology Enhanced Learning.

Baker, R.S. (2016) Stupid tutoring systems, intelligent humans. International Journal of Artificial Intelligence and Education, 26 (2), 600-614.

Baker, R.S., Ocumpaugh, J.L., Andres, J.M.A.L. (2020) BROMP quantitative field observations: A review. In R. Feldman (Ed.), Learning Science: Theory, Research, and Practice, pp. 127-156. McGraw-Hill.

Cloude, E. B., Dever, D. A., Hahs-Vaughn, D., Emerson, A., Azevedo, R., & Lester, J. (in press). Affective dynamics and cognition during game-based learning. IEEE Transactions on Affective Computing.

D'Mello, S. (2017). Emotional learning analytics. Handbook of Learning Analytics, 115.

Forner-Phillips, N. A., Mills, C., & Ross, R. S. (2020). Tendency to ruminate and anxiety are associated with altered alpha and beta oscillatory power dynamics during memory for contextual details. Cognitive, Affective, & Behavioral Neuroscience, 20(4), 698-716.

Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards Strengthening Links between Learning Analytics and Assessment: Challenges and Potentials of a Promising New Bond. Computers in Human Behavior, 107304.

Hutt, S., Grafsgaard, J. F., & D'Mello, S. K. (2019). Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-14).

Mills, C., Gregg, J., Bixler, R., & D'Mello, S. K. (2021). Eye-mind reader: An intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. Human–Computer Interaction, 36(4), 306-332.

Pekrun, R., & Marsh, H. W. (2022). Research on situated motivation and emotion: Progress and open problems. Learning and Instruction, 81, 101664.

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. Annual Review of Psychology, 70(1), 719-745.

Zamecnik, A., Kovanović, V., Grossmann, G., Joksimović, S., Jolliffe, G., Gibson, D., & Pardo, A. (2022). Team interactions with learning analytics dashboards. Computers & Education, 185, 104514.

Human-Centred Learning Analytics (HCLA): towards trustworthy learning analytics

This Section is Left Blank for Review

ABSTRACT: The term human-centred learning analytics (HCLA) is an emerging subcommunity of learning analytics (LA) researchers and practitioners interested in creating reliable and trustworthy LA systems that amplify and augment the abilities of educational stakeholders and which are aligned to intentions, revealed preferences, ideal preferences, interests and values. This is the fourth edition of this HCLA workshop which seeks to build on the momentum from recent years within the LA and TEL communities around the contributions that Human-Centred Design and Human-Centred Artificial Intelligence theory and practice should make to LA system conception, design, implementation and evaluation.

Keywords: human-centred design, participation, co-design, human-centred AI,

1 INTRODUCTION

Although there is a growing interest in designing LA systems with students and teachers (Sarmiento and Wise, 2022), several questions still remain regarding how the LA community can appropriate design approaches from other communities such as human-centred design (HCD) (Giacomin, 2014) and human-centred artificial intelligence (HCAI) (Shneiderman, 2022); and identify best practices that can be more suitable for educational developments. This workshop intends to address some of these questions. Some LA researchers have started to involve various educational stakeholders in the design process of LA innovations. Whilst most advancements have included teachers in the design process (Echeverria, 2020; Holstein, 2019; Tsai et al., 2022) others have also advocated for the value of involving students in this process (Chatti et al., 2021; Dollinger, 2019; Prieto-Alvarez, 2020; Sarmiento et al., 2020). While the value of involving these stakeholders as “participants” or “collaborators” in the design process is increasingly becoming a point of debate (Buckingham Shum, et al. 2019; Sarmiento and Wise, 2022), little work has been done in proposing the steps that other researchers or designers can use as a guidance to structure design sessions to understand critical aspects of the envisaged use of LA tools in order to make them fair and include proper safeguards against bias.

The aim of the workshop is to consolidate the subcommunity of LA researchers and practitioners interested in the human factors related to the effective design of LA innovations. In doing so, we plan to address questions such as: What is the state of the art in HCLA, and what have we learned from these experiences? Within the context of our field, how do we appropriate concepts such as "participatory", "co-design" and "human-centeredness", which point at strong bodies of literature and communities beyond LA? How can we design LA systems for and/or with educational stakeholders? Finally, grounding on broad HCAI principles (Shneiderman, 2022): how can we create LA systems that are reliable, safe and trustworthy? and how do we achieve balance between human agency and AI agency in learning analytics? Outcomes of this workshop include: (1) the formation of a network of LA researchers and practitioners interested in HCLA; and (2) the publication of short proceedings of emerging HCLA works.

2 BACKGROUND

(Mis)understandings of real-world users, stakeholders, contexts, and routines can make or break LA tools and systems. However, the extent to which existing human-centred design methods, processes, and tools are suited to address such human and societal factors in the context of LA is a topic that remains under-explored by our community. In response, the term human-centred learning analytics (HCLA) was recently coined (Buckingham Shum et al., 2019) to refer to the subcommunity of LA researchers and practitioners interested in utilising the body of knowledge and practice from design communities, such as participatory design and co-design, into data-intensive educational contexts. Holstein et al. (2017) were the first in adapting various co-design techniques to identify teachers' data needs and build prototypes of awareness tools with them. In fact, teachers have been the most commonly involved stakeholders in LA co-design studies. For example, Dollinger et al. (2019) discussed implications for the use of participatory semi-structured interviews with teachers in long-term LA projects. Wise and Jung (2019) combined LA interface walkthroughs and transcript analysis to make design decisions for a dashboard intended to be used by teachers. Holstein et al. (2019) featured a number of co-design techniques, namely card sorting, directed storytelling, semi-structured interviews, prototyping and behavioural mapping, to co-design a classroom analytics innovation with teachers. Whilst some examples of LA design processes have focused on engaging with students, these are starting to emerge (Chatti et al., 2020; Chen & Zhu, 2019; de Quincey et al., 2019; Prieto-Alvarez et al., 2018, Tsai et al., 2022; Sarmiento et al., 2020).

2.1 Evidence of interest

This workshop seeks to build on the momentum from recent years within the LAK and 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 and LAK22. Some of the co-organisers of this workshop are also involved in the publication of a Special Section in the British Journal of Educational Technologies, so the workshop can serve as a platform for some attendees to also pitch their works submitted to the journal.

3 ORGANISATIONAL DETAILS

3.1 Workshop format, participation, and pre-workshop task

The workshop is envisioned to be a half-day, hybrid workshop. Between 12 and 24 participants, with a shared interest in human-centred learning analytics, are expected to be part of this workshop. This workshop welcomes everyone with an interest in the field, from beginners to experts. We will have, for the first time, a call for papers. This is because the time is ripe to welcome more elaborated contributions to this area. The 2-4 pages workshop papers will be peer-reviewed by members of the organisation team and authors of papers. All the participants of the workshop will gain access to the submitted and accepted papers before the workshop to be discussed during the workshop.

Participants will also be asked to fill a survey which will capture previous experiences in HCLA and current understandings of design aspects that will be relevant for the discussions during the workshop.

In particular, participants will be asked to share their experience with human centred design or human centred AI; and current and future plans to adopt human-centred design methods in LA projects.

3.2 Workshop activities

The workshop is planned to take place during the pre-conference activities of the main conference and is planned for a half-day format of up to 4 hours (March 13 or 14, 2021). The workshop is divided into four parts:

1. **Overview of HCLA**. In the first part of the workshop, and based on the survey results, we will present a number of processes, frameworks and examples for engaging in participatory/co-design processes with students, faculty or administrators, emphasising both opportunities and challenges.

2. **Modified pecha-kucha poster presentation**. The second part will be for authors of papers to provide a brief overview of their works as flash presentations. They will be able to prepare 6 slides to be presented in 20 seconds each so each will provide a brief 2-minutes presentation.

3. **Sharing and guided critique**. The third part will be a discussion based on the experience co-designing the human-centred papers. A number of discussants from other communities (e.g. human-computer interaction, interaction design, participatory design and information visualisation), and some that critique human-centred design methods, will be invited to the workshop to give their critical points of view on the ideas posed in the design plans. We expect that this will lead to a discussion of the pros and cons of human-centred design techniques, what needs to be adapted to fit LA purposes and the provision of feedback to the presenters.

4. **Discussion on next steps**. All participants will be invited to contribute with ideas to set a potential HCLA research agenda.

3.3 Dissemination strategy

A workshop website will be made available upon acceptance of this workshop. A call for participation will be generated and published via the website, and through the twitter accounts and mailing lists the workshop organisers have access to. The website will also include an overview of the aims of the workshop, information about the workshop organisers, contact details and reports and other outputs from the workshop. The accepted papers will be published as workshop proceedings or as a part of the LAK companion proceedings or a CEUR proceedings set.

3.4 Logistics and tools

The workshop will be hybrid. We will use a multicamera online meeting tool (i.e. MultiCam in Zoom). Slack will be leveraged to engage in asynchronous discussion and share tools, papers, and resources before, during, and after the workshop. One author will be the point person for the hybrid experience including synthesizing and sharing asynchronous discussions and ensuring virtual and in-person attendees' perspectives are valued in discussions (e.g.. monitoring chat and Slack).

REFERENCES

Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in Context: Reaching beyond Usability in Learning Analytics Dashboard Design. Journal of Learning Analytics, 6(2), 70-85.

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

Chatti, M. A., Muslim, A., Guesmi, M., Richtscheid, F., Nasimi, D., Shahin, A., & Damera, R. (2020). How to design effective learning analytics indicators? A human-centered design approach. In European conference on technology enhanced learning (pp. 303-317). Springer, Cham.

Chatti, M. A., Yücepur, V., Muslim, A., Guesmi, M., & Joarder, S. (2021). Designing Theory-Driven Analytics-Enhanced Self-Regulated Learning Applications. In Visualizations and Dashboards for Learning Analytics (pp. 47-68). Springer, Cham.

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics, 6*(2), 10–26.

Echeverria, E. (2020). *Designing Feedback for Collocated Teams using Multimodal Learning Analytics.* Doctoral Dissertation, Connected Intelligence Centre, University of Technology Sydney, Australia.

Giacomin, J. (2014). What is human centred design?. *The Design Journal*, *17*(4), 606-623.

Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 257-266).

Holstein, K. (2019). *Designing Real-time Teacher Augmentation to Combine Strengths of Human and AI Instruction.* Doctoral Dissertation, Human-Computer Interaction Institute, Carnegie Mellon University, USA.

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: co-designing wearable cognitive augmentation for K-12 teachers. In *Proceedings of the 8th international conference on learning Analytics and knowledge* (pp. 79-88).

Prieto-Alvarez, C. (2020), *Engaging Stakeholders in the Learning Analytics Design Process.* Doctoral Dissertation, Connected Intelligence Centre, University of Technology Sydney, AUS.

Prieto-Alvarez, C., Martinez-Maldonado, R., & Shum, S. B. (2018). *Mapping Learner-Data Journeys: Evolution of a Visual Co-Design Tool*. Paper presented at the ACM Australian Computer-Human Interaction Conference, OzCHI'18.

Sarmiento, J. P., Campos, F., & Wise, A. (2020). Engaging Students as Co-Designers of Learning Analytics. *Companion Proceedings 10th International Conference on Learning Analytics & Knowledge* (LAK20).

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).

Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press.

Tsai, Y. S., Singh, S., Rakovic, M., Lim, L. A., Roychoudhury, A., & Gasevic, D. (2022, March). Charting Design Needs and Strategic Approaches for Academic Analytics Systems through Co-Design. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 381-391).

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.

Causal Reasoning with Graphical Causal Diagrams

ABSTRACT: Actionable Learning Analytics requires causal claims; to take well-informed action implies we have an understanding of the causal effect of that action. Causal claims that are made in the field of Learning Analytics (LA) are most commonly made in an experimental framework, such as a randomised control trial (RCT), but RCTs are not always feasible, ethical or practical. Without an RCT we need other ways to build trustworthy LA systems from observational data. Graphical causal models can be used to make causal inferences from observational data, but they require some principled scientific reasoning. This half-day hands-on workshop will introduce you to drawing such graphical causal models, using them to think about the causal assumptions underlying your LA system, and utilising the causal structure to identify and minimise potential bias in order to make stronger scientific claims.

Keywords: Causal inference, directed acyclic graphs, causal reasoning, bias, observational research

1 BACKGROUND

One central goal of Learning Analytics is the improvement of student learning. A major pathway to improving student learning comes from *actionable insight* or *actionable intelligence* (Clow, 2013; Jørnø & Gynther, 2018), frequently derived via data-intensive methods of modeling and prediction in online environments. However, the step from prediction to actually improving learning success via data-informed learning design necessitates causal knowledge about the effects of interventions. Yet, as has been noted in the literature, robust causal knowledge remains relatively scarce in Learning Analytics (Kizilcec et al. 2016; Motz et al., 2018; Wong et al., 2019).

Crucially, data analysis alone cannot provide causal information (Bareinboim et al., 2022). The leap from statistical patterns to causal claims is a qualitative one that requires knowledge from "outside" the data (Hernán et al., 2019). Reasoning about the data-generation process, i.e. causal assumptions, is one type of this knowledge. Without it, claims of improved learning or effective LA systems, for instance, must remain in doubt (Prosperi et al., 2020). For example, a statistical comparison of two distributions becomes a vehicle for causal claims if we learn the data were generated in a randomized experiment. Contrary to common opinion, however, this does not imply that making a causal conclusion in the absence of experimentation is impossible.

Recent years have seen rapid development in how to approach causal inference, even in the absence of randomized experiments. Of the many different research lines, directed acyclic graphs (DAGs) are particularly popular, as originally developed by Judea Pearl (Pearl, 1995; 2009). DAGs are a principled approach of non-parametrically encoding causal assumption within a set of variables of interest. Because construction of DAGs follow a small set of construction rules and parametric considerations do not apply, DAGs are excellent tools for communication and co-construction between researchers and educational stakeholders (Hick et al., 2022).

Crucially, DAGs are useful in identifying central causal inference pitfalls that apply in many different research contexts: Confounder Bias, Overcontrol Bias, and Collider Bias. Weidlich et al., (under review) provided indications that these pitfalls may also be present in the published Learning Analytics literature. Crucially, careful causal reasoning with DAGs in study planning and data analysis may avoid committing these pitfalls. Weidlich et al., (under review) demonstrate how DAGS can be used in different phases of the empirical research process.

2 WHO IS THIS WORKSHOP FOR?

Although there will be options for utilising software such as R to run the models in code, this workshop is intended to be suitable for a non-technical audience that is happy to think and draw simple pictures to represent their thinking. Anyone who has been puzzled about the 'real' effect that an education intervention is having, given the complexities of learning, will likely find the thinking process presented here stimulating and the methods useful as well. As this graphical approach can be used for study planning, data analysis, and appraisal of the literature, the workshop is suitable for researchers at all stages. Further, as a non-technical tool, causal diagrams are suitable as a communication tool between researchers and, for example, educational stakeholders.

3 WORKSHOP ORGANIZATION

This will be a half-day hands-on workshop with a mixture of conceptual background, puzzles, practical activities, discussion and (very optional) statistical coding in R. First, we will provide a brief introduction to the topic of DAGs via a presentation, introducing major construction rules and principles of adjustement. Then, using an online tool or simply pen-and-paper participants, either individually or in small groups, will construct causal diagrams and reason about causal effects and sources of bias.

3.1 Part One – Theory of graphical causal models

DAGs: We will explore DAGs (Directed Acyclic Graphs) and how they can be used to represent a causal model. Looking at an elementary set of DAGs we will examine causal models to understand how causal effects are transmitted through the model, how association arises from a model, and how to uncover possible sources of bias.

Finding adjustment sets: Once we can visually inspect a DAG, we can then make a plan for minimising bias. In a statistical model, this involves finding the right mix of covariates to make causal inferences from association. We will do this both visually and with the free software DAGitty (Textor, et al. 2016).

Construction of models: Building on earlier work from Hicks et al. (2022) we will outline a process for co-constructing a graphical causal model. This will include how to get started drawing a model, and how to interrogate the model with key questions.

3.2 Part Two – Application of graphical causal models

Build your own causal model: Using pen and paper, or the web based DAG drawing software DAGitty, we will construct a causal model relevant to your work. If possible, bring a question about a causal effect that you want to try and understand along, such as how much does X affect Y?

Understanding the implications of your causal model: Once we have built some causal models we will workshop as a group the possible implications of the models. For instance, is it possible to estimate a causal effect? If so, how? If not, what new data would we need to do so?

[Optional] Coding the model: If you are a little comfortable coding in R there will be opportunities to move the model from the graphical interface of DAGitty into R and begin analysis, or simulation.

3.3 Part Three – Discussion and Reflection

Further scope for causal models in LA? These models are primarily used for causal inference, commonly within a regression model. However, they have applications in machine learning, and potentially could inform other LA tools such as dashboard design.

4 RESULTS AND GOALS

A main goal of this workshop is to bring the crucial role of causal inference to the forefront of discourse in the Learning Analytics community. Because research contexts are often diverse and applied, not always lending themselves to rigorous experimentation, researchers will leave this workshop equipped with a set of tools to reason about causal inference, no matter the research design at hand. Participants will learn how to construct causal graphs, interrogate them, and derive implications for their own research goals. As robust causal knowledge is central to building trustworthy LA systems, this workshop contributes to the conference theme and the LAK community by equipping participants with a methodology to improve their causal inferences and making stronger scientific claims.

REFERENCES

Bareinboim, E., Correa, J. D., Ibeling, D., & Icard, T. (2022). On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl* (pp. 507-556).

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: a classification of data science tasks. *Chance*, *32*(1), 42-49.

Hicks, B., Kitto, K., Payne, L., & Buckingham Shum, S. (2022, March). Thinking with causal models: A visual formalism for collaboratively crafting assumptions. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 250-259).

Jørnø, R. L., & Gynther, K. (2018). What Constitutes an" Actionable Insight" in Learning Analytics?. *Journal of Learning Analytics*, *5*(3), 198-221.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016, April). Recommending self-regulated learning strategies does not improve performance in a MOOC. In *Proceedings of the third (2016) ACM conference on learning@ scale* (pp. 101-104).

Motz, B. A., Carvalho, P. F., de Leeuw, J. R., & Goldstone, R. L. (2018). Embedding experiments: Staking causal inference in authentic educational contexts. *Journal of Learning Analytics*, *5*(2), 47-59.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669-688.

Pearl, J. (2009). *Causality*. Cambridge university press.

Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., ... & Bian, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, *2*(7), 369-375.

Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT (2016). "Robust causal inference using directed acyclic graphs: the R package 'dagitty'." *International Journal of Epidemiology*, **45**(6), 1887–1894.

Weidlich et al. (under review). Causal inference and bias in Learning Analytics: A primer on pitfalls using Directed Acyclic Graphs. *Manuscript submitted for review at Journal of Learning Analytics*

Weidlich et al. (under review). Causal reasoning with causal graphs in Educational Technology Research. *Manuscript submitted for review at Educational Technology Research and Development*

Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational theories and learning analytics: From data to knowledge. In *Utilizing learning analytics to support study success* (pp. 3-25). Springer, Cham.

LAK Theory 2023: Workshop on Theory and Learning Analytics

Author 1: Please Leave This Section Blank for Review
Institution
Email

Author 2: Please Leave This Section Blank for Review
Institution
Email

Author 3: Please Leave This Section Blank for Review
Institution
Email

ABSTRACT: This half day workshop is the 4th annual pre-conference meeting at LAK to discuss the ways in which theory informs and arises from learning analytics research. The organisers will briefly set the scene then hand over to Andy Nguyen to give the keynote presentation on a new generic Framework for Analysing Regulation in Collaborative Learning (FARCL). The second half of the workshop will be devoted to conversation. Participants are invited to nominate a current research project or new research idea that would benefit from a roundtable-style discussion with colleagues, including a theoretical framework that is of interest. In groups, participants will consider how nominated projects can demonstrate the role of theory in their design, model validation, and interpretation of findings.

Keywords: Conceptual rationale, framework, paradigm, theoretical model

1 BACKGROUND

LAK Theory will run for its fourth consecutive year in 2023. Enrolment in previous years has been healthy, even amidst the COVID-19 pandemic. The workshop is very welcoming includes return attendees who are keen to continue the conversation along with those joining for the first time.

This workshop is founded on the premise that the quality of learning analytics (LA), both research and practice, rests on the strength of its connection to theory (Gašević, Dawson, & Siemens, 2015). This is because theory creates concrete conceptual bridges between LA and more established areas of educational research, the broader social sciences, mathematical and computer sciences, and related disciplines. Through this annual workshop we hope to build an ongoing community of scholars interested in both using educational (and other) theory in learning analytics research and practice, and contributing to further development of theory through their work.

Theory provides a common language through which to communicate about research, it gives a frame of reference to understand the type of knowledge being generated, and what may be legitimately claimed (Reimann, 2016). In a typical research cycle, we suppose that theory influences the questions we ask, design of data collection, analysis approach and method, and interpretation and reporting of results (Wise & Shaffer, 2015). In this way we are arguing for a move away from the primacy of method in learning analytics, that is, away from pragmatism to theory-driven paradigms for research where theory underpins method and the two cannot be separated (Bartimote, Pardo, Reimann, 2019). This adds the possibility for explanation – for an observed pattern, for a prediction, for why an intervention or pedagogical strategy works – in research, and in practice. Use of theory also means we can better understand the nature of educational data.

Theory allows for informed practice by a range of actors that support learning in educational settings, such as teachers, student support officers, advisors, and academic managers. If the objective of learning analytics is actionable information, then theory-driven analytics enables choices and decisions that are situated in defensible frameworks (Bartimote et al., 2019). Further, it means we have a starting point for explanation when things do or don't work, and a basis for adaption of tactics and strategies shown to be effective in one context to other contexts. For analysts, data scientists, and software developers, theory can guide what usage activity to capture, the development of indicators and measures, the display of information, and the form of personalised messages and automated nudges. We need to focus on providing information about constructs that matter, and learning (and other) theories substantiated by empirical research can serve as useful starting points.

The LAK community is increasingly drawing on ideas from the learning sciences, educational psychology, sociology, and social psychology. This is demonstrated in recently published learning analytics work referring to theories such as social cognitive theory and self-efficacy beliefs, various self-regulated learning models, measurement theory, collaborative learning theory, human-computer interaction (HCI) and activity theory, etc.

Increasingly, the nitty gritty of processes and procedures employed when we work with particular methods is informed by theory. One example is the use of theory to inform trace data selection and curation as explicated by Winne, 2020 and Fan et al. (2022). Our keynote speaker for 2023 also attends to this in his presentation of a new generic Framework for Analysing Regulation in Collaborative Learning (FARCL), where in each dimension theory, data type, analysis approach, and analytical technique and considered iteratively and coherently in the design and critique of LA studies.

We consider the time is ripe for a call across the community to gather to consider more explicitly the role of theory in learning analytics. Given the interdisciplinary nature of the learning analytics community, it's important that researchers are able to articulate their stances and begin to create some common understandings in the field. Coming together to support this work is the purpose of the LAK Theory workshops.

2 ORGANISATIONAL DETAILS

2.1 Half-Day Workshop Schedule

Table 1: Schedule.

Timing	Description	Contributors
10 minutes	Welcome and plan for today, introductions	Organiser 1
20 minutes	'Setting the scene: Why focus on theory in learning analytics' 10 minutes presentation, 10 Q&A	Organiser 2
30 minutes	'A new generic Framework for Analysing Regulation in Collaborative Learning (FARCL)' 20 minutes presentation, 10 Q&A	Andy Nguyen
40 minutes	Roundtable (Part 1). Work in progress roundtables: 10 minutes to introduce project, summarise progress to date, outline challenges to be overcome, and input that would be useful from the group, followed by 10 minutes discussion with colleagues [x2 before break]	Participants: 4 research teams per roundtable group
30 minutes	*tea/coffee*	*All*
40 minutes	Roundtable (Part 2). Continued [x2 after break]	Participants continued
30 minutes	Roundtable report back: group representatives to summarise conversation and potential impact on the work	Participants
10 minutes	Next steps plenary discussion, and close	Organiser 3

2.2 Other Details

The event will be an open workshop. All attendees will have the opportunity to give a short presentation in their roundtable group on either work in progress or idea in development, should they wish to. Abstract submissions of 300-600 words for these short presentations will be handled via the event's Google Form: *{link to be inserted following blind review}*. Please use #LAKtheory when referencing this event on social media.

This workshop can be adapted to be either blended with both online and face-to-face participants, or online only, depending on the final format of the conference.

3 OBJECTIVES/INTENDED OUTCOMES

The workshop will provide a space for both capacity building and connection, and it is hoped that the event will continue to be a catalyst for the growth of a community of practice. The outcomes of the event will be housed on the Google Site *{link to be inserted following blind review}*. This event will serve as a template for an ongoing workshop initiative on theory and learning analytics.

4 WEBSITE STRUCTURE AND CONTENT

The Google website will: 1. support pre-workshop data gathering and planning materials; 2. act as a collection point for materials, group interactions and archive for the workshop; and, 3. support ongoing dissemination and group activities. It is the aim for the workshop to be ongoing, in which case the website will be a continuing hub for year on year activities and building field memory.

REFERENCES

Bartimote, K., Pardo, A., & Reimann, P. (2019). The perspective realism brings to learning analytics in the classroom. In J. M. Lodge, J. Cooney Horvath, & L. Corrin (Eds.), *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers*. London: Routledge.

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., ... & Gašević, D. (in press). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71.

Reimann, P. (2016). Connecting learning analytics with learning research: The role of design-based research. *Learning: Research and Practice*, *2*(2), 130-142.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior*, *112*, 106457.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5-13.

InnovateDesign: Workshop on Learning Design Analytics: Balanced Planning with an Innovative, Free-to-Use Software Tool

Organizers 1: Blaženka Divjak, Darko Grabar
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 aim of this half-day face-to-face workshop, organized in cooperation of three universities, is twofold. First, it will provide a platform for sharing of experiences, research and challenges related to the link between learning analytics (LA) and learning design (LD). Second, the workshop will enable participants to engage with an innovative, free-to-use LD tool (learning-design.eu), and create advanced learning analytics on LD using the tool. Participants will be invited to work collaboratively on LD of their own courses, reflect on the LA generated by the LD tool and improve the course LD accordingly. This will also contribute to the further development of the concept and tool, based on a pre-established research protocol. Participants will take away recommendations for improvement of their own courses, as well as know-how on how to use an innovative LD tool at their own institutions. Ahead of the workshop, if interested, participants will be invited to apply for a short presentation (5-10 minutes). They will also be asked to consider their courses and particular learning outcome(s) which could be redesigned at the workshop.

Keywords: learning design concept and tool, learning analytics, curriculum analytics, assessment

1 INTRODUCTION TO LEARNING DESIGN CONCEPTS AND TOOLS

Learning Analytics (LA) has been increasingly used to support sound learning design (LD) (Rienties et al., 2017), in particular in ensuring constructive alignment between learning outcomes (LOs), teaching and learning activities and assessment (Divjak et al., 2022). LD has been defined as “the documented design and sequencing of teaching practice“ (Lockyer et al., 2013, p. 1439), describing the order of learning tasks, resources and related support. LD has been presented by Conole (2013) as a methodology helping teachers and designers in more informed decision-making related to the design of learning activities, that is “pedagogically informed” and uses resources and technologies in an effective way. In essence, LD has two dimensions - a conceptual and a technological one. The concept of LD has been related to the notions of sharing and reusing: it helps make the design process “more explicit and shareable” (Conole, 2013) and aims at “reusability” in different educational contexts (Lockyer et al., 2013). It has been argued (Conole, 2013) that more explicit and shareable design enables more effective learning environments and teachers’ interventions, while enabling learners to understand their learning paths better. In terms of linking LD with LA, five main types of analytics have been identified in previous research: temporal, comparative, tool specific, cohort dynamics and contingency (Bakharia et al., 2016).

Considering the recognized benefits of LD in supporting and enhancing teaching and learning in a digital age and the 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). These approaches presented a valuable input for the development of the BDP concept and tool, but the BDP also aimed to introduce innovation in several ways. It also introduces innovation in terms of linking course LOs with the study program LOs, providing an institutional perspective. In relation to this, research has indicated that students benefit from long-term study program level planning (Raković et al., 2022). Furthermore, the BDP tool focuses strongly on ensuring constructive alignment between LOs, types of teaching and learning activities, assessment, feedback and student workload, supporting a student-directed approach. The BDP tool 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). In particular, these analytics provide detailed analyses and visualizations of assessment, minding its alignment with the prioritization, level and weights of LOs. These analytics are provided in real-time, through a dedicated dashboard, and can be used during the design process as a valuable input directing the LD process. The tool enables collaborative work and sharing of LDs, as well as export of LDs. Finally, the tool can be used in a simple and 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 250 courses and MOOCs, by over 550 users from more than 20 countries, including within four European funded projects. Based upon the initial pilot testing (Divjak et al., 2022) further functional and design modifications have been made, and at LAK 2023 we aim to share some additional functionality in terms of LA features.

2 LEARNING OUTCOMES, WORKSHOP STRUCTURE AND WEBSITE

Based on the capacity-building at the workshop, participants will be able to (1) analyse the benefits of LA for improvement of LD, (2) effectively use a free-to-use LD tool, including an innovative approach to LD, and (3) upgrade initial LD based on available LA. The workshop, organized in cooperation of three universities, will be held face-to-face, taking half a day and consisting of the parts presented in the table below. The expected number of participants is 20, and the maximum 40.

Table 1. The proposed agenda of the workshop

Duration	Description	Responsible
10 min	INTRODUCTION	Organizer 1
	SHARING OF EXPERIENCES, RESEARCH AND CHALLENGES	Organizer 2
40 min	Presentations of participants' experiences	Organizer 2
20 min	Presentation of the BDP concept and tool	Organizer 1 & 2
30 min	BREAK	
	HANDS-ON COLLABORATION ON LEARNING DESIGN	
90 min	Collaboration on LDs in groups	Organizer 1, 2 & 3
30 min	Presentation of LDs and discussion	Organizer 2
20 min	FUTURE STEPS AND CONCLUSIONS	Organizer 1, 2 & 3

The workshop 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 recruit participants, along with the website, social networks and media will be used. 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 from participants or, alternatively, a few short presentations from the workshop organizers, focusing on the current research, practices and experiences in the use of LD. A special focus will be on how LA can support sound LD.

Therefore, participants will be invited to submit abstracts outlining short presentations (5-10 minutes) ahead of the workshop. The workshop organizers will review the applications and choose interesting and diverse examples. The presentations will be followed by a discussion of all participants, leading to open questions and challenges, providing introduction to the next phase of the workshop. Finally, the BDP concept and software tool will be presented by the workshop organizers.

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. At the workshop, participants will be divided into groups, based on their own preferences and similarity of courses/LOs they would like to work on.

The groups will be invited to access the BDP tool, open and design their courses and LOs (Figure 1). Furthermore, they will work on the detailed planning of teaching and learning activities, assessment, feedback, modes of delivery, etc. In the process, they will consult the analyses provided by the tool (Figure 2), 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 2 hours

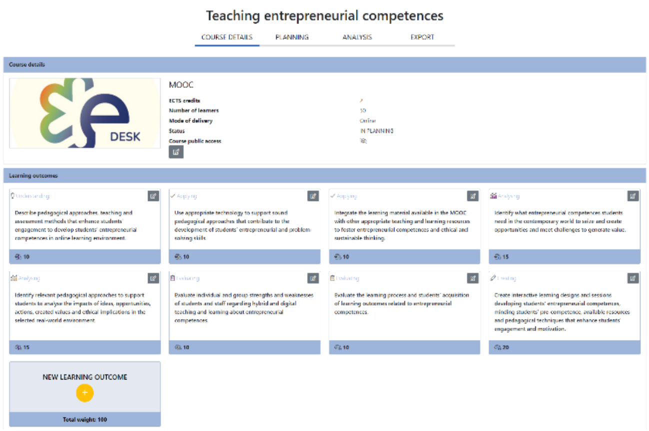


Figure 1. Example of a course with LOs planned in the BDP tool

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.

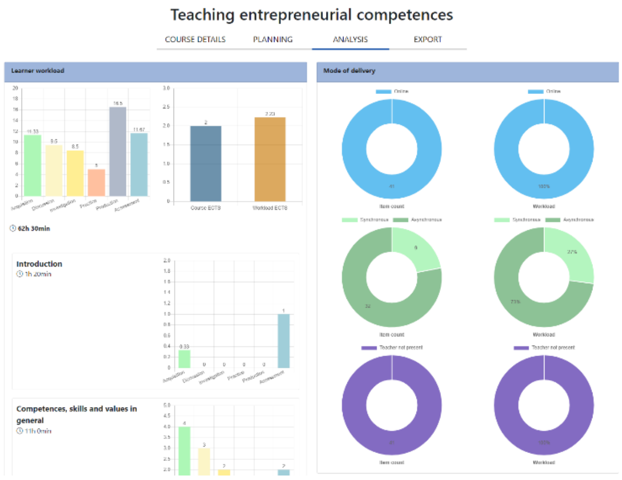


Figure 2. Example of a part of the analytics available in the BDP tool

5 FUTURE STEPS AND CONCLUSIONS

Finally, the participants will be asked to take part in the evaluation of the tool, prepared in line with the approved research protocol (ethically approved by one of the workshop organizers' universities). Conclusions of the workshop will be shared with the participants after the workshop, in the form of a workshop summary published on the workshop website. 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

Conole, G. (2013). *Designing for Learning in an Open World*. Springer New York. https://doi.org/10.1007/978-1-4419-8517-0

Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2022). Balanced Learning Design Planning: Concept and Tool. *Journal of Information and Organizational Sciences.*

Laurillard, D., Charlton, P., Craft, B., Dimakopoulos, D., Ljubojevic, D., Magoulas, G., Masterman, E., Pujadas, R., Whitley, E. A., & Whittlestone, K. (2013). A constructionist learning environment for teachers to model learning designs. *Journal of Computer Assisted Learning*, *29*(1), 15–30. https://doi.org/10.1111/j.1365-2729.2011.00458.x

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics with Learning Design. *American Behavioral Scientist*, *57*(10). https://doi.org/10.1177/0002764213479367

Raković, M., Bernacki, M. L., Greene, J. A., Plumley, R. D., Hogan, K. A., Gates, K. M., & Panter, A. T. (2022). Examining the critical role of evaluation and adaptation in self-regulated learning. *Contemporary Educational Psychology*, *68*, 102027. https://doi.org/10.1016/j.cedpsych.2021.102027

Rienties, B., Nguyen, Q., Holmes, W., & Reedy, K. (2017). A review of ten years of implementation and research in aligning learning design with learning analytics at the Open University UK. *Interaction Design and Architecture(S)*, *33*, 134–154. https://doi.org/10.55612/s-5002-033-007

Lowering the Technical Barriers to Trustworthy NLP

Organizers:
Pete Smith, Elizabeth Powers, Henry Anderson
The University of Texas at Arlington
Justin Dellinger
Texas A & M University
George Siemens
University of South Australia &
The University of Texas at Arlington
James Pennebaker
The University of Texas at Austin

ABSTRACT: Learning Analytics research increasingly involves large amounts of complex language data. This coincides with a widespread surge in interest in natural language processing (NLP), with models like BERT, GPT-3, and DALL-E making headlines. Such tools excel at raw predictive power, but often fall short on other important measures such as ease of use, explainability, and strong theoretical foundations.

Fortunately, tools for NLP such as LIWC, Coh-Metrix and CGA do not suffer from these drawbacks. These tools may have slightly lower accuracies than cutting-edge NLP models, but their ease of use, explainability, and theoretical foundations make them compelling options for LA researchers and practitioners.

This workshop will serve to highlight such tools, the research being done with them, and the roles they can play in advancing Learning Analytics research. This workshop also welcomes work focusing on the explainability, auditability, and trustworthiness of cutting-edge NLP models.

Building on a successful foundational NLP workshop at LAK22, this year's gathering will continue to build NLP capacity within Learning Analytics, and develop lasting networks for future scholarly exchange.

Keywords: natural language processing (NLP), no-code, low-code, linguistics, language datasets, NLP algorithms, educational contexts, large language models, AI, ChatGPT.

1 INTRODUCTION AND BACKGROUND

Learning Analytics (LA) has long made use of language data to better understand learners, learning processes, and learning environments. Tools such as LIWC, TAACO, and Coh-Metrix have long been features of language-focused LA work. Interest in language data has increased in recent years, coinciding with a surge of interest in Natural Language Processing (NLP) research, and many LA researchers and practitioners have begun to utilize cutting-edge NLP tools.

In NLP, the recent focus has been on developing large neural network models that offer high predictive accuracies on a wide range of tasks. This has led to considerable (and well-deserved) attention and hype: BERT (Devlin et al., 2019) set new state-of-the-art performance for tasks ranging from document classification to question answering, as well as document and image creation. T5 (Raffel et al., 2020) showed further leaps in performance across many tasks and domains. DALL-E 2 generates highly detailed images based on a textual description. ChatGPT-3 and other generative AI can produce text nearly indistinguishable from that written by a human.

Yet as these models continue to play a central role in NLP, the technical requirements to create and to use them continue to climb, putting them out of reach of many researchers who could benefit from them. Specialized technical knowledge in machine learning and neural networks, a fair bit of programming skill, and potentially expensive computer hardware are all requirements to build and use these models effectively. This is not a big ask for companies like Google, but these requirements are a bigger ask for LA researchers, and put these tools beyond the reach of many people in the field. As Learning Analytics increasingly begins to incorporate larger language datasets and models into its analyses, there is a need for NLP tools that are more widely accessible to researchers and practitioners without such high levels of technical expertise.

There is also a need for tools that satisfy important criteria other than pure predictive accuracy. There has been a rapidly growing interest in the topics of algorithmic transparency, auditability, and trustworthiness in recent years. Recent LAK and Educational Data Mining (EDM) conferences have hosted dedicated workshops (Holstein & Doroudi, 2019; Lynch et al., 2022) and the Handbook of Learning Analytics devoted a chapter to the subject (Uttamchandani & Quick, 2022). The current generation of large language models has traded interpretability and transparency for accuracy, and despite considerable efforts by NLP researchers (e.g. Rogers, Kovaleva & Rumshisky, 2020), the models remain black boxes. This tradeoff may not always be acceptable for LA researchers and practitioners.

Fortunately, there are many tools and approaches that maintain high usability and transparency, even though their on-paper accuracy metrics may not look as impressive when compared to cutting-edge NLP models. Linguistic Inquiry and Word Count (LIWC) can process large amounts of text on fairly modest computer hardware, and its predictions can be easily audited by reading the word lists associated with each of its output categories (Boyd, 2021: Pennebaker, 2022; Pennebaker et al, 2014). The TAACO family of tools (including TAALES, TASSC, etc.; Crossley et al., 2019) similarly, provide straightforward, well-documented metrics about texts and group interactions and are accessible through a simple graphical interface. Other tools provide analyses of online discourse patterns among groups of interlocutors (Dowell et al., 2019).

This workshop will focus on language-centric LA work and seeks to highlight the role of approaches that range from simpler word-count methods up through modern, state-of-the-art AI methods.

2 WORKSHOP OVERVIEW

2.1 Type of Event

This full-day workshop was held as an in-person event.

2.2 Workshop Schedule

The Psychology of Verbal Behavior
Dr. Ryan Boyd, TikTok & Obelus Institute

LIWC Introduction
Dr. James Pennebaker, The University of Texas at Austin

Dictionary-based Natural Language Processing: Top-down and Bottom-up Approaches to Understanding Psychological Dynamics
Dr. David Markowitz, University of Oregon

Teaching and Learning in the Age of AI: What are the Opportunities?
Dr. Vitomir Kovanovic, University of South Australia

Leveraging Natural Language Processing to Detect Gaming the System in Open-ended Questions in a Math Digital Learning Game
Jiayi Zhang, University of Pennsylvania

Applied Techniques in Natural Language Processing
Stefan Slater, University of Pennsylvania

NLP and the Reduction of Complexity
Henry Anderson, Elizabeth Powers, and Haein Won, The University of Texas at Arlington

3 WORKSHOP OBJECTIVES

3.1 Objectives

The broader goals of this workshop were to: 1) continue to build capacity for emerging research in NLP around trustworthy, reliable, valid approaches and tools; and 2) forge connections between future and existing NLP researchers within the LA community through the use of tools with ease-of-use.

3.2 Hosting, Sharing, and Communication

LALN will host and openly share materials via its website and resource hub. Hosting will include standard presentation artifacts (slide sets, recordings, etc.) as well as the event report at this location: https://learninganalytics.net/laln/

REFERENCES

Boyd, R. L., & Schwartz, H. A. (2021). Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field. *Journal of Language and Social Psychology*, *40*(1), 21–41. https://doi.org/10.1177/0261927X20967028

Crossley, S.A., Kyle, K. & McNamara, D.S. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behav Res* **48**, 1227– 1237 (2016). https://doi.org/10.3758/s13428-015-0651-7

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv:1810.04805 [Cs]*. http://arxiv.org/abs/1810.04805

Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, *51*(3), 1007–1041. https://doi.org/10.3758/s13428-018-1102-z

Holstein, K., & Doroudi, S. (2019). Fairness and equity in learning analytics systems (FairLAK): *Companion proceedings of the ninth international learning analytics & knowledge conference (LAK 2019)* (pp. 500-527).

Gibson, A., Kitto, K., & Bruza, P. (2016). Towards the Discovery of Learner Metacognition from ?? Text missing?

Lynch, C., Marras, M., Pechenizkiy, M., Rafferty, A., Ritter, S., Swamy, V., & Yu, R. (2022). FATED 2022: Fairness, accountability, and transparency in educational data. In *Proceedings of the 15th international conference on educational data mining* (No. CONF, pp. 848-849).

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLoS ONE*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. (2022). *LIWC 22*. LAK 22.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, *21*(140), 1-67.

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in bertology: What we know about how bert works. *Transactions of the Association for Computational Linguistics*, *8*, 842-866.

Routley, N. (2023, February 1). Infographic: Genrative AI explained by AI. *Visual Capitalist*. https://www.visualcapitalist.com/generative-ai-explained-by-ai/

Uttamchandani, S., & Quick, J. (2022). An Introduction to fairness, absence of bias, and equity in learning analytics. *Handbook of Learning Analytics*.

Analytics Revealing Systemic Inequities At Different Scales

Stefano Fiorini, Montserrat Valdivia Medinaceli
Indiana University
sfiorini@iu.edu, mbvaldiv@indiana.edu

Rebecca L. Matz, Meaghan Pearson, Nita Tarchinski and Xin Tong
University of Michigan
rlmatz@umich.edu, meaghanp@umich.edu, nitaked@umich.edu, xstong@umich.edu

Sarah Castle
Michigan State University
gadysara@msu.edu

Angel Sylvester
University of Minnesota
sylve057@umn.edu

Kameryn Denaro
University of California, Irvine
kdenaro@uci.edu

Victoria Farrar
University of California, Davis
vsfarrar@ucdavis.edu

ABSTRACT: **Mini-Symposium with Tutorial.** Analytics applied to the student record, although limited for describing individual student learning and experiences, are effective in revealing institutional systemic inequities at different levels and in different spaces. In this session we will discuss and share analytical tools developed by students, faculty, and staff engaged in the Measurement Working Group of the Sloan Equity and Inclusion in STEM Introductory Courses (SEISMIC) Collaboration. We have been engaged in establishing metrics for measuring equity and inclusion in foundational STEM courses, conducting the measurements, and identifying actionable data to promote change. Along this path of discovery, the group has established theoretically informed guidelines for framing, analyzing, and interpreting quantitative data in support of equity goals in STEM. Building on this understanding, we have focused analyses on institutional structural inequities at different scales, revealing inequities in the students' experience in their first STEM course, at the curriculum level, and at the course level. The goal of this Mini-Symposium and Tutorial is to provide participants with these theoretical and practical tools first by presenting the methods and results of the studies and then by providing hands-on training, ultimately empowering participants to apply analyses like these to their own institutional data.

Keywords: Diversity, equity, inclusion, measurement, student success, systemic inequities.

1 ORGANIZATIONAL DETAILS

We propose to host an Interactive Mini-Symposium followed by a Tutorial. Each of these interconnected parts will be approximately two hours for a total of a half-day session.

1.1 The Mini-Symposium will consist of eight presentations:

1.1.1 An introduction to the session:

An overview of the collaboration, its history, trajectory, and positionality.

1.1.2 Integrating critical approaches into quantitative science, technology, engineering, and mathematics (STEM) equity work:

A discussion of how researchers and educational practitioners should more critically approach STEM equity analyses and why modifying our approaches matters for STEM equity goals (Pearson et al., 2022).

1.1.3 Exposing inequity: A multi-institutional analysis of systematic advantages in introductory STEM courses:

This study introduces the systemic advantage index (developed using race/ethnicity, gender, income status, and first-generation status) and shows its effectiveness in revealing grade disparities in STEM as a manifestation of systemic inequities across seven large, public, research-intensive US universities over ten years (Castle et al., 2021).

1.1.4 Cross-institutional comparison of curricular pathways to reveal minoritizing structures:

Here, we apply process analytics (see e.g. Salazar-Fernandez et al., 2021) and social network analysis (see e.g. Dawson and Hubball, 2014) to students' enrollments as analytical tools to reveal minoritizing structures. In the process we address two questions: 1) Is the analysis of students' progression through a curriculum effective in revealing common impacts of degree structures across universities? And 2) Does student progression through the curriculum structure result in minoritization? The analysis of the curriculum as a tool for disciplinary acculturation is effective in revealing aspects of students' transitions through education systems not captured by commonly applied course or retention analysis.

1.1.5 Exploring the role of class and college composition on performance in students' first STEM classes:

The contributors to retention and performance in STEM-based courses are multifaceted and influenced by the presence of external recognition, classroom climate, and general confidence that students might possess as a result of their general expectations within that class. Even with an increased presence of diversity, the lack of explicit support and opportunities to nurture growth may only impede the pursuit of increased representation in the classroom.

1.1.6 A multi-institutional analysis of opportunity gaps amongst biology students across STEM courses revealing differential outcomes for undergraduate students across disciplines:

In this work we reveal the presence of opportunity gaps emerging from an educational context that hampers the opportunity for students to express their full learning potential. We show this by applying three methodologies (Denaro et al., 2022): 1: We calculate the proportion of As and Bs awarded compared to that for Cs, Ds, and Fs between two subgroups and calculate the difference between the percentage of As and Bs awarded to each group. 2: We calculate the difference between the average grade received by the two groups on a 4.0 scale. 3: We use quantile regression to calculate normalized regression rankscores (NRR) and then calculate the difference in NRR. Course performance should be independent of demographic characteristics; that is, the difference in academic performance between two groups of students based on demographics should be centered around zero.

1.1.7 Equity gaps associated with identities that confer systemic advantages persist into upper-division biology courses:

We assessed whether equity gaps associated with demographic factors that confer systemic advantages are present in upper-division, biomedical prerequisite courses across five public, research-intensive institutions in the US. We also examined whether grade disparities are present for transfer students and how transfer status interacts with other demographic identities. Given that such data is hierarchical (i.e., university, course, section) and the response variable (grade) is highly skewed, we used robust multilevel modeling, and compared model estimates across institutions.

1.1.8 Closing remarks and a facilitated short question and answer period:

A discussion of how researchers and educational practitioners should more critically approach STEM equity analyses and why modifying our approaches matters for STEM equity goals (Pearson et al., 2022).

1.2 The Tutorial:

The Tutorial will provide the participants with an overview of the data structures used for these analyses as well as a synthetic dataset for the application of the analytical methodologies presented during the Mini-Symposium. Following this brief overview of the data, the rest of the tutorials (approximately 25 minutes each) will focus on the methods applied in the studies presented in the first half of the session. Following the numbering above, the order will be: 1.1.3, 1.1.4, 1.1.5 and 1.1.7 combined, and 1.1.6.

2 MINI-SYMPOSIUM AND TUTORIAL OBJECTIVES AND INTENDED OUTCOMES

With this Mini-Symposium and Tutorial, we intend to provide a theoretical and practical approach to equity analysis that builds on data that is regularly collected and readily and widely available to institutions of higher education. The diffusion of these practices, by revealing the presence of systemic inequities using common data structures and analysis can provide benchmarking opportunities and shared knowledge of the effectiveness of practices adopted to remediate inequities. Importantly, this work can and should be done at scale.

Leading to the event we will develop a section of the SEISMIC Collaboration's website containing abstracts of the papers and links to more extended resources. A Google Drive space, accessible to the participants, will contain in-depth descriptions of the data structures used in the analysis and the methodologies applied. Code for the analyses used during the tutorials will be available via a public GitHub site.

Communication with the participants, before and after the event will be supported by a SEISMIC mailing list as well as website updates.

Participants will be encouraged to share results of the analyses based on data from their home institutions to contribute to our understanding of the pervasiveness of inequities in STEM education. Participants will also be encouraged to share case studies of impactful interventions that mitigated or removed systemic inequalities in the education system.

3 PLANNED MECHANISM(S) FOR COMMUNICATING INFORMATION AND RESOURCES

The event will be promoted through the SEISMIC Collaboration as well as other professional and academic networks such as the SoLAR Newsletter, Unizin, discipline-based education research (DBER) professional societies, etc. Links to an event website and a GitHub repository will be provided in the announcement.

The website and GitHub repository will be one source of information. A Google Drive space, restricted to the event's participants, will make available descriptions of data and methodologies applied in the papers that will be presented at the Mini-Symposium. The Drive space and mailing list will provide support for communication and information sharing after the event.

REFERENCES

Castle, S. D., Byrd, W. C. , Koester B. P., Bonem, E., Caporale N., Cwik S., Denaro, K., Fiorini, S., Matz, R. L., Mead, C., Whitcomb, K. M., Singh, C., Levesque-Bristol, C. & McKay, T. (2021). Equity in the STEM Landscape: A Multi-institutional Approach to Mapping Systemic Advantages Within STEM Courses. Paper presented at the 2021 annual meeting of the American Educational Research Association. https://doi.org/10.3102/1689325

Dawson, S., & Hubball, H. (2014). Curriculum analytics: Application of social network analysis for improving strategic curriculum decision-making in a research-intensive university. Teaching and Learning Inquiry, 2(2), 59-74. https://doi.org/10.2979/teachlearninqu.2.2.59

Denaro, K., Dennin, K., Dennin, M., & Sato, B. (2022). Identifying systemic inequity in higher education and opportunities for improvement. Plos one, 17(4), e0264059. https://doi.org/10.1371/journal.pone.0264059

Pearson, M. I., Castle, S. D., Matz, R. L., Koester, B. P., & Byrd, W. C. (2022). Integrating critical approaches into quantitative STEM equity work. CBE—Life Sciences Education, 21(1), es1. https://doi.org/10.1187/cbe.21-06-0158

Salazar-Fernandez, J. P., Munoz-Gama, J., Maldonado-Mahauad, J., Bustamante, D., & Sepúlveda, M. (2021). Backpack Process Model (BPPM): A Process Mining Approach for Curricular Analytics. Applied Sciences, 11(9), 4265. https://doi.org/10.3390/app11094265

Getting started with Learning Analytics for Digital Games

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

ABSTRACT: Game data provide a rich source of information about learner interaction that can be used to understand learners, teaching and design. During this full-day workshop we will onboard new researchers (with or without programming experience) into an existing collection of datasets from a variety of games, analysis infrastructure, and code samples. The results will be a new community of researchers that have the access, tools and vision to participate in game data learning analytics in the near future.

Keywords: Game-based learning, educational data mining, open science, secondary analysis

1 Background

Games are extremely popular learning tools, therefore improvements to how they are theorized, designed and used will have widespread effects. A 2020 national survey of teachers conducted by Project Tomorrow reported that 50% of teachers in the United States use games weekly, up 7% from a corresponding 2019 study and 20% from a 2012 study (Evans, 2020). A similar study (Takeuchi & Vaala, 2014) reports that 55% of teachers claim to use games at least weekly in their instruction. While the surveys above hold a loose definition of games, here we use "game" to describe a family of interactive digital media that includes video games, virtual worlds, multi-user virtual environments (MUVEs), virtual laboratories, educational simulations, interactive narratives, and virtual reality experiences.

In addition to being popular, games are unique in their ability to facilitate and instrument complex learning behavior. Each game defines a unique set of player verbs, or actions that take place within a specific context and toward a goal. Players' "chat" with other players or non-player characters can be captured in context of their other activities. Combined with frameworks such as evidence-centered design (Mislevy & Haertel, 2007) or quantitative ethnography (Shaffer, 2017) these design elements can be used to instrument human thinking *in the context of problem solving*. Each action within the game, or interaction with other learners, provides evidence for understanding the individual's thinking, skills, attitudes, and decision-making process. Signals such as gaze, gesture, and facial expression can be captured and synchronized with their actions. Together, these data provide a fine-grained, millisecond scale of measurement that can be scaled to large numbers of learners.

1.1 Challenge

Unfortunately, the application of learning analytics to the data that games provide has significant challenges and barriers. While the data captured is rich with context and evidence for thinking, it is inherently complex in structure. Worse yet, games are a wildly varied medium meaning and analysis from one game genre may have little use to another game or another genre.

Research with game data also suffers from a tight coupling of game production and research teams. This means that most potential learning science or analytics researchers simply don't have access to datasets until their team was also the designer of the game. This is compounded by the fact that games are expensive to produce, requiring teams with capacities in a number of technical and artistic domains such as illustration, modeling, animation, engineering, writing, mechanics design, UI design, and sound design. For researchers that are able to develop their own game content, significant expense and expertise are required to develop the infrastructure to capture, transfer, store, analyze and visualize game data. Often this infrastructure is developed in an ad hoc manner for a specific research agenda and not utilized across projects. Popular analytics tools such as game analytics and google analytics are optimized for sales and are not sufficient for exploring *learning*.

From a scientific standpoint, these issues lead to a lack of repeatability in studies that utilize game data. As the datasets are often unavailable and the algorithms used to explore and visualize these data were created in house and not easily shared, individual research products are nearly impossible to critique and repeat. While projects such as ADAGE (Stenerson et al., 2014) and TERC's Data Arcade (Rowe et al., 2017) attempt to solve some of these problems, no systems currently exist that allow for open distribution of game datasets and analysis methods that enable study replication or secondary analysis generally. The final result is that a very small number of institutions and a very small number of researchers are able to participate in research with game data.

1.2 Connections to the LAK Community

Despite these challenges, there has been consistent, significant interest in games from the LAK community. LAK 2021 included a workshop on the design of game-based learning analytics for classroom use (Kim et al, 2021). Additionally, submissions have explored the use of in-game, clickstream data (Liu, 2022) and peripheral, game-surrounding data (Carpenter et al, 2021; Park et al, 2021). This workshop builds on the growing interest in game-based analytics and brings together researchers to share tools and provide new opportunities for collaboration.

1.3 Objectives

During the proposed workshop, participants will learn about the significant promise of using game data to improve the design of learning experiences, to serve educators using games and to support students in new ways.

2 Organizational Details

This workshop will take place over a full-day interactive session with a mix of presentation, discussion, lab time and focus group sessions.

2.1 Schedule

Timing	**Activities**	**Leader**
30 min	Introductions and Goals	ALL
30 min	Intro to Game Data	BLINDED
45 min	Discussion: Develop a potential game data research question	BLINDED

30 min	Demo: Game data analysis for research, teaching and design	BLINDED
15 min	Break	
60 min	Lab Time: Visualizing Data	BLINDED
60 min	Lunch Break	
30 min	Introduction to Data Collection & Feature Creation Pipeline	BLINDED
45 min	Lab Time: New Feature Engineering	BLINDED
60 min	Focus Group: What infrastructure would support your research?	BLINDED
15 min	Discussion: Closing & Next Steps	BLINDED

2.2 Recruitment and Selection

We plan on recruiting participants to this workshop using a mix of social media from the proposers' significant audience and through posts to existing game learning analytics adjacent community channels such as the Learning Engineering Google Group, IDGA and CHI Play discord servers.

Optimal participants will come from institutions that do not have existing game research studios but do have support in learning analytics and data science. We will be looking to have significant representation from international and traditionally underserved populations.

3 INTENDED OUTCOMES

This workshop begins work in expanding the audience of researchers that are able to participate in game data analytics, supported by significant investments by the National Science Foundation and the Learning Agency to develop new game data research infrastructure. The high-level objective of this workshop is that a community of new researchers would submit new original research using game data at LAK 2024.

Goals exist for two audiences. The first audience are early career researchers and graduate students who have familiarity with basic learning analytics methods but have rudimentary programming skills. This audience is likely served by the recent influx of tools such as Rapidminer, Tablou, Google Data Studio, and Google Colab that have enabled new audiences to ingest and filter datasets, train and evaluate models, and develop visualizations with very little to no programming. For this audience the workshop will provide insights into the use of game data and access to raw and feature engineered datasets from a collection of studios. The second audience are researchers that are also proficient in data science-related software development. For this audience, we will not only provide new datasets but also tools, samples and workshop time to perform their own feature engineering of the existing datasets, expanding the range of how these datasets can be used for new research. For both audiences a key objective is to form a new community of LAK researchers that are interested in game data for learning analytics, inviting these new members into an existing multi institutional community that has been incubating over recent years.

A final objective is to solicit user feedback from early career researchers interested in game data that will inform the development of funded research infrastructure development and the organization of a new research coordinating community.

4 Infrastructure

Open Game Data is a repository of event and feature data from over a dozen educational games. Its infrastructure unifies event stream logging and feature engineering. Game events include player actions, game feedback, and player progression, in line with Owen & Baker's recommendations (2020). Feature engineering is handled with small code modules to ensure repeatability and easy modifiability. The workshop will utilize a subset of the data to introduce participants to potential applications through Binder. External to the workshop, participants will be able to obtain additional datasets from the Open Game Data website and create their own features using the open source Github repository.

5 References

Carpenter, D., Cloude, E., Rowe, J., Azevedo, R., & Lester, J. (2021, April). Investigating student reflection during game-based learning in middle grades science. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 280-291).

Evans, J. A. (2020). Digital Learning during the Pandemic: Emerging Evidence of an Education Transformation (p. 14). S*peak Up*.

Gauthier, A., Mavrikis, M., Benton, L., & Vasalou, A. (2022, March). Adoption and usage challenges of a Learning Analytics Dashboard for game-based learning: design and implementation implications. In *Companion Proceedings 12th International Conference on Learning Analytics & Knowledge (LAK22)*. Solar Research.

Liu, T. (2022). Using Gameplay Data to Investigate Students' Problem-Solving Behaviors in Zoombinis. *Companion Proceedings 12th International Conference on Learning Analytics & Knowledge (LAK22)*. Solar Research.

Mislevy, R. J., & Haertel, G. D. (2007). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, 25(4), 6–20.

Owen, V. E., & Baker, R. S. (2020). Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games. *Technology, Knowledge and Learning*, 25(2), 225–250.

Park, K., Sohn, H., Mott, B., Min, W., Saleh, A., Glazewski, K., ... & Lester, J. (2021, April). Detecting disruptive talk in student chat-based discussion within collaborative game-based learning environments. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 405-415).

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., Brown, R. A., & Edwards, T. (2017). Assessing implicit science learning in digital games. Computers in Human Behavior, 76, 617–630.

Shaffer, D. W. (2017). Quantitative Ethnography. Cathcart Press.

Takeuchi, L. M., & Vaala, S. (2014). Level up learning: A national survey on teaching with digital games (p. 67). The Joan Ganz Cooney Center at Sesame Workshop.

Interactive Workshop: Collaboration Analytics

Peter W. Foltz, Jason Reitman, Sidney K. D'Mello
University of Colorado, Boulder
{Peter.Foltz, Jason.Reitman, Sidney.DMello}@Colorado.edu

Sadhana Putambekar
University of Wisconsin, Madison
Putambekar@wisc.edu

Jamie Gorman
Arizona State University
Jcgorman@asu.edu

ABSTRACT: Collaboration is central to learning. However, analytic methods applied to analysis of small group collaboration are still in research stages and have yet to have significant impact in supporting students and educators in the classroom. In addition, collaboration analytics methods have been developed across a wide range of field, focusing on different aspects of group interaction, and cognitive, social, and affective states. This half-day interactive workshop brings together a diverse group of researchers who are working with student collaboration data and developing collaborative analytics. Workshop participants will share their methodologies as well as learn about other approaches that may come from different perspectives of collaborative analytics. In a series of guided discussions and interactive sessions, participants will work with their own and others' sets of data to try different approaches to analyzing student collaborative work.

Keywords: Collaboration Analytics, Collaborative Problem Solving (CPS), Computer Supported Cooperative Learning (CSCL), Multimodal analytics, Skill frameworks, Team Science.

1 INTRODUCTION

Learning is inherently collaborative and social (e.g., Bransford et al., 2000; Vygotsky, 1978). As preparation for careers, graduates will be expected to work closely with others (Levy & Murmane, 2005). Indeed, 94 percent of employers surveyed as part of the MetLife Survey of the American Teacher characterized working in teams as either "very important" or "absolutely essential" (Markow & Pieters, 2011). The need for developing collaboration skills has also been reflected in national and international educational standards for math and science (NGA, 2010; NGSS, 2013) as well as in standards and frameworks for developing skills for the 21st century workforce (Griffin, McGaw & Care, 2012; OECD 2015). A common theme across these standards emphasize engaging students in collaborative knowledge-building and problem solving to develop disciplinary ideas and reasoning through a range of activities and types of tasks.

Despite this increased emphasis on incorporating collaborative learning throughout the curriculum, techniques for teaching, assessing, and providing feedback are not widely implemented in classrooms or well embedded within curricula. This is largely due to the fact that it can be challenging for a teacher to orchestrate rich collaborative learning activities in the classroom and monitor and support teams of students all interacting simultaneously in real-time. Thus, there remain large opportunities in developing theories and methods of collaboration analytics that can be turned into effective tools and techniques to support students and teachers.

Learning analytics applied to student interaction data provide a means to instrument, measure, and understand the rich collaborative experiences that can unfold in educational settings. Methods for measuring collaborative learning have been researched, developed, and implemented across different disciplines using varied theoretical and methodological perspectives. These include:

- *Skills frameworks*, such as the internationally recognized PISA (OECD, 2015) and ATC21s (Griffin, Care & McGaw, 2012), provide detailed descriptions of human-human collaborative problem-solving skills and links to concrete indicators of behaviors related to those skills.
- *Computer Supported Cooperative Learning* (CSCL) has worked to understand how individuals learn in groups, how groups of learners construct shared knowledge, and how technology interacts with that learning (e.g., Dillenbourg, 1999; Puntambekar, Erkens & Hmelo-Silver, 2011); Wise & Jung, 2019.
- *Conversational agent* and *Natural Language Understanding* research has analyzed student discourse to mine the rich linguistic content generated by students and incorporated frameworks such as "academically productive talk", APT (Michaels & O'Conner, Kumar & Rosé, 2010) which examines discourse moves to support and facilitate collaborative conversations where students share and build on each other's ideas.
- *Team Science* has emphasized measuring real-time constructs of individual and team cognition and the dynamics of change over situations (e.g., Cooke et al., 2013; Gorman et al., 2020) and team communication frameworks have focused on ways to measure communication style (e.g., Foltz & Martin, 2008) to address both team cognition and peer mentoring theory that show that how teams communicate can impact their functioning more than how much they communicate (e.g., Marlow, Lacerenza, & Salas, 2017).
- *Distributed cognition* views cognition as processes that go beyond any individual's mind and incorporates other individuals and artifacts of the work environment to examine them as an interacting whole system (e.g., Hutchins, 1995; Wright et al., 2000).
- *Multimodal, multiparty methods* have focused on characterizing the rich sources of information from modalities such as gesture, body movement, eye-gaze, paralinguistics, body movement and linked these markers to social/cognitive/affective states related to collaborative performance (e.g., Praharaj et al., 2019).

Each perspective provides valid, yet slightly different analytic-based windows that elucidate our understanding of collaboration as whole. However, research efforts seldom incorporate more than one perspective or technical approach. The workshop is designed to bring together researchers from different perspectives to discuss their approaches as well as work interactively and hands-on with shared data sets.

2 WORKSHOP OBJECTIVES AND INTENDED OUTCOMES

2.1 Objectives

The objectives of the workshop are to bring together a diverse group of researchers who are developing collaborative analytics and to share their methodology and illustrate techniques on their own data. It is expected that the workshop will help the LAK community by improving researchers' measures of collaborative skills, generate generalizable insights around techniques for assessing collaboration, and improve the linking of analytic methods to theories of cognitive performance, pedagogy, and social/affective functioning of individuals and groups, and teams.

2.2 Dissemination of outcomes

The outcomes of the conference workshop will be a set of short position papers which outline each researcher's perspective on collaboration analytics, theoretical background, analytic techniques applied, and tasks and contexts to which the analytics is applied. The organizers will compile the position papers along with a larger summary document that outlines the full space of collaboration analytics. These papers will be made available through the workshop organizer's webpage at: https://sites.google.com/colorado.edu/collaborativeanalytics/home. Information and methods developed from the workshop will further be made available through the different participating collaborative analytic communities (e.g., LAK, Team science, CSCL, CPS, NLU).

REFERENCES

Bransford, Brown, D. A. and Cocking, R. eds. (2000) *How people learn: Brain, mind, experience, and school committee on developments in the science of learning*, National Academy Press: Washington, DC.

Brannick, M.T. and Prince, C. (1997), An overview of team performance measurement, in *Team performance assessment and measurement. Theory, methods, & applications*, M.T. Brannick, E. Salas, and C. Prince, Editors., Lawrence Erlbaum Associates: Mahwah, NJ. p. 3-16.

Cooke, N.J., Gorman, J. C., Myers, C. W. and Duran, J. L. (2013), Interactive team cognition. *Cognitive Science,*. **37**(2): p. 255-285.

Dillenbourg, P. (1999), What do you mean by collaborative learning, in *Collaborative-learning: Cognitive and computational approaches*, P. Dillenbourg, Editor.: Oxford. p. 1-15.

Foltz, P. W., & Martin, M. J. (2008). Automated communication analysis of teams. In *Team effectiveness in complex organizations* (pp. 445-466). Routledge.

Griffin, P., Care, E. and McGaw, B. (2012). *The changing role of education and schools*, in *Assessment and teaching of 21st century skills*, P. Griffin, B. McGaw, and E. Care, Editors. 2012, Springer: Heidelberg. p. 1-15.

Gorman, J.C., Grimm, D.A., Stevens, R.H., Galloway, T., Willemsen-Dunlap, A.M., & Halpin, D.J. (2020). Measuring real-time team cognition during team training. *Human Factors, 62*, 825-860.

Hutchins, E.L. "How a cockpit remembers its speed." Cognitive science, 1995, Vol. 19, pp. 265-288.

Kumar, R. and Rose, C. P. (2010) Architecture for building conversational agents that support collaborative learning. IEEE Transactions on Learning Technologies, **4**(1): p. 21-34.

Markow, D., & Pieters, A. (2011). *The MetLife survey of the American teacher: Preparing students for college and careers*. New York.

Marlow, S. L., Lacerenza, C. N., & Salas, E. (2017). Communication in virtual teams: A conceptual framework and research agenda. *Human Resource Management Review*, *27*(4), 575-589.

Michaels, S. and O'Connor, C. (2015) Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion, in *Socializing intelligence through talk and dialogue*, L.B. Resnick, C.S.C. Asterhan, and S.N. Clarke, Editors., American Educational Research Association. p. 33-248.

OECD, *PISA 2015 Collaborative Problem Solving Framework*, (2015), Organisation for Economic Cooperation and Development (OECD).

Praharaj, S., Scheffel, M., Drachsler, H. and Specht, M. (2019). Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics. Can We Go the Whole Nine Yards? *IEEE Transactions on Learning Technologies*. 14, 3 (2019), 367–385.

Puntambekar, S., Erkens, G., & Hmelo-Silver, C. (Eds.). (2011). *Analyzing interactions in CSCL: Methods, approaches and issues (Vol. 12)*. Springer Science & Business Media.

Vygotsky, L., (1978) *Mind in society: The development of higher psychological processes*, Cambridge. MA: Harvard University Press.

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.

Wright, P.C., Fields, R.E. and Harrison, M.D. (2000). Analyzing Human-Computer Interaction as Distributed Cognition: The Resources Model. *Human-Computer Interaction*, Vol.15.

Participatory Co-Design of Platform-Embedded Learning Experiments: LAK 2023 Workshop

Ryan S. Baker
University of Pennsylvania

Stephen E. Fancsali
Carnegie Learning, Inc.

Neil Heffernan
Worcester Polytechnic Institute

Rene Kizilcec
Cornell University

Debshila Basu Mallick
Rice University

Danielle McNamara
Arizona State University

Benjamin Motz
Indiana University

Steve Ritter
Carnegie Learning, Inc.

Jeremy Roschelle
Digital Promise

ABSTRACT: This half-day interactive workshop focuses on participatory design of future learning experiments that could be embedded within emerging digital learning platforms that are guided by the Standards of Excellence in Education Research (SEER) principles published by the U.S. Department of Education, Institute of Education Sciences (IES). Researchers, including students, practitioners, policy makers, and others attending the workshop, will learn about the SEER principles and opportunities to conduct platform-based learning research on six widely used digital learning platforms before getting the opportunity to participate in co-design activities with representatives/developers of the platform of their choice.

Keywords: participatory design, co-design, digital learning platforms, SEER principles, experiments, A/B tests

1 INTRODUCTION

This half-day workshop introduces a suite of emerging digital learning platforms for conducting learning and education research, unified by the ideals of SEER principles (Standards for Excellence in Education Research) (U.S. Department of Education, 2022). With coordination by Digital Promise and Empirical Education, this network of platforms seeks to connect developers, researchers, and educators to share ideas, build knowledge, and strengthen dissemination. Six digital learning platforms will be represented at the workshop: E-Trials (Heffernan & Heffernan, 2014), UpGrade/MATHia (Ritter et al., 2020), Terracotta, OpenStax Kinetic, Arizona State University (ASU): Learning @ Scale, and Behavioral Intervention Research Infrastructure (BIRI)/Realizeit.

Following introductions, workshop participants will closely collaborate with the attending digital learning platform developers to brainstorm and co-design future platform-embedded experiments that could be deployed within these systems. The workshop will build awareness and knowledge of the digital learning platforms' efforts to drive collaborative, platform-based learning research, the SEER principles, as well as cultivate potential future collaborations and partnerships between learning analytics/science researchers and the developers of widely used digital platforms for learning. At the conclusion of the workshop, the findings will be synthesized and disseminated broadly in the form of a short paper and a series of blog posts from each of the learning platforms that are represented at the workshop. Moreover, the workshop organizers will create a Slack workspace for continuing conversation and connecting researchers and other LAK community members with each of the participating digital platforms. Discussion at the workshop (e.g., during the Read Out and closing) and subsequent Slack channel conversations will focus on next-steps to continue cultivating partnerships and collaborations between researchers and digital learning platforms.

2 DIGITAL LEARNING PLATFORMS

- E-Trials (https://www.etrialstestbed.org/)
- UpGrade/MATHia (https://www.upgradeplatform.org/)
- TerraCotta (https://www.terracotta.education/)
- Kinetic/OpenStax (https://openstax.org/kinetic)
- ASU: Learning @ Scale (https://learningatscale.asu.edu/)
- Behavioral Intervention Research Infrastructure/Realizeit (https://biri-research.org/)

3 WORKSHOP ACTIVITIES

Session #1: Introduction to SEER Principles

- Introduction to the workshop
- Introduction to the SEER principles
- Discussion of the SEER principles

Session #2: Introduction to Digital Learning Platforms

- E-Trials
- UpGrade/MATHia

- TerraCotta
- Kinetic/OpenStax
- ASU: Learning @ Scale
- Behavioral Intervention Research Infrastructure/Realizeit
- Introduction to participatory co-design activity

Session #3: Co-Design Activity

- Breakout groups focus on each digital learning platform; participants choose the platform they'd like to work with and propose platform-based research/experiments that may be appropriate for the platform.
- Groups elect a "reporter" to participate in the "read out" session.

Session #4 Read Out & Closing

- Read Out: Groups report out on the results of the co-design activity.
- Closing: Opportunities for future collaboration, next-steps, etc., are discussed.

REFERENCES

Department of Education. (2022, June 22). *Standards for Excellence in Education Research: About*. United States Government. https://ies.ed.gov/seer/

Heffernan, N. & Heffernan, C. (2014). The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. *International Journal of Artificial Intelligence in Education, 24*(4), 470-497.

Ritter, S., Murphy, A., Fancsali, S. E., Fitkariwala, V., Patel, N., and Lomas, J. D. (2020). UpGrade: An open source tool to support A/B testing in educational software. *Proceedings of the First Workshop on Educational A/B Testing at Scale*. EdTech Books.

LAK23 Assess: The 3rd Workshop on Learning Analytics and Assessment

Dragan Gašević and Mladen Raković
Monash University
{dragan.gasevic, mladen.rakovic}@monash.edu

Naif Aljohani
King Abudlaziz University
nraljohani@kau.edu.sa

José A. Ruipérez Valiente
University of Murcia
jruiperez@um.es

Sandra Milligan
University of Melbourne
s.milligan@unimelb.edu.au

Blaženka Divjak
University of Zagreb
bdivjak@foi.hr

ABSTRACT: The first two editions of the Workshop on Learning Analytics and Assessment were successfully organized at LAK21 and LAK22 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. An open call for contributions will be distributed to solicit brief descriptions of current research and practice projects for roundtable-style discussions with workshop participants. Expected outcomes are the formation of a community of practice and possible follow-up publications.

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 et al., 2015; Gardner & Brooks, 2018; Greene et al., 2019), analysis of learning strategies and 21st century skills (e.g., Jovanović et al., 2017; Matcha et al., 2019), adaptive learner support and personalized feedback at scale (e.g., McNamara et al., 2012; Molenaar et al., 2012), and frameworks for ethics, privacy protection, and adoption (e.g., Tsai et al., 2018).

1.1 Challenge

Regardless of many promising results, the field still needs to address some critical challenges, including those at the intersection between learning analytics and assessment. For example, how can learning analytics be used to monitor learning progress? How can learning analytics inform formative and summative assessment as learning unfolds? In which ways can validity and reliability of data collection and analysis in learning analytics be improved? These challenges are of high significance in contemporary society that more and more requires development and use of complex skill sets (Greiff et al., 2017). Therefore, learning and assessment experience are closely associated. A growing body of research in educational data mining has been done on developing techniques that can support intelligent tutoring systems with the mechanisms for skill development (Corbett & Anderson, 1994; Desmarais & Baker, 2012). Yet, there is limited research that looks at how data collected and methods applied in learning analytics can be used and possibly constitute a formative or summative assessment. Moreover, can such data and methods satisfy requirements for assessments articulated in psychometric properties, methodological models, and different types of validity and reliability?

The role of learning analytics in analysis of assessment trustworthiness is another open research challenge. This has particularly been emphasized during the COVID19 pandemic with the emergency transition to distance and online education that also required different approaches to assessment that go beyond proctored exams. Several studies proposed the use of data analytic methods for detection of potential academic dishonesty and cheating behaviors. Although some interesting insights are ported and a strong potential to detect suspicious behaviors is demonstrated, there are many open challenges related to technical, ethical, privacy, practical, and policy issues of the development, implementation, and use of such data analytic methods.

1.2 Prior Accomplishments of LAK Assess

The first two editions of the Workshop on Learning Analytics and Assessment were successfully organized at LAK21 and LAK22 conferences. At each workshop, we gathered around 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 third edition of the workshop.

1.3 Objectives

The main objective of this workshop will be to continue promoting research and practice that looks at the intersection of learning analytics and assessment. We will examine approaches that build upon established principles in educational assessment to improve reliability, validity, usefulness of data collection and analysis in learning analytics. In the workshop, we will also look into the ways how learning analytics can contribute to the future developments in assessment for summative and formative purposes. In addition, we will examine practices for the use of learning analytics to assure assessment trustworthiness, with particular attention to the socio-technical nature of potential challenges. The workshop will also be an opportunity to further frame and shape special issues as important products for the connections between LA and assessment.

2 ORGANISATIONAL DETAILS

2.1 Proposed Half-Day Workshop Schedule

Table 1: Proposed schedule.

Timing	Description	Contributors
5 minutes	Welcome, introductions and plan for today	Organizers
5 minutes	In Memoriam - Dr. Saeed Ul-Hassan	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 writing	Presenters 3 & 4
25 minutes	Formative assessment of writing - roundtable	Participants
30 minutes	*Morning Tea*	
30 minutes	Learning analytics for assessment in games 20 minutes for presentation + 10 minutes for Q&A	Keynote
20 minutes	Analytics of assessment 7 minutes per presentation + 3 minutes for Q&A per presentations	Presenters 5 & 6
25 minutes	Analytics of 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 1st November 2022, deadline for abstract submissions 16th December 2022, and notification of acceptance 13th January 2023. 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.

Desmarais, M. C., & Baker, R. S. (2012). A Review of Recent Advances in Learner and Skill Modeling in Intelligent Learning Environments. *User Modeling and User-Adapted Interaction*, *22*(1–2), 9–38.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. User Modeling and User-Adapted Interaction, 28(2), 127-203.

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., ... & Panter, A. T. (2019). Modeling temporal self-regulatory processing in a higher education biology course. Learning and Instruction, 101201.

Greiff, S., Gasevic, D., & von Davier, A. A. (2017). Using process data for assessment in Intelligent Tutoring Systems. A psychometrician's, cognitive psychologist's, and computer scientist's perspective. *Design Recommendations for Intelligent Tutoring Systems. 5*, 171–179.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*(4), 74-85.

Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 461-470).

McNamara, D. S., Raine, R., Roscoe, R., Crossley, S. A., Jackson, G. T., Dai, J., ... & Dempsey, K. (2012). The Writing-Pal: Natural language algorithms to support intelligent tutoring on writing strategies. In Applied natural language processing: Identification, investigation and resolution (pp. 298-311). IGI Global.

Molenaar, I., Roda, C., van Boxtel, C., & Sleegers, P. (2012). Dynamic scaffolding of socially regulated learning in a computer-based learning environment. Computers & Education, 59(2), 515-523.

Tsai, Y. S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA framework: Informing institutional strategies and policy processes of learning analytics. *Journal of Learning Analytics*, *5*(3), 5-20.

7th Educational Data Mining in Computer Science Education (CSEDM) Workshop

Bita Akram, Thomas Price, Yang Shi
North Carolina State University
{bakram, twprice, yshi26}@ncsu.edu

Peter Brusilovsky
University of Pittsburgh
peterb@pitt.edu

Sharon I-Han Hsiao
Santa Clara University
ihsiao@scu.edu

Juho Leinonen
Aalto University
juho.2.leinonen@aalto.fi

ABSTRACT: There is a growing community of researchers at the intersection of data mining, AI and computing education research. The objective of the CSEDM workshop is to facilitate a discussion among this research community, with a focus on how data mining can be uniquely applied in computing education research. For example, what new techniques are needed to analyze program code and CS log data? How do results from CS education inform our analysis of this data? The workshop is meant to be an interdisciplinary event at the intersection of EDM and Computing Education Research. Researchers, faculty and students are encouraged to share their AI- and data-driven approaches, methodologies and experiences where data is transforming the way students learn Computer Science (CS) skills. This **full-day** workshop will feature paper presentations and discussions to promote collaboration.

Keywords: Computer Science Education, Educational Data Mining, AI in Education, Learning Analytics.

1 Workshop Goals

Computing is an increasingly fundamental skill for students across disciplines. It enables them to solve complex, real and challenging problems and make a positive impact in the world. Yet, the field of computing education is still facing a range of problems from high failure and attrition rates, to challenges training and recruiting teachers, to the under-representation of women and students of color.

Advanced learning technologies, which use data and AI to improve student learning outcomes, have the potential to address these problems. However, the domain of CS education presents novel challenges for applying these techniques. CS presents domain-specific challenges, such as helping students effectively use tools like compilers and debuggers, and supporting complex, open-ended problems with many possible solutions (Hsiao et al., 2010; Akram et al. 2020; Sarsa et al., 2022). CS also presents unique opportunities for developing learning technologies, such as abundant and rich log data, including code traces that capture each detail of how students' solutions evolved over time (Shi et al., 2022; Mao et al., 2021). It also provides opportunities for learning analytics researchers to analyze student learning more effectively (Mangaroska et al., 2020; Somyürek et al., 2020).

These domain-specific challenge and opportunities suggest the need for a specialized community of researchers, working at the intersection of AI, data-mining and computing education research. The goal of this 7th Educational Data Mining for Computer Science Education (CSEDM) is to bring this community together to share insights for how to support and understand learning in the domain of CS using data. This field is nascent but growing, with research in computing education increasingly using data analysis approaches, and researchers in the EDM and LAK community increasingly studying CS datasets. This workshop will help these researchers learn from each other, and develop the growing sub-field of CSEDM.

The workshop will build on six successful prior CSEDM workshops at: the International Educational Data Mining Conference (EDM) in 2018, 2020, 2021, and 2022, the International Learning Analytics and Knowledge Conference (LAK) in 2019, and the International Conference on AI in Education (AIED) in 2019.

Each were fruitful and well-attended. Our past in-person workshops have been well attended, and our virtual events have **had over 100 people registered and over 70 simultaneous attendees**! The last proceedings were published in Zenodo.

We hope to keep our momentum with a 7th CSEDM Workshop, returning to LAK in 2023.

1.1 Relevant Topics

The workshop encourages contributions from the following topics of interest:

- Predictive and descriptive modelling for CS courses
- Adaptation and personalization within CS learning environments
- Intelligent support for collaborative CS problem solving
- Machine learning approaches to analyze massive CS datasets and courses
- Online learning environments for CS: implementation, design and best practices
- Multimodal learning analytics and combination of student data sources in CS Education
- Affective, self-regulation, and motivational modeling of students as related to CS learning
- Adaptive feedback and adaptive testing for CS learning
- Discourse and dialogue research related to classroom, online, collaborative, or one-on-one learning of CS
- Teaching approaches using AI tools
- Visual Learning Analytics and Dashboards for CS
- Network Analysis for programming learning environments
- Classification of student program code
- Natural Language Processing for CS forums and discussions
- Analysis of programming design and trajectory paths
- Recommender systems and in-course recommendations for CS learning
- Adaptive educational technology and CS pedagogy for non-majors

We will invite researchers who are interested in further exploring, contributing, collaborating and developing data- and AI-driven techniques for building educational tools for Computer Science to submit paper on any of these topics.

2 Workshop Organization

The workshop will be organized by a team of organizers and program committees with a history of CSEDM research.

3 Call for Participation

We will solicit three types of research contributions:

- **4-8 page Research Papers**: Original, unpublished work, addressing any of the topics of interest above.
- **4-6 page position Papers:** Papers that present a coherent discussions related to computer science educational data mining including but not limited to diversity and equity, future research and practice directions, and impacts on CS education.
- **2 page Descriptions of CS Tools/Datasets/Infrastructure (2 pages):** Descriptions of shareable Computer Science (CS) datasets; Descriptions of data mining / analytics approaches applied to specifically Computer Science datasets; Case studies of collaboration where reproducible practices were used to integrate or compose two or more data analysis tools from different teams; Descriptions of infrastructures that could collect and integrate data from multiple learning tools (e.g. forum posts, LMS activity and programming data).

4 Workshop Activities

The workshop will be a **full day** workshop. It will primarily consist of paper presentations, discussions to facilitate collaboration. Interactive sessions include multiple parallel, short presentations, where participants can float around to the presentations they are interested in, similar to a poster session. (See Section 5 for details on remote attendees). A tentative schedule is available on the workshop website when available.

5 Plans for Supporting Remote Attendees

We propose a hybrid format, where participants, including presenters, can participate remotely as needed. If the conference is held fully online, we can switch to an online format, as we did in CSEDM 2022. We will take the following steps to support remote attendees:

- The workshop will occur concurrently on Zoom (or another online platform) and in-person.
- All presenters will join the Zoom meeting and share their screen while presenting, giving remote attendees full access to presentations. If possible, we will integrate microphones into the Zoom meeting as well.
- For remote presenters, we will project the Zoom meeting to the in-person participants, so they can see it presented live, and then hold a live Q&A over Zoom.
- For interactive presentations, we will ask all presenters to bring computers and share their while presenting, so that remote attendees can join via Zoom breakout rooms. Remote presenters will be remote-only, and present their work in Zoom breakout rooms.

6 SOLICITATION PLAN

Building on our growing network of contributors to prior workshops, we intend to solicit participation on the workshop through the following mailing lists and research networks:

- ACM's Special Interest Group on Computer Science Education (SIGCSE)
- Computer Science Education (CSED) research list (from the ICER community)
- European Association of Technology-Enhanced Learning (EATEL) community
- User Modeling (UM) mailing list
- Asia-Pacific Society for Computers in Education (APSCE) community
- PSLC community list
- Relevant EU project consortia
- The International Educational Data Mining Society
- The Society for Learning Analytics Research (SoLAR)

We will also reach out to prior contributors to CSEDM Workshops to solicit additional submissions.

7 REFERENCES

Hsiao, I-H., Sergey Sosnovsky, and Peter Brusilovsky. "Guiding students to the right questions: adaptive navigation support in an E-Learning system for Java programming." Journal of Computer Assisted Learning 26.4 (2010): 270-283.

Akram, B., Min, W., Wiebe, E., Navied, A., Mott, B., Boyer, K. E., & Lester, J. (2020, January). Automated assessment of computer science competencies from student programs with gaussian process regression. In *Proceedings of the 2020 Educational Data Mining Conference*.

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

Shi, Y., Chi, M., Barnes, T., & Price, T. (2022). Code-DKT: A Code-based Knowledge Tracing Model for Programming Tasks. In *Proceedings of the 15th International Conference on Educational Data Mining* (p. 50).

Mao, Y., Shi, Y., Marwan, S., Price, T. W., Barnes, T., & Chi, M. (2021). Knowing" When" and" Where": Temporal-ASTNN for Student Learning Progression in Novice Programming Tasks. *International Educational Data Mining Society.*

Mangaroska, K., Sharma, K., Gaševic, D., & Giannakos, M. (2020). Multimodal Learning Analytics to Inform Learning Design: Lessons Learned from Computing Education. *Journal of Learning Analytics*, *7*(3), 79-97.

Somyürek, S., Brusilovsky, P., Çebi, A., Akhüseyinoğlu, K., & Güyer, T. (2020). How do students perceive their own and their peers' progress in e-learning?. *The International Journal of Information and Learning Technology*, *38*(1), 49-74.

Partnerships for Cocreating Educational Content

Steven Moore[1], Anjali Singh[2], Hassan Khosravi[3], Paul Denny[4], Chris Brooks[2], John Stamper[1]
Carnegie Mellon University[1], University of Michigan[2], University of Queensland[3], University of Auckland[4]
StevenJamesMoore@gmail.com

ABSTRACT: We propose the first annual workshop on Partnerships for Cocreating Educational Content. This in-person workshop explores ample opportunities in leveraging humans, AI, and learning analytics to generate content, particularly appealing to instructors, researchers, learning engineers, and many other roles. The process of humans and AI cocreating educational content involves many stakeholders (students, instructors, researchers, instructional designers, etc.), thus multiple viewpoints can help to inform what future generated content might be useful, new and better ways to assess the quality of such content and to spark potential collaborative efforts between attendees. We ultimately want to show how everyone can leverage recent advancements in learnersourcing, AI, and learning analytics, and engage all participants in shaping the landscape of challenges and opportunities in this space. Our hope is to attract attendees interested in scaling the generation of instructional and assessment content and those interested in online learning platforms.

Keywords: Educational content creation, human-AI partnerships, learnersourcing

1 BACKGROUND

Globally, as educational delivery continues to transition towards online platforms hastened by the COVID-19 pandemic, the need for scalable and effective assessments has emerged as a pressing issue for instructors and educators. Amid many other logistical issues that arise from emergency online education (Hodges et al., 2020), instructors often find themselves having to generate large banks of resources such as practice and assessment questions to accommodate this new learning format. The continual creation and improvement of assessment items allows for a greater breadth of topic coverage, helps to identify well-constructed and valid assessments, and as a result, enables improved learning analytics. However, instructors and teaching staff rarely have the time or incentive to develop quality questions for formative assessments that are often used for personalization and adaptive learning; instead their efforts are often focused on creating high-stakes assessments such as quiz or exam questions (Jones, 2019). This challenge motivates the need for supporting the efforts of educational content creation via partnerships that involve pairings of instructors, students, and AI.

Partnerships for cocreating educational content often involve four distinct and iterative phases: creation, evaluation, utilization, and instructor/expert oversight. A popular student-student and student-instructor partnership that is widely becoming adopted and involves all of these phases is learnersourcing. Learnersourcing involves students generating their own educational resources and content that can be leveraged by future learners (Khosravi et al., 2021). This offers a domain agnostic way to help scale the creation of high-quality assessments, while also helping students learn the course content. Learning analytics plays a key role in this process by providing insight into how we might effectively leverage students to create educational content. This can include optimally selecting students to create questions that target topics where those students have demonstrated expertise, as well as recommending student-generated content on specific topics to learners who are struggling with those topics.

Partnerships between student-AI and instructor-AI also present ample opportunity in terms of content creation and evaluation (Singh et al., 2022). Advances in natural language processing and generative models provide space for AI to play a fundamental role in the co-creation of content with humans or to assist with the automated evaluation of its quality. The quality evaluation of this content can be further supported by learning analytics related to how students perform on these human-AI cocreated questions, compared to traditional assessments. For instance, one newly emerging area where human oversight may be needed in the educational content creation space is with the use of generative language models. (Sarsa et al., 2022) propose the idea of robosourcing, where content generated by large language models can be used as a starting point for students to accelerate the educational content creation process. Related work has also leveraged natural language processing (Moore et al., 2020), trust-based networks (Darvishi et al., 2021), and deep learning methods (Ni et al., 2022) to assist students in the evaluation of both student- and AI-generated content. While human input remains critical in this creation and evaluation process, more work needs to look at using artificial intelligence to further support students and instructors as they create educational content.

2 CALL FOR SUBMISSION

While no submission is required to participate in the workshop, we encourage 2 page submissions of work-in-progress or position papers that are related to partnerships for co-creating educational content. Some related challenges are highlighted in Figure 1. When it comes to the evaluation process of having students or AI review and revise other student-generated questions, there is a challenge regarding how we can assist students in optimally acting on the provided feedback. How to best incorporate student evaluation of the materials into the learning process, such as through learner models used to power learning analytics, remains an open problem (Abdi et al., 2020). While research indicates the learning benefits of students generating questions, oftentimes the quality of student-generated questions requires improvement. Recent work demonstrated that MCQs authored by students performed as well as those authored by academics, but further work remains to investigate how we might leverage AI to assist students in making consistently high-quality learnersourced contributions (Huang et al., 2021).

Among these challenges with humans and AI cocreating educational content lie many opportunities to explore ways of making it more accessible and beneficial to student learning. A clear opportunity regarding the creation of student-generated content is the different ways we can encourage students to make high-quality contributions, such as leveraging self regulated learning interventions (Lahza et al., 2022). While much of the existing research around students and AI creating educational content involves the creation of multiple-choice questions, there are limitless activity types that can be created and evaluated using a plethora of techniques. For instance, students could work in conjunction with a large language model, like GPT-3, to develop and refine assessment questions or explanations of learning content (Sarsa et al., 2022, Moore et al., 2022). This can help them quickly improve the content they generate, while also engaging them in critical thinking as they review the model's suggestions, such as recommended distractors. On the one hand, the increasing automation supported by such models may suggest less need for human input, but there is a need for caution. In their review of the opportunities and risks offered by foundation models, Bommasani et al. explicitly warn against the removal of teachers from such a loop (Moore et al., 2021). Large language models are trained on broad data produced by humans, and thus are known to suffer from biases similar to humans. Using automatically generated content without human oversight for educational content generation runs the risk of perpetuating some of these biases. We see a human-in-the-loop approach, involving both students and instructors, as essential for moderating biases and improving and tailoring the performance of the underlying generative models for suitability in learnersourcing contexts.

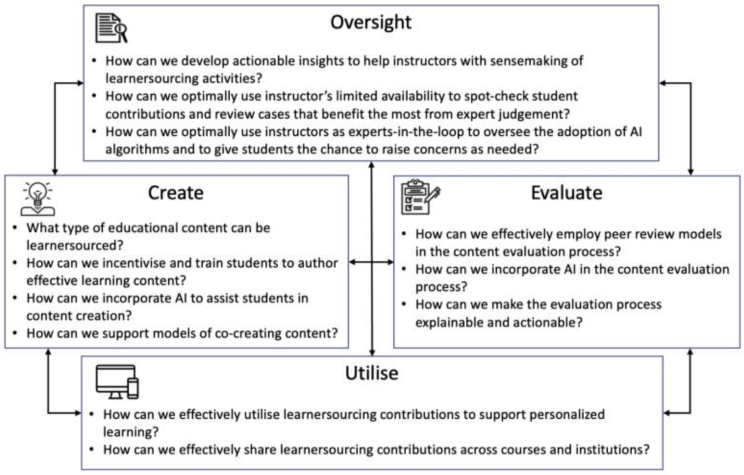


Figure 1: Challenges and opportunities relating to the four key aspects involved in the creation of educational content involving students, instructors, and AI.

3 WORKSHOP STRUCTURE

The workshop will run as an interactive half-day session with mini-presentations and round-table discussions on the theme. The provisional schedule is given below:

- **Introductions**: Introductions of workshop organizers and participants, and a background to the focus of the workshop.
- **Short Presentations**: Authors of accepted submissions present their work which would be followed by a Q&A session
- **Round-table discussion**: Participants will move around specific topics of interest related to various types of partnerships for creating educational content including partnerships between: student-student, student-AI, student-instructor, and instructor-AI.
- **Open discussion**: An open discussion will be facilitated among all participants summarizing activities from the round table discussions and building consensus using the co-creation of shared notes and resources.
- **Concluding remarks and community engagement:** Closing remarks on the workshop will be made with future steps. In addition, a Slack channel has been created to keep the participants involved and promote collaboration between attendees.

4 OUTCOMES

The main goal of this workshop is to explore how partnerships between students, instructors, and AI can be leveraged for creating educational content and how learning analytics informs these interactions. We believe participants from a wide range of backgrounds and prior knowledge on learnersourcing, machine learning, and learning analytics can both benefit and contribute to this workshop. As this creation of educational content involves many stakeholders (students, instructors, researchers, instructional designers, etc.), multiple viewpoints can help to inform what future student- and AI-generated educational content might be useful, new and better ways to assess the quality of the content, and spark potential collaboration efforts between attendees. The accepted submissions will be published as part of a CEUR proceedings.

REFERENCES

Abdi, S., Khosravi, H., & Sadiq, S. (2020). Modelling learners in crowdsourcing educational systems. *International Conference on Artificial Intelligence in Education*, 3–9.

Darvishi, A., Khosravi, H., & Sadiq, S. (2021). Employing peer review to evaluate the quality of student generated content at scale: A trust propagation approach. *Proceedings of the Eighth ACM Conference on Learning@ Scale*, 139–150.

Hodges, C. B., Moore, S., Lockee, B. B., Trust, T., & Bond, M. A. (2020). *The difference between emergency remote teaching and online learning*.

Huang, A., Hancock, D., Clemson, M., Yeo, G., Harney, D., Denny, P., & Denyer, G. (2021). Selecting student-authored questions for summative assessments. *Research in Learning Technology*.

Jones, J. A. (2019). Scaffolding self-regulated learning through student-generated quizzes. *Active Learning in Higher Education*, *20*(2), 115–126.

Khosravi, H., Demartini, G., Sadiq, S., & Gasevic, D. (2021). Charting the design and analytics agenda of learnersourcing systems. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 32–42.

Lahza, H., Khosravi, H., Demartini, G., & Gasevic, D. (2022). Effects of Technological Interventions for Self-regulation: A Control Experiment in Learnersourcing. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 542–548.

Moore, S., Nguyen, H. A., & Stamper, J. (2020). Evaluating Crowdsourcing and Topic Modeling in Generating Knowledge Components from Explanations. *International Conference on Artificial Intelligence in Education*, 398–410.

Moore, S., Nguyen, H. A., & Stamper, J. (2021). Examining the Effects of Student Participation and Performance on the Quality of Learnersourcing Multiple-Choice Questions. *Proceedings of the Eighth ACM Conference on Learning@ Scale*, 209–220.

Moore, S., Nguyen, H. A., Bier, N., Domadia, T., & Stamper, J. (2022). Assessing the quality of student-generated short answer questions using GPT-3. In the European conference on technology enhanced learning, 243-257.

Ni, L., Bao, Q., Li, X., Qi, Q., Denny, P., Warren, J., Witbrock, M., & Liu, J. (2022). Deepqr: Neural-based quality ratings for learnersourced multiple-choice questions. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(11), 12826–12834.

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*, 27–43.

Singh, A., Brooks, C., & Doroudi, S. (2022). Learnersourcing in Theory and Practice: Synthesizing the Literature and Charting the Future. In Proceedings of the Ninth ACM Conference on Learning@ Scale (234-245).

Workshop for Learning Analytics Graduate Programs

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

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

ABSTRACT: As the field of learning analytics has matured over the past decade, the demand for individuals with the necessary knowledge and skills to investigate and improve learning experiences and outcomes has increased dramatically. Over the past few years, universities have launched a number of graduate programs that aspire to fill the training gap for researchers and practitioners alike with additional institutions currently in the development process. To date, there have been few broad discussions of competencies, curriculum, and instructional approaches for teaching learning analytics and this session will bring together existing graduate programs to discuss key elements that can improve current efforts and provide guidance for institutions that plan to develop degrees and certificates. The goal of this workshop is to identify common and divergent program elements, successes and challenges, effective and ineffective strategies and approaches, and commit to the sharing of curricular resources for existing and future programs.

Keywords: Curriculum, Teaching, Graduate Programs

1 BACKGROUND

The growth of digital learning and maturation of the field of learning analytics has led to the increased demand for individuals and teams with the necessary knowledge and skills in data science methods and cyberinfrastructure, and competence and experience with educational data, research, and practice. Until recently, there have been few universities with learning analytics programs to fill the training gap for researchers and practitioners alike. In a recent contribution, Kizilcec & Davis (in press) found that there is no standard learning analytics curriculum and that “while most programs emphasized data literacy and an awareness of common analytic methods and systems as part of their learning goals, there was no common set of topics covered across all programs.” While there have been some broader discussions of competencies, curriculum, and instructional approaches for teaching learning analytics (e.g., LAK’19 Workshop on Building the LA Curriculum 2020 & Beyond; LALN Session on Designing Learning Analytics Courses, Programs, and High-Impact Practices), this session will bring together graduate learning analytics programs to discuss key elements with the goal of improving current efforts and providing guidance for institutions that plan to develop degrees and certificates in the future. The organizers seek to identify common and divergent program elements, successes and challenges, and effective and ineffective strategies and approaches as well as commit to the sharing of curricular resources for existing and future programs.

2 ORGANIZATIONAL DETAILS

2.1 Type of Event

This session will be an interactive workshop.

2.2 Duration

This workshop would follow a half-day format.

2.3 Workshop Activities

The workshop will include presentations by all programs, a guided activity, and small- and large-group discussions.

2.4 Proposed Schedule

Introduction (5)
Graduate Program Presentations (50)
Break (5)
Guided Activity (55)
Break (5)
Discussion (50)
Wrap-Up and Next Steps (10)

2.5 Expected Number of Participants

This workshop expects to attract approximately 30 participants.

3 OBJECTIVES AND OUTCOMES

3.1 Objectives

This workshop seeks to: 1) Convene learning analytics graduate programs to discuss competencies, curriculum, and instructional approaches for teaching learning analytics; 2) produce a report to share with participants and post publicly on the workshop website for broad dissemination; 3) build a commitment to support the improvement and growth of learning analytics graduate programs through collaboration, resource sharing, and future sessions.

3.2 Dissemination Plan

The workshop organizers will create a website for program identification, sharing of resources and workshop report(s), and future events.

4 COMMUNICATION PLAN

4.1 Recruitment

The organizers will recruit through individual invitations, social media platforms, networks, the Learning Analytics Google Group, and the conference website.

4.2 Information Sharing

The organizers will communicate via email prior to and following the event.

4.3 Tools

The organizers plan to make use of a website and shared Google Drive.

REFERENCES

Dellinger, J. T., Vovides, Y., Davis, J., Kellogg, S., & Chen, B. (2022). *Designing learning analytics courses, programs, and high-impact practices*. Learning Analytics Learning Network. Retrieved from https://learninganalytics.net/laln/previous-events/designing-la-courses-programs/

Kizilcec, R. F., & Davis, D. (in press). Learning analytics education: A case study, review of current programs, and recommendations for instructors. In O. Viberg & Å. Grönlund (Eds.), *Practicable learning analytics*. Springer Nature.

Lang, C., Teasley, S., & Stamper, J. (2019). *Building the learning analytics curriculum 2020 & beyond*. Retrieved from https://blogs.cuit.columbia.edu/cl3584/

Subversive Learning Analytics: Imagining LA Futures to Support Equitable Educational Practices

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

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

ABSTRACT: The challenges that emerge from the use of educational data in sociotechnical systems have received increasing attention in recent years. However, much of the discussion has centered around analysis of problems, such as issues of ethics and equity, or solutions to specific local quandaries. We lack holistic examination of embedded assumptions and values contributing to such problems and radical innovation of LA that might shift us away from such paradigms. In this half-day interactive workshop, we take up the idea of a “subversive stance” as a tool for generative discussions and insights around these questions and to spur ideas for unorthodox possibilities in LA that support equitable educational practices. Participants will reflect about deep seated assumptions of our discipline, consider alternatives, and ideate LA artifacts that could deal with these challenges.

Keywords: Learning Analytics, Equity, Socio Technical Systems, Innovation

1 OVERVIEW AND RATIONALE

Innovation has been said to stem from one of two processes. Incremental innovation emerges from evaluating existing situations and pragmatically detecting the gaps for improvement which themselves depend on its embedded value systems (Norman & Verganti, 2013). “Scaling up” an existing practice or making it “more efficient” are examples of this. To date, the bulk of LA work has been conducted within this tradition, accepting the frame of existing educational institutions and their historical value systems, and attempting to use data for “understanding and optimising” their practice (Siemens et. al., 2011). There is another way to innovate, however. A radical approach to innovation creates the possibility for bigger change: a change in core meaning or a shift in paradigms (Norman & Verganti, 2013). One way to foster such innovation is by consciously calling into question the values that are embedded in a system. Instead of being restricted to operating from within, such a process for innovation invites critical change by reflecting on the whys behind practice, meaning behind assumptions, and unorthodox possibilities for futures that might support alternative value systems.

The current landscape of socio-technical challenges surrounding the generation, analysis and use of data in education, including critical concern with the intersecting ways we engage with algorithms (Noble, 2018), AI (Broussard, 208), ethics (Prinsloo, 2017), surveillance (Benjamin, 2019) and equity (Scholes, 2016) suggest that although an incremental approach to LA development has value, it is not enough. To move towards developing a space for radical innovation in the field, Wise, Sarmiento and Boothe Jr. (2021) presented the idea of a "subversive stance" for LA. The concept draws from a number of critical traditions, including speculative design (Wong & Khovanskaya, 2018), the study of socio-technical systems (Achiume, 2014), feminism (D'Ignazio & Klein, 2021) and design activism (Costanza-Chock, 2018) and proposes ways to help us identify taken-for-granted assumptions-in-practice, ask generative questions about design processes and consider new models of creation to produce tools that may operate differently in educational ecosystems.

Such activities can help us move beyond recognition of problems related to the use of data in education and narrow attempts to ameliorate them in a specific application to take a broader perspective on the different roles that LA could play in educational data ecosystem; for example, illuminating the boundaries of where technology can and cannot help, making the dimensions of problems more salient and vivid (Abebe et. al., 2020), or being an agent for equitable practices (Williamson & Kizilcec, 2022).

In this workshop we will introduce the community to this approach, exploring together some of the deep-seated assumptions in educational practice, and collectively imagining unconventional tools and artifacts. We do so in this instance with a particular focus on equity and the support of equitable practices, noting that a subversive stance could also be taken with respect to other foci. The workshop activities will be guided by three objectives:

Objective 1: Inquire collaboratively into some of the assumptions and values embedded in educational systems that currently guide LA work.

Objective 2: Explore how different forms of thinking, questioning, or valuing can open spaces of possibility for future LA design.

Objective 3: Generate novel avenues for development of LA to support equitable practice.

This work will lay the foundation for both people- and idea-focused outcomes.

Outcome 1 Kickstart a community of LA practitioners, researchers and educators interested in exploring how different educational frameworks and value systems can contribute to, and be used to innovate the practice of LA.

Outcome 2 Develop the foundation for a publication that disseminates key ideas about the relationship between values, design and LA that emerged in the workshop.

2 ORGANISATIONAL DETAILS

2.1 Format, Recruitment, and pre-workshop activities

This first Subversive Analytics workshop is envisioned as a half-day interactive workshop. It will follow a hybrid in-person and remote format, with up to 25 participants. The hybrid model is important since it allows for more equitable participation by those who may not be able to travel to Texas for the conference. Participants can be academics, researchers, practitioners, or students with an interest in Learning Analytics and values-based innovation, as well as critical frameworks and education. Because of the nature of the topic, both those who are highly specialized in LA and those who are beginners in the field are equally welcome. We particularly encourage participation of those who have experience and expertise in considering questions of equity in education.

We will distribute information about the workshop to potential attendees through the LA Google Group mailing list, over Twitter and our individual professional networks. A survey will be distributed amongst participants before the event to better understand their background, and what value systems or experiences they can bring to the discussion. A website will be created with a suggested short list of pre-readings to help stimulate discussion in the workshop.

2.2 Synchronous activities

In the workshop our initial ideas of what a subversive LA might look like will be presented to the participants, to build on the concept collaboratively. Then,

1. **Whole group presentation and grounding discussion (45 min):** We share the existing conceptualization of taking a "subversive stance" and engage with participants to unpack its meaning and areas for further elaboration. We present a range of value systems (dominant and alternative) that may be relevant for educational design.

2. **Small group discussion activity (60 min).** Small groups of 3-5 participants will split up and discuss examples of current LA practice based on the generative questions outlined in Sarmiento, Wise & Boothe Jr (2021). Groups will work to develop a map of underlying assumptions and value systems that are embedded in everyday practice in LA and traditional education systems.

3. **Consolidation of Ideas (15 min).** Groups will share their mapping from the discussion activity, comparing perspectives and considering implications for equity in the education space.

4. **Short Break (15 min).**

5. **Co-creation activity (60 min)** Using creativity facilitation techniques, participants will work in small groups to ideate LA artifacts that challenge underlying assumptions, incorporate alternative value systems, or in some other way use LA to subvert traditional educational practice.

Sharing, Discussion and Planning for Next Steps (45 min): Groups present their ideas and their experiences in generating the ideas. In the discussion we will also discuss next steps for the development of the community of practice.

2.3 Logistics

The workshop will be hybrid. Both facilitators will be on site, but one will focus on working with the in-person participants, while the other will facilitate the processes of the discussion of online participants. Whole group activities will offer the opportunity to share across face-to-face and online participants. Zoom, Slack and Mural will be used to facilitate and communicate the discussions and sharing activities.

2.4 Dissemination strategy

A workshop website will be used to disseminate information about the workshop to the community, including suggested readings to prepare for the event. After the workshop, the website will be used to disseminate key artifacts that emerged from the discussions as well as a synthesis of the insights. These will also serve as the foundation for a conceptual publication as described above.

REFERENCES

Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020, January). Roles for computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 252-260).

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code*. Social forces.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Costanza-Chock, S. (2018). Design justice: Towards an intersectional feminist framework for design theory and practice. *Proceedings of the Design Research Society*.

D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.

Green, B. (2021). Data science as political action: Grounding data science in a politics of justice. *Journal of Social Computing*, 2(3), 249-265.

Noble, S. U. (2018). *Algorithms of oppression.* New York University Press.

Norman, D. A., & Verganti, R. (2014). Incremental and radical innovation: Design research vs. technology and meaning change. *Design issues*, 30(1), 78-96.

Prinsloo, P. (2017). Fleeing from Frankenstein's monster and meeting Kafka on the way: Algorithmic decision-making in higher education*. E-Learning and Digital Media*, 14(3), 138-163.

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development,* 64(5), 939-955.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.

Williamson, K., & Kizilcec, R. (2022). A review of learning analytics dashboard research in higher education: Implications for justice, equity, diversity, and inclusion. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 260-270). New York: ACM.

Wong R.Y., Khovanskaya V. (2018) Speculative Design in HCI: From Corporate Imaginations to Critical Orientations. In: Filimowicz M., Tzankova V. (eds) *New Directions in Third Wave Human-Computer Interaction: Volume 2 - Methodologies. Human–Computer Interaction Series*. Springer, Cham

Data Storytelling in Learning Analytics

Lujie Karen Chen[1], Jiaqi Gong[2], Louise Yarnall[3], and John Fritz[1]
[1]University of Maryalnd Baltimore County, Maryland, USA
[2]The University of Alabama, Tuscaloosa, Alabama, USA
[3]SRI International, Menlo Park , California, USA

ABSTRACT: Data storytelling has seen exponential growth in real-world demand in recent years. Its growing interest in the field of learning analytics (LA) is not an exception. In order for the learning analytics product to make real impacts, LA researchers and practitioners need to be equipped with the competence to construct coherent, unbiased, and compelling stories for various types of LA stakeholders. In recent years, there are emerging themes of data storytelling in the community of LA research and practices. Many LA-related data stories have been created, shared, and reflected on. Research processes and products have been explored around data storytelling. In this workshop, we will invite LA researchers and practitioners to create, share and reflect on their own LA stories and think critically and creatively about data storytelling in LA: What does a good LA story look like? What are the patterns of effective LA stories? What are the success and failures of storytelling in LA? How we could train LA researchers and practitioners to be better storytellers? This workshop will bring together researchers and practitioners in learning analytics and data storytelling to explore the strategies and tactics for telling effective LA data stores and the related challenges and opportunities.

Keywords: data science, data analytics, communication, data storytelling

1 BACKGROUND

Storytelling draws on underlying theories of rhetoric. Theorists have framed data storytelling as blending rhetorical forms of information presentation and persuasion with concepts of scientific exploration and sensemaking (Segel & Heer, 2010). To generate a story, an analyst needs to think in terms of “behaviors, events, and plots such that it leads to comprehension, discovery, hypothesis generation, and communication” (Eccles et. el 2008) and authorial messages, audience needs, visual narrative, and narrative structure (Segel & Heer, 2010). Data storytelling involves developing a story logic, synthesizing story elements, understanding visual representation needs, and checking for bias. A good data storyteller can construct coherent, unbiased, and compelling stories or narratives for audiences of various types.

Learning analytics (LA) is a field built around the generation, analysis, and sense-making of quantitative information, guided by deliberate goals of turning data into actionable insights to induce changes to improve learning. As such, it has a natural connection to data storytelling. Data stories may serve as a tool of communication to connect stakeholders, including researchers, designers, data analysts, data owners, and users, and various kinds of stakeholders such as instructors, learners, parents, and administrators. Building a community that values the role of data stories in the LA ecosystem, and helps to support the development of data storytelling competency may have a profound impact on the building of LA systems with high levels of transparency and trust.

2 OBJECTIVE

In recent years, LA researchers have explored the power of storytelling in designing dashboards and user engagement in general (Echeverria et al., 2018; Fernandez Nieto et al., 2022; Martínez-Maldonado et al., n.d.; Martinez-Maldonado et al., 2020). In LAK 2021, a workshop on "A Tutorial on Data Storytelling for Learning Analytics Dashboards" was conducted. This proposal builds on those existing works and broadens the scope of LA data storytelling beyond the goal of designing dashboards. We are interested in engaging a multidisciplinary group of LA researchers and practitioners, as well as those with cross-disciplinary interests in learning science, computer science, data science, information visualization, communication science, and human-centered design to explore the strategies and tactics in telling effective LA data stores and the related challenges and opportunities.

The overarching goal of the workshop is to stimulate discussions on the theory and practice of data storytelling in learning analytics. Specifically,

a. **Theory: (re)defining data storytelling in LA -** what are the standards of good LA stories?
b. **Practice: How to tell good LA stories** - what are the principles and techniques of telling good LA stories?

3 OPEN QUESTIONS

In this workshop, we are interested in engaging participants in exploring the following list of questions:

1. In what part of the LA ecosystem could data storytelling have a high impact? e.g., design, implementation, evaluation, and adoption?
2. Who can benefit from storytelling in LA communities?
3. Who are the main actors of the LA data storytelling processes? e.g., who are the storytellers, and who is the audience?
4. What are the training/education challenges to support researchers and practitioners in LA communities to acquire data storytelling competencies?
5. Whare the common patterns of effective storytelling in LA communities?
6. What is the connection between data storytelling and LA system adoption?
7. What is the connection between LA data storytelling and data literacy and LA literacy education?

4 INTENDED AUDIENCE

This workshop welcomes participants from multiple disciplines interested in exploring the power of data storytelling in LA, regardless of their experience level in data storytelling. This group could include LA researchers and practitioners and those engaged in data storytelling-related disciplines such as learning science, computer science, data science, information visualization, communication science, and human-centered design.

5 PROPOSED ACTIVITIES

5.1 WORKSHOP ACTIVITIES

The workshop will be opened with a presentation on an overarching framework of LA storytelling and followed by a series of interactive talks (20-30 mins), each with a short presentation and interactive activities. We plan to have 4-5 talks, each with a specific focus along those two dimensions : (1) components of the overarching storytelling framework (data analytics, visualization, or narrative); (2) target audiences (e.g. students or instructors).

5.2 PRE-workshop activities

5.2.1 Pre-workshop Readings

SEGEL, E., & HEER, J. (2010). NARRATIVE VISUALIZATION: TELLING STORIES WITH DATA. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, 16(6), 1139–1148. HTTPS://DOI.ORG/10.1109/TVCG.2010.179

KOSARA, R., & MACKINLAY, J. (2013). STORYTELLING: THE NEXT STEP FOR VISUALIZATION. COMPUTER, 46(5), 44–50. HTTPS://DOI.ORG/10.1109/MC.2013.36

OJO, A., & HERAVI, B. (2018). PATTERNS IN AWARD WINNING DATA STORYTELLING: STORY TYPES, ENABLING TOOLS AND COMPETENCES. *DIGITAL JOURNALISM*, *6*(6), 693–718. HTTPS://DOI.ORG/10.1080/21670811.2017.1403291

5.2.2 Pre-workshop reflection

Before the conference, participants are asked to reflect on their understanding and experience of LAK storytelling at the interaction of ***learning, data, and stories.***

For those with data storytelling experience: who are the audiences? what are the communication objectives? what is the communication medium? what was the creation process? etc.

For those without LAK storytelling experience: in your mind, what is data storytelling in learning analytics? what are strategies for effective storytelling? what are some good examples of LAK stories?

REFERENCES

Chen, Q., Li, Z., Pong, T.-C., & Qu, H. (2019). Designing Narrative Slideshows for Learning Analytics. *2019 IEEE Pacific Visualization Symposium (PacificVis)*, 237–246. https://doi.org/10.1109/PacificVis.2019.00036

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Buckingham Shum, S. (2018). Driving data storytelling from learning design. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 131–140. https://doi.org/10.1145/3170358.3170380

Echeverria, V., Martinez-Maldonado, R., Shum, S. B., 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), Article 3. https://doi.org/10.18608/jla.2018.53.6

Fernandez Nieto, G. M., Kitto, K., Buckingham Shum, S., & Martinez-Maldonado, R. (2022). Beyond the Learning Analytics Dashboard: Alternative Ways to Communicate Student Data Insights Combining Visualisation, Narrative and Storytelling. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 219–229. https://doi.org/10.1145/3506860.3506895

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., & Buckingham Shum, S. (2020). From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376148

Martínez-Maldonado, R., Yan, L., Deppeler, J., Phillips, M., & Gašević, D. (n.d.). *Classroom Analytics: Telling Stories about Learning Spaces using Sensor Data*. 20.

Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Stories in GeoTime—Ryan Eccles, Thomas Kapler, Robert Harper, William Wright, 2008. (n.d.). Retrieved October 7, 2022, from https://journals.sagepub.com/doi/abs/10.1057/palgrave.ivs.9500173?casa_token=GzPv_LcCpQoAAAAA:9OQyVKY2uyoOj0soCCtaDcjiSdYh2hSKNzuRGMO3JOyxvC7siy0OS9gJsjuJRuCJAreh5frkw-rqKw

Dykes, B. (2019). Effective data storytelling: how to drive change with data, narrative and visuals. John Wiley & Sons.

DesignLAK23: Towards an integrated framework for design-aware learning analytics

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

ABSTRACT: The 8th Annual workshop brings together work by the learning analytics community over the past decade to assess what has been achieved in bridging learning design (LD) and learning analytics (LA) and explore what further challenges need to be overcome in order to guide the development of systems that can support the provision of pedagogically meaningful learning analytics to teachers and learners. A working paper written for the purpose of this workshop will be used to introduce key frameworks proposed in literature, and situate how these contributions provide building blocks in relation to what would be needed to create an integrated, continuous bridge between LD and LA. The workshop activities will involve learning scenarios to enable the critique of a proposed integrated framework for LD/LA. Participants will work in and share across groups to identify the usefulness of the framework presented in the working paper, as well as the gaps and challenges that remain to be resolved to realise an operationalisable link between LD and LA. Another outcome of the workshop will be the development of evaluation projects that will be conducted throughout the year and showcased in next year's DesignLAK workshop.

Keywords: Learning analytics, learning design, frameworks, evaluation

1 BACKGROUND

For the past decade there has been a broad recognition in the learning analytics community of the importance of the connection between learning design (LD) and learning analytics (LA). In response, there have been multiple attempts to conceptualise and articulate this connection in the form of frameworks and models to facilitate the design of systems that can realise the potential of LA to improve teaching and learning in authentic educational settings. These frameworks have represented the connection of LA and LD from different perspectives with some that are broad in their view of the elements and stakeholders involved (e.g., Bakharia et al., 2016), while others have examined the LD/LA elements in more detail providing taxonomies and layers to link design and analytics (e.g., Law & Liang, 2010; Hernández-Leo et al., 2019). There have also been efforts to model the approaches needed to operationalise these frameworks so that practitioners and teachers can work through from design to analytics in practical ways (e.g., the development of a layered storytelling approach for communicating the interpretation of LA findings to teachers and learners in pedagogically relevant, non-technical language (Martinez-Maldonado et al., 2020)).

However, there is a need for broader conversations related to the need for articulation and greater consensus around the frameworks and the vocabulary used by researchers and practitioners in the field when bringing together LD and LA. Importantly, there is a need for an integrated framework for contextualising the contributions of emerging research and development outcomes to accelerate the appropriation of new advances by the LA and LD communities for the construction of design-aware learning analytics systems.

Over the past seven years the DesignLAK workshops have focused on a range of perspectives on the relationship between learning design and learning analytics. This has included workshops on particular aspects of learning design such as feedback processes (Authors, 2016) and elements related to assessment design (Authors, 2017; Authors, 2019). Other workshops have profiled tools that have been designed to provide a link between learning designs and analytics (Authors, 2018; Authors, 2022), or prototyping tools which enable the visualisation of learning analytics with reference to design patterns (Authors, 2021). The lively and constructive discussions held in these workshops often came back to a realisation that there are still significant conceptual and technological gaps to be addressed to provide an operationalisable foundation for the construction of systems that can provide pedagogically meaningful, LA-grounded feedback for teachers and learners in common, authentic learning scenarios. Addressing this need is an intent of the DesignLAK23 workshop design so that a contribution can be made to benefit all stakeholders in the way that learning analytics can be used in educational environments.

2 OBJECTIVES OF THE WORKSHOP

The primary objective of the DesignLAK23 workshop is to explore and connect existing work in the area of learning analytics and learning design to work towards an integrated framework that can be used to inform the development and use of design-aware learning analytics. Participants will be given the opportunity to engage with and critique a draft integrated framework through the exploration of authentic learning scenarios. The resulting feedback on the framework will be used to inform the development of an ongoing program of work to continue the conversation of the connection between LD and LA in ways that can be easily operationalised by key stakeholders in educational settings.

3 WORKSHOP DESIGN

The DesignLAK23 workshop is designed to be a half-day workshop held face-to-face (although hybrid delivery could be accommodated if there is a strong demand and supported by the conference). The workshop will be highly interactive in nature with participants working together to discuss, apply, critique, and provide feedback on the integrated LD/LA framework proposed in a working paper provided to participants in advance of the workshop. The evaluation of the framework will be conducted through its application to an authentic learning scenario using an assessment platform, Ruby, developed by the University of [blinded for review]. The number of participants for the workshop should be capped at 40 to allow inclusive full-group discussions and sharing of critiques and suggestions for future development of the integrative framework for LD/LA.

3.1 Pre-workshop preparation

Prior to the workshop the organisers will work on the development of a working paper that draws together what has been learnt as a field over the last decade related to the connection between learning design and learning analytics. Existing frameworks and commentary will be reviewed to provide a comprehensive exploration of the conversation so far and to contribute to the design of a proposed integrated LD/LA framework to consolidate this work across the field. The working paper will be prepared in collaboration with a group of key scholars from the field who have proposed frameworks/models. The inclusion of invited collaborators is to capture a range of perspectives and considerations that have contributed to conversations to date relating to the link between LD and LA. The working paper will be distributed to workshop participants two weeks prior to the workshop to allow them time to read and engage with the ideas presented in preparation for the workshop discussions. To attract participants with an interest in contributing to this critique and conversation, the workshop will be promoted through several channels including via social media (e.g., Twitter, LinkedIn) and mailing lists of learning analytics-related groups that the organisers are associated with (e.g., SoLAR, ASCILITE, etc.). A website will also be established to host more detailed information about the workshop and its design, as well as guidance on how to participate.

3.2 The workshop

The workshop will be structured in five parts. Part 1 will include an introduction to the workshop aims and outcomes, as well as an icebreaker activity to allow participants to get to know each other and why they chose to engage in this LD/LA conversation. This will be followed in Part 2 by a discussion of the proposed integrated LD/LA framework and exploration of the key elements and layers for consideration, as presented in the working paper. In Part 3, participants will have a chance to work together in groups to apply the integrated framework to authentic learning scenarios. In order to engage with these scenarios, access will be provided to the Ruby platform, which is an easily navigable assessment platform that can monitor, track and support developmental learning, and can demonstrate students' progress at any point in time. Based on the learning design information provided for the scenarios, participants will have an opportunity to explore whether they could use the integrated framework described in the working paper to operationalise the process of generating insights and feedback to teachers and learners for the scenario relevant to the design principles and targeted outcomes. An advantage of using the Ruby system to anchor the workshop activities is that it supports assessment designs that relate to the development of skills such as communication, empirical reasoning, or quantitative reasoning. The learning scenarios used in the workshop will explore a range of assessment task designs involving interactions with mock assessments such as projects, exhibitions, theses, or internships.

Once the groups have had a chance to evaluate how well the proposed integrated framework can be applied to these learning scenarios, the whole workshop group will reconvene (after a short break) in Part 4 to discuss and critique the proposed integrated framework. Participants will be encouraged to generate feedback on what works and what improvements could be made to strengthen the usefulness of the framework for LA practitioners, teachers, and students. To conclude the workshop, in Part 5 participants will be encouraged to generate ideas and identify collaborators for ongoing projects to evaluate the proposed framework within their own contexts or across contexts. This is to ensure the applicability and scalability of the integrated framework to a range of learning designs and educational contexts, and to demonstrate how well (or not) the links of the bridge between LD and LA hold up when applied to various authentic scenarios. This approach aligns with calls in the LA literature for a greater focus on replication studies (Dawson et. al., 2019) and collaboration between experts to confirm and move the conceptualisation of foundational concepts forward for the benefit of the field as a whole.

4 WORKSHOP OUTCOMES

There will be several outcomes from this workshop for participants and the broader LA community. For workshop participants, they will gain an insight into the historical conversations and proposed frameworks that relate to the intersection between learning design and learning analytics. Their feedback and critique of the proposed integrated framework will be used to refine the framework so it can be presented to the LA field. This will be done through the submission of a revised version of the working paper to the Journal of Learning Analytics with acknowledgement given to all those who participated in the workshop discussions. The organisers will continue to engage with participants through the identified projects (see Part 5 above) to continue the evaluation of the framework across different contexts, maintaining communications and sharing of experiences between groups throughout the year. The outcomes of these projects will form the basis of the DesignLAK24 workshop, where project teams will be given an opportunity to showcase their findings and continue the conversation and evaluation of the design of the integrated framework. The overall goal is to help consolidate the conceptual and practical investigations, proposed frameworks, and technology tools to address challenges around how best to build the bridge to connect learning design and learning analytics.

REFERENCES

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic, D., Mulder, R., Williams, D., Dawson, S., Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. In *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (pp. 409-413). New York: ACM. https://doi.org/10.1145/2883851.2883944

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the Impact of Learning Analytics. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, (pp. 446–455). https://doi.org/10.1145/3303772.3303784

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

Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design. *Journal of Learning Analytics*, *7*(3), 98-117. https://dx.doi.org/10.18608/jla.2020.73.8

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516-534. https://doi.org/10.1109/TLT.2018.2868673

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., & Buckingham Shum, S. (2020). From data to insights: A layered storytelling approach for multimodal learning analytics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). https://doi.org/10.1145/3313831.3376148

Hi, LA! - Highly Informative Learning Analytics

Daniele Di Mitri
DIPF, Germany
d.dimitri@dipf.de

Marcel Schmitz
University of Applied Sciences Zuyd, Netherlands
marcel.schmitz@zuyd.nl

Ioana Jivet
Goethe University of Frankfurt, Germany
jivet@sd.uni-frankfurt.de

Sebastian Gombert
DIPF, Germany
s.gombert@dipf.de

Atezaz Ahmad
DIPF, Germany
a.ahmad@dipf.de

Hendrik Drachsler
DIPF, Germany
h.drachsler@dipf.de

ABSTRACT: In this workshop, we investigate the concept of highly informative learning analytics and propose a methodology for designing an environment that delivers highly informative learning analytics. The workshop is designed as a hands-on, interactive session that allows participants to test the methodology's potential in a realistic use case. The proposed approach is based on the four-stage process of the Design Cycle for Education (DC4E). We exemplify practical tools that were designed in-house for each stage, including a tool to support teachers while designing learning activities - the Fellowship of Learning Analytics (FoLA²), a learning analytics infrastructure integrated with Moodle - Edutex, and two Moodle plugins for learning activities that enable the collection of rich trace logs - Hyperchalk and the Concept Mapping Plugin. Finally, we discuss potential use cases that can be suitable for the methodology.

Keywords: collaborative design, structured approach, learning analytics indicators, feedback

1 BACKGROUND

Learning management systems such as Moodle distribute learning activities and materials among students (Dougiamas & Taylor, 2003). They also enable tracking of individual students' learning and can be a data source for learning analytics.

With this workshop proposal, we want to take this achievement even further and aim to develop *Highly-Informative Feedback with Learning Analytics* (HILA).

According to Hattie (2009), feedback has a powerful effect on learning success, with a mean effect size of d = 0.75. Wisniewski, Zierer and Hattie (2020) even report a mean effect of d = 0.99 for highly informative feedback (on right/wrong, correct solution, type of processing, possibilities for improvement, hints on self-regulation and learning strategies). Such feedback provides good conditions for self-directed learning (e.g. Winne & Hadwin, 2008) and effective metacognitive control of the learning process (Nelson & Narens, 1994). Until a few years ago, it was simply not possible in terms of personnel to provide highly informative and competence-oriented feedback at large university lectures. Nowadays, however, computers and other digital devices open up far-reaching possibilities that have not yet been fully exploited. This feedback has a high potential for improving individual study success and reducing dropouts, thus effectively supporting students in their learning process.

Within the HILA workshop, we will work along an emerging design and development process for Highly-Informative Learning Analytics based on various project experiences. We will first identify relevant LA indicators for different learning designs. From there, we will demonstrate how we turned the designs of learning into data-enriched learning activities that have the potential to provide highly-informative feedback. Finally, we discuss different types of feedback messages given to the students and future challenges for the HILA field.

2 PROPOSED SOLUTION

A team of researchers from different European universities have developed a *Highly Informative Learning Analytics* methodology. This process is based on the Design Cycle for Education (DC4E) model proposed by Scheffel et al. (2021). The methodology interprets the DC4E model more pragmatically: the first two quadrants in red and green (Identify & Combine phase) are accomplished by the FoLA^2 methodology for collaboratively designing LA-powered learning activities (Schmitz et al., 2022). The third quadrant (Realise phase) is accomplished with the Edutex LA infrastructure (Ciordas-Hertel et al., 2021) and all the applications created within the Edutex framework, i.e. Hyperchalk, the Concept Mapping Tool, and Edutex Android. These applications can be used as instances of learning activities designed with FoLA^2. These applications generate a wealth of data stored/processed within the Edutex infrastructure. Finally, the fourth quadrant (The research phase) is achieved by defining the process data indicators, as Goldhammer et al. (2021) explained.

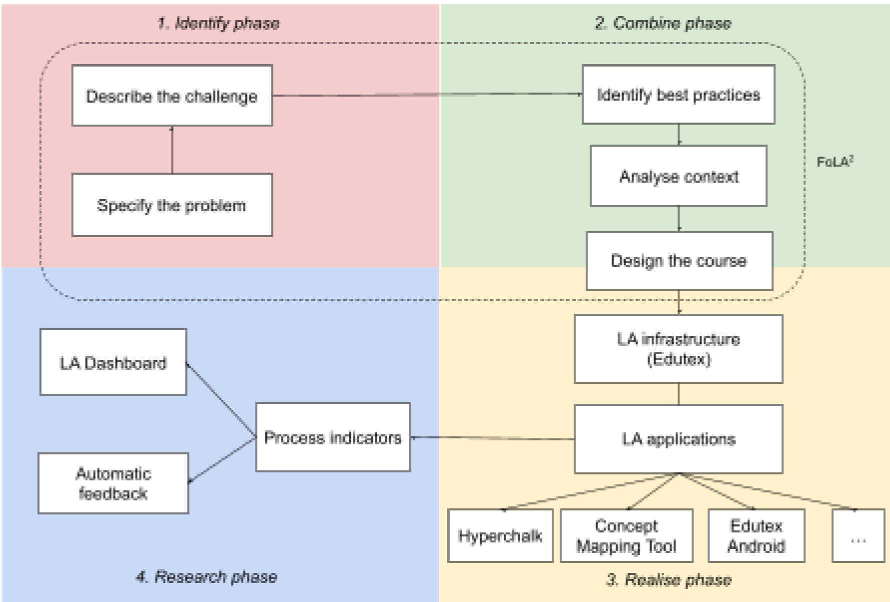


Figure 1 - The *Highly Informative Learning Analytics* process in a nutshell.

The Fellowship of Learning Activities and Analytics (FoLA^2) is a methodology for designing learning activities with "analytics in mind". We established a method — reinforced by a gameboard and cards, to provide structure and inspiration. Recently, it is also available with a digital version. The method enables several participants with different roles to collaboratively interact with a set of card decks to create an LA-supported learning design. Using this method helps to design learning activities collaboratively and practically; it also raises awareness about the benefits of multidisciplinary co-design and connections between learning analytics and learning design. FoLA^2 can be used to develop, capture, and systematize design elements and to incorporate LA systematically.

Edutex is a Learning Analytics infrastructure integrated with Moodle Learning Management System. Edutex is a context-aware learning analytics architecture that evolved at the beginning of the pandemic to discover more about the physical context of learners in distance education without losing track of data protection. EduTex has been designed to support learners in their preferred physical learning environment and in their individual home learning processes. With the help of adaptive interventions and integration of commodity Android smartphones and smartwatches, Edutex allows integrating their sensor data with questionnaire data obtained on the devices with learning management system data.

HyperChalk (Menzel et al., 2022) is a digital collaborative whiteboard built using the open-source component Excalidraw and a custom back-end. It was awarded the best demo award at EC-TEL 2022. The software can be self-hosted, collects rich log data appropriate for learning analytics purposes, and integrates with learning management systems – such as Moodle – using the LTI 1.3 Advantage standard. Similar to commercial whiteboard software such as Miro, Hyperchalk can be used to implement a wide range of creative collaboration tasks, but unlike commercial software, it allows researchers unlimited access to user data. It collects data appropriate for qualitative and quantitative studies on user behaviour demonstrated during collaboration tasks. Through a replay mode, the collaboration processes can be closely studied. Moreover, we plan to implement a corresponding annotation tool that can be used to code what is happening on the whiteboards explicitly.

The Concept Mapping Tool comes from a Moodle plugin which implements concept mapping as a learning activity. Concept maps visualize the relationships between concepts in a given domain via a network-like graph. They were introduced by Joseph D. Novak as means to represent students' developing knowledge (Nesbit & Adesope, 2006). By letting learners construct their own domain models through concept maps, one can gain insight into how they view a domain and think its ontologies are structured. This can provide researchers with insights into the structure of students' knowledge and also helps to identify misconceptions which can be used to provide personalized, targeted feedback.

3 Objectives

In this workshop, we investigate the concept of *highly informative learning analytics*. The workshop is thought of as a hands-on, interactive session. We plan to demonstrate the proposed LA cycle in this workshop and allow the participants a hands-on experience. The workshop activities are divided as follows:

in the morning - Part 1:

- Welcome and initial remarks
- A discussion of a representative task with the FoLA^2 methodology. The participants are divided into groups, each group is given a FoLA^2 board with which they need to design the learning session choosing among a set of available activities.

coffee break

- Each group presents their resulting designed sessions with the chosen design elements.

- The groups engage in a discussion in which they map the chosen activities with a set of existing tools.

in the afternoon - Part 2:

- The participants explore the collaborative whiteboard tool Hyperchalk and the Collaborative concept mapping tool.
- How to define the right process data indicators from the learning activities

coffee break

- Groups discuss
- A tour of the existing application use cases using the proposed process

4 Organisation

The HILA workshop is organised as an interactive, full-day workshop. For the logistics, we need a large room (30 participants) with a reliable internet connection, projector, separate tables for group exercises, and, if possible, stationery such as sticky notes and pens. The organisers will provide technical tools and Slack channels and disseminate progress and outcomes via blogs and the Twitter hashtag #HiLA23.

5 References

Brusilovsky, P. (2012). Adaptive Hypermedia for Education and Training. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive Technologies for Training and Education* (1st ed., pp. 46–66). Cambridge University Press. https://doi.org/10.1017/CBO9781139049580.006

Ciordas-Hertel, G.-P.; Rödling, S.; Schneider, J.; Di Mitri, D.; Weidlich, J.; Drachsler, H. Mobile Sensing with Smart Wearables of the Physical Context of Distance Learning Students to Consider Its Effects on Learning. Sensors 2021, 21, 6649. https://doi.org/10.3390/s21196649

Dougiamas, M., & Taylor, P. (2003). *Moodle: Using Learning Communities to Create an Open Source Course Management System*. 171–178. https://www.learntechlib.org/primary/p/13739/

Goldhammer, F., Hahnel, C., Kroehne, U., & Zehner, F. (2021). From byproduct to design factor: On validating the interpretation of process indicators based on log data. *Large-Scale Assessments in Education*, *9*(1), 20. https://doi.org/10.1186/s40536-021-00113-5

Hattie, J., & Timperley, H. (2007). The power of feedback. [References]. *Review of Educational Research*, *.77*(1), 16–17. https://doi.org/10.3102/003465430298487

Menzel, L., Gombert, S., Di Mitri, D., & Drachsler, H. (2022). Superpowers in the Classroom: Hyperchalk is an Online Whiteboard for Learning Analytics Data Collection. In I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (pp. 463–469). Springer International Publishing. https://doi.org/10.1007/978-3-031-16290-9_37

Nelson, T. O., & Narens, L. (1994). Why investigate Metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), Metacognition. Knowing about Knowing. (pp. 1-25). Cambridge, Mass: Cambridge University Press.

Nesbit, J. C., & Adesope, O. O. (2006). Learning With Concept and Knowledge Maps: A Meta-Analysis. *Review of Educational Research*, *76*(3), 413–448. https://doi.org/10.3102/00346543076003413

Scheffel, M., Schmitz, M., van Hooijdonk, J., van Limbeek, E., Kockelkoren, C., Joppe, D., & Drachsler, H. (2021). *The Design Cycle for Education (DC4E)*. Gesellschaft für Informatik e.V. http://dl.gi.de/handle/20.500.12116/37016

Schmitz, M., Scheffel, M., Bemelmans, R., & Drachsler, H. (2022). FoLA2 — A Method for Co-creating Learning Analytics–Supported Learning Design. Journal of Learning Analytics, 9(2), 265-281. https://doi.org/10.18608/jla.2022.7643

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10. 1-14. doi: 10.3389/fpsyg.2019.03087

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), Motivation and self-regulated learning: Theory, research, and applications (pp. 297–314). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

6 Acknowledgements

Part of this work was funded by the Hessian Ministry for Science and the Arts (HMWK) as part of the project HessenHub – Netzwerk digitale Hochschullehre Hessen, Trusted Learning Analytics.