Year: 2022

Examining pedagogical data literacy: results of a survey among school teachers at upper secondary level in Switzerland

Michos, Konstantinos ; Petko, Dominik

Abstract: One of the main concerns of learning analytics studies that involve teachers is the concept of data literacy. Pedagogical data literacy includes collecting and analyzing student data and deriving educational interventions, the success of which is again assessed using data. In this paper, we examine this construct to understand pedagogical data literacy levels within a representative survey of N = 1059 teachers in upper secondary schools in Switzerland. Preliminary results reveal that more than half of the secondary school teachers indicate having access to digital student data but only one-third is making use of this data to inform their teaching. Only one-fourth of the teachers indicate that they consider themselves proficient in using digital student data to improve their teaching. In-depth analyses will explore different conditions and characteristics that are associated with different teacher profiles of pedagogical data literacy

Originally published at:
Michos, Konstantinos; Petko, Dominik (2022). Examining pedagogical data literacy: results of a survey among school teachers at upper secondary level in Switzerland. In: LAK22: The 12th International Learning Analytics and Knowledge Conference, Online, 21 March 2022 - 25 March 2022. ACM + SIG Web, 79-81.

Companion Proceedings of the 12th International Conference on Learning Analytics & Knowledge LAK22

Learning Analytics for Transition, Disruption and Social Change

March 21-25, 2022
Online,
Everywhere

Organized by:

LAK22 Program Chairs' Welcome

We are very pleased to welcome you to the Twelfth International Conference on Learning Analytics and Knowledge (LAK22), organized by the Society for Learning Analytics Research (SoLAR). This year's conference, while originally planned to be hosted by University of California, Irvine at the Newport Beach Marriott, is held virtually March 21-25, 2022 in an effort to protect the LAK community from the ongoing COVID-19 pandemic.

The theme for the 12th annual LAK conference is "Learning Analytics for Transition, Disruption and Social Change." This theme brings to the forefront both the dynamic world situation in which learning analytics now operate and the potential role of learning analytics as a driving force for change within it. In a moment when questions about transparency, fairness, equity and privacy of analytics are being raised in many areas of application, there is both an opportunity and an imperative to engage with these issues in support of ethical pedagogical transitions and transformative social justice. In addition, as LAK itself explores changing formats for knowledge exchange and generation, this theme offers the opportunity for reflection on how to make the conference more sustainable and accessible for people around the world. We have three excellent keynotes who will address this theme across the complementary lenses of education, data science and social change: Lorri J. Santamaría from California Lutheran University who will speak about how equity-driven and culturally sustaining leadership can enhance learning analytics, Catherine D'Ignazio from the Massachusetts Institute of Technology who will speak about how feminist thinking can help envision more ethical and equitable data practices, and Pierre Dillenbourg from the Swiss Federal Institute of Technology who will speak about the future of classroom analytics to smooth orchestration.

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 6 Practitioner Track Papers, 27 Posters, 4 Demos, 17 Workshops and 8 participants to the Doctoral Consortium, each of which is represented in this Companion Proceedings. In addition, we accepted 2 Tutorials to be held at the conference: Processing and Visualizing Clickstream Data Using R and Aligning Decision-Making Models with Curriculum Theory.

We would like to emphasize our ongoing gratitude for the efforts made by everyone involved in our community during these difficult COVID times. We recognize that as we move through the second full year of the pandemic that the students, researchers and staff in our community face continued physical and emotional challenges, including stress, uncertainty and fear. These are difficult times for us all and we want to thank each one of you for the important efforts you have devoted that have allowed this conference to continue as a scientific event and scholarly exchange of ideas of the highest caliber.

We hope that LAK22 participants and other readers of these companion proceedings will find value in the many varied contributions to the field of learning analytics contained within and also recognize their positionality with respect to the different interdisciplinary fields from which we draw. This includes contributions to theory, methods and practice as well as both basic and applied work. As we continue to navigate these challenging times, the power and potential for analytics to help us achieve deeper understandings of learning and create better support for students, teachers and other educational stakeholders motivate us to persevere in this important work.

Alyssa Friend Wise
New York University, USA

Roberto Martinez-Maldonado
Monash University, Australia

Isabel Hilliger
Pontificia Universidad Católica de Chile, Chile

LAK22 Conference Organization

Program Chairs
Alyssa Friend Wise, New York University, USA
Roberto Martinez-Maldonado, Monash University, Australia
Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile

Practitioner Track Chairs
Mingyu Feng, WestEd, USA
Yi-Shan Tsai, Monash University, Australia

Workshop and Tutorial Chairs
Alejandra Martínez-Monés, Universidad de Valladolid, Spain
Olga Viberg, KTH Royal Institute of Technology, Sweden
Rafael Ferreira Mello, Universidade Federal Rural de Pernambuco, Brazil

Posters and Demonstrations Chairs
Elizabeth Koh, Nanyang Technology University, Singapore
Shayan Doroudi, University of California, Irvine, USA

Doctoral Consortium Chairs
Michael Brown, Iowa State University, USA
Dragan Gašević, Monash University, Australia
Oleksandra Poquet, University of South Australia, Australia
Jelena Jovanovic, University of Belgrade, Serbia

Proceedings Chairs
Hassan Khosravi, The University of Queensland, Australia
Abhinava Barthakur, University of South Australia, Australia

Publicity Chair
Yizhou Fan, University of Edinburgh, UK

Website Manager
Yancy Vance Paredes, Arizona State University, USA

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

LAK22 Companion Proceedings Committee

Practitioner Track Program Committee
Samantha Ahern, University College London, United Kingdom
Alan Berg, University of Amsterdam, Netherlands
Alex Bowers, Columbia University, USA
Justin T. Dellinger, University of Texas at Arlington, USA
Stephanie Frost, University of Saskatchewan, Canada
Samuel Greiff, University of Luxembourg, Luxembourg
Wolfgang Greller, Vienna University of Education, Austria
Dai Griffiths, The Universidad Internacional de La Rioja, Spain
Maureen Guarcello, San Diego State University, USA
Christothea Herodotou, The Open University, United Kingdom
Martin Hlosta, The Open University, United Kingdom
Hazel Jones, Griffith University, Australia
Danny Liu, The University of Sydney, Australia
Leah Macfadyen, The University of British Columbia, Canada
Rachel Niemer, University of Michigan, USA
Rob Nyland, Air University, USA
Joshua Quick, Indiana University Bloomington, USA
Mike Sharkey, Data & Graphs, USA
Kp Thai, Age of Learning, USA
Craig Thompson, The University of British Columbia, Canada
Steven Williams, University of California, Berkeley, USA

Posters & Demo Program Committee
June Ahn, University of California, Irvine, USA
Farhan Ali, National Institute of Education, Singapore
Ainhoa Alvarez, University of the Basque Country, Spain
Ishari Amarasinghe, Universitat Pompeu Fabra, Spain
Kimberly Arnold, University of Wisconsin-Madison, USA
Juan Asensio-Pérez, Universidad de Valladolid, Spain
Roger Azevedo, University of Central Florida, USA
Ryan Baker, University of Pennsylvania, USA
Alan Berg, University of Amsterdam, Netherlands
Matthew Bernacki, The University of North Carolina at Chapel Hill, USA
Nigel Bosch, University of Illinois Urbana-Champaign, USA
Fabio Campos, New York University, USA
Guanliang Chen, Monash University, Australia
Heeryung Choi, University of Michigan, USA
Linda Corrin, Swinburne University of Technology, Australia
Mutlu Cukurova, University College London, United Kingdom
Yannis Dimitriadis, University of Valladolid, Spain
Vanessa Echeverria, Escuela Superior Politécnica del Litoral, Ecuador
Yizhou Fan, The University of Edinburgh, United Kingdom
Elaine Farrow, The University of Edinburgh, United Kingdom
Gloria Milena Fernandez-Nieto, University of Technology, Sydney, Australia
Rafael Ferreira Mello, Federal Rural University of Pernambuco, Brazil
Ed Foster, Nottingham Trent University, United Kingdom
Stephanie Frost, University of Saskatchewan, Canada
Andrew Gibson, Queensland University of Technology, Australia
Geraldine Gray, Technological University Dublin, Ireland
Julio Guerra, University of Pittsburgh, USA

Ella Haig, University of Portsmouth, United Kingdom
Areum Han, University of Luxembourg, Luxembourg
Martin Hlosta, The Open University, United Kingdom
Xiao Hu, The University of Hong Kong, Hong Kong
Srecko Joksimovic, University of South Australia, Australia
Jelena Jovanovic, University of Belgrade, Serbia
Yeonji Jung, New York University, USA
Ean Teng Khor, Nanyang Technological University, Singapore
Rene Kizilcec, Cornell University, USA
Malgorzata Korolkiewicz, University of South Australia, Australia
Vitomir Kovanovic, University of South Australia, Australia
Elise Lavoué, Université Jean Moulin Lyon 3, France
Alwyn Vwen Yen Lee, Nanyang Technological University, Singapore
Lisa Lim, University of Technology Sydney, Australia
Yiwen Lin, University of Michigan, USA
Mirko Marras, University of Cagliari, Italy
Alejandra Martínez-Monés, Universidad de Valladolid, Spain
Agathe Merceron, Berliner Hochschule für Technik, Germany
Sandra Milligan, The University of Melbourne, Australia
Caitlin Mills, University of New Hampshire, Canada
Negin Mirriahi, University of South Australia, Australia
Pedro Manuel Moreno-Marcos, Universidad Carlos III de Madrid, Spain
Kousuke Mouri, Tokyo University of Agriculture and Technology, Japan
Rachel Niemer, University of Michigan, USA
Xavier Ochoa, New York University, USA
Alejandro Ortega-Arranz, Universidad de Valladolid, Spain
Zacharoula Papamitsiou, SINTEF Digital, Norway
Luc Paquette, University of Illinois at Urbana-Champaign, USA
Ward Peeters, Kanda University of International Studies, Japan
Mar Perez-Sanagustin, Université Paul Sabatier Toulouse III, France
Héctor Javier Pijeira-Díaz, Maastricht University, Netherlands
Oleksandra Poquet, University of South Australia, Australia
Justin Reich, Massachusetts Institute of Technology, USA
Bart Rienties, The Open University, United Kingdom
Fernando Rodriguez, University of California, Irvine, USA
José A. Ruipérez Valiente, University of Murcia, Spain
Agnes Sandor, Naver Labs Europe, France
Mohammed Saqr, University of Eastern Finland, Finland
Juan Pablo Sarmiento, New York University, USA
Bruce Sherin, Northwestern University, USA
Atsushi Shimada, Kyushu University, Japan
Zachari Swiecki, Monash University, USA
Michelle Taub, University of Central Florida, USA
Craig Thompson, The University of British Columbia, Canada
Stefan Trausan-Matu, University Politehnica of Bucharest, Romania
Yi-Shan Tsai, Monash University, Australia
Cristina Villa-Torrano, Universidad de Valladolid, Spain
April Wang, University of Michigan, USA
Elle Yuan Wang, Arizona State University, USA
Phil Winne, Simon Fraser University, Canada
Masanori Yamada, Kyushu University, Japan
Gaoxia Zhu, Nanyang Technological University, Singapore

Table of Contents

Doctoral Consortium

Workshops

Procrastination is a key indicator of success or failure in some online business courses

Brent Benson
Harvard Business School Online
bbenson@hbs.edu

ABSTRACT: A strong correlation between procrastination and key course outcomes like completion and course grade has been observed across a study of over 50,000 students and over 200 course offerings of mostly-asynchronous online business courses. These relationships hold across many different cohorts of students across multiple years. Previous published work on procrastination has not shown such a definitive effect. An effort is made to identify reasons for the strong correlation based on pedagogy and course design and ideas are presented on how procrastination metrics might be surfaced to students in an attempt to improve course completion and grade outcomes.

Keywords: Procrastination, Academic Performance, Self-Regulated Learning, Learning Analytics

1 PROCRASTINATION OVERVIEW

Is procrastination—in the form of delaying completion of an academic assignment until close to a deadline—detrimental to academic performance? Previous research has been somewhat equivocal on this question, and the answer may depend on learning context, measurement technique, and student cohort.

A meta-analysis of 33 quantitative procrastination studies (Kim & Seo, 2015) did not find a consistent relationship between procrastination and outcome, but it did find a measurable negative correlation between procrastination and performance in contexts where the both procrastination and performance measures were externally assessed rather than self-reported.

With this in mind, an effort was made to determine the role of procrastination on student outcomes in a large historical group of online business course offerings.

2 LEARNING CONTEXT AND MEASUREMENT

The online business courses considered in this research (*Online Business Courses & Certifications | HBS Online*, 2021) are targeted at adults (college age and up) and built around modules that are open between one and three weeks with strict completion deadlines.

While these courses are asynchronous—students can login and work on them whenever they want—all students in a course have the same module deadlines providing a regular course cadence. All of the module material must be completed by the module deadline and access to upcoming material is restricted until the student completes the previous module.

The tasks to be accomplished in a module before the deadline are based on an active, social, case-based pedagogy and include reading expository material, videos, interactive simulations, short pieces of reflective writing, social interaction with the cohort around reflections, and a module quiz.

The precise (anti-)procrastination metric for the purposes of this research is *the percentage of active module time the student spends before the last 24 hours approaching the module deadline*. Higher values mean more work was completed before the last day with less procrastination. The choice of metric and number of hours is discussed in a later section.

Most online learning platforms provide analytics support for measuring the amount of time students spend on learning activities with varying levels of precision (Kovanovic *et al.*, 2015; Nguyen, 2020). The LMS which is the subject of this research has a particularly fine-grained time measurement subsystem that uses a signal sent every 60 seconds to understand whether the student is active and on which element of the course they are working.

This time-spent data at per-second granularity can be very large and unwieldy to work with across hundreds of course offerings and thousands of students. In order to make this data easier to work with, a data table was created that aggregates time-on-platform information by hour-before-deadline. The table holds the number of seconds the student spent in each hour leading up to the module deadline. The representation is sparse, meaning that if the student spent no time in a particular hour, the record is not stored.

3 SAMPLE POPULATION

The data used for this research looks at 57,151 students across 201 HBS Online program offerings from 2017 through the beginning of 2021. Thirty-five of those offerings were instances of the Credential Of Readiness (CORe) program—a credential program with higher stakes grading and a final exam—and the remaining 166 offerings were for 12 certificate programs where grades are pass/fail and participants are granted a certificate based on satisfactory and on-time completion of the material.

Students in the CORe program are graded based on a combination of on-time completion, final exam grades, quiz grades, and a measure of participation. The final grade for CORe is one of *Fail*, *Pass*, *Honors*, or *High Honors*. This final grade is the performance measure used for CORe offerings.

The 12 certificate/pass-fail programs in this research offer a certificate of completion if the course materials were satisfactorily completed within the required deadlines. Completion is used as the performance measure for these certificate courses.

4 PROCRASTINATION VS. OUTCOME

First, a look at the relationship between the procrastination metric and Credential Of Readiness (CORe) final grade. This summary is across 35 offerings of CORe from 2017 through early 2021.

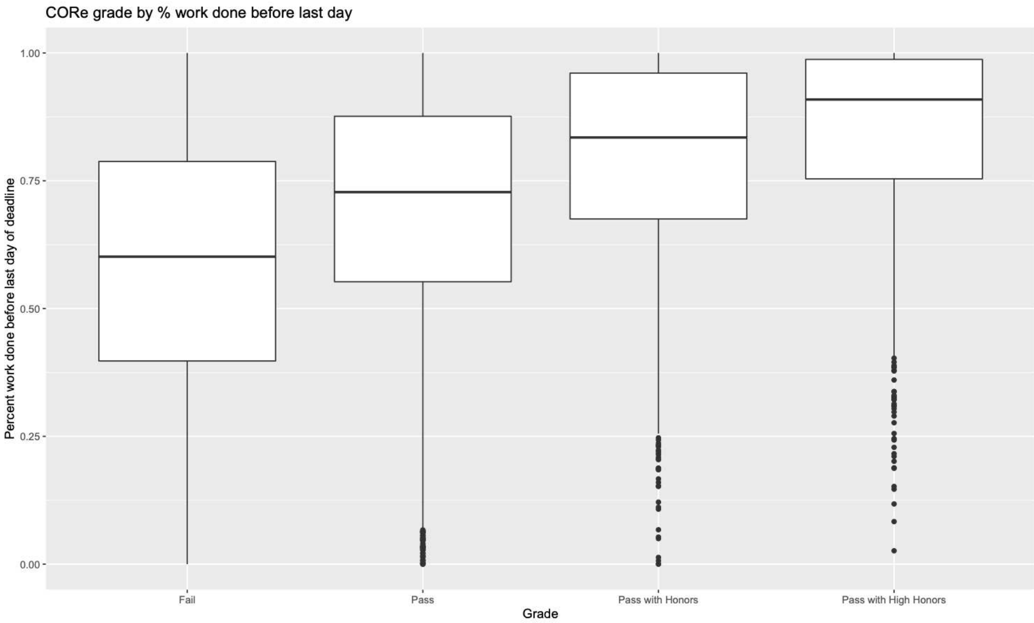


Figure 1: Procrastination by grade

The relationship between the percentage of work done before the last 24 hours and the grade is clear from Figure 1. The median procrastination metric value (the horizontal line in the middle of the boxplot box) for the grades are *Fail*: 60%, *Pass*: 73%, *Pass with Honors*: 84%, and *Pass with High Honors:* 91%. While the tendency is clear, there are a some exceptional students achieving passing grades while exhibiting evidence of procrastination, deserving deeper examination.

The cutoff value of 24 hours was chosen by creating a logistic regression model between the procrastination metric as the exogenous variable and the grade outcome as the endogenous variable. A separate logistic regression model was created using a procrastination metric variable that was calculated for each of the hourly cutoff values of 1 through 40. The regression models were compared using the Akaike Information Criterion (AIC) values of the models. The 24-hour model minimized AIC and prediction error, and thus became our cutoff value.

Alternative procrastination metrics were considered, including the measure of *median hour* of module work where half of the work was done before the median hour, and half of it after. This procrastination metric was less predictive in the logistic regression models on both CORe grade and certificate course completion outcome. The amount of work concentrated at the very end of the module deadline turns out to be more correlated with student outcomes than this measure of central tendency.

It may be that setting a cutoff time before a deadline and calculating the percentage of work done before and after the deadline can be useful as a procrastination measure in other learning contexts, although the cutoff may need to be adjusted based on platform and course particulars.

The next visualization (Figure 2) considers the same 24-hour procrastination metric, but looks at the average metric value for each grade within each offering of CORe from 2017 through early 2021.

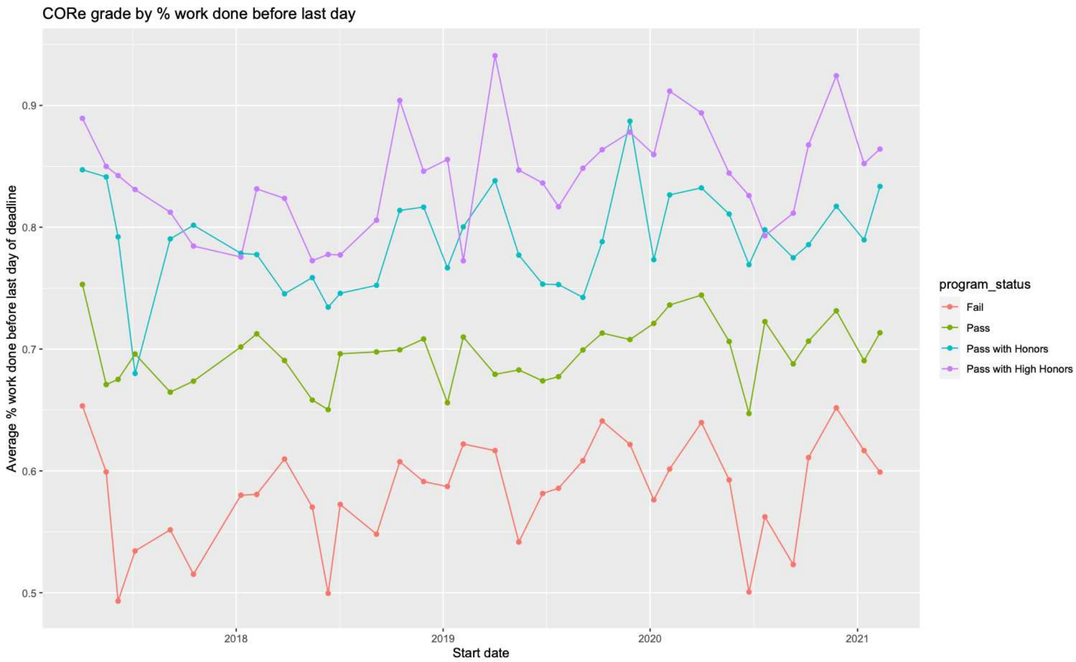


Figure 2: Procrastination by grade over time/offering

While there are several CORe offerings where the average values for *Pass with Honors* and *Pass with High Honors* are close, or even cross each other, and one case where the average value for *Pass* is slightly higher than *Pass with Honors*, the overall situation is clear: on average, students who complete more of their work before the deadline day are more likely to get a better grade.

We see a very similar pattern with respect to completion likelihood in certificate courses (Figure 3). These data are from 166 certificate offerings from 2017 through early 2021. The median procrastination value for *Did not Complete* students was 69%, while *Completed* students had a median procrastination metric of 89%.

This relationship holds across years, courses, and offerings. The two outlier cases are both first time offerings—June 2018 Sustainable Business Strategy and May 2019 Leadership Principles. The first offering of a new course is often small and is usually filled with motivated students and very few non-completers. The June 2019 SBS offering had only one student out of 179 that did not complete.

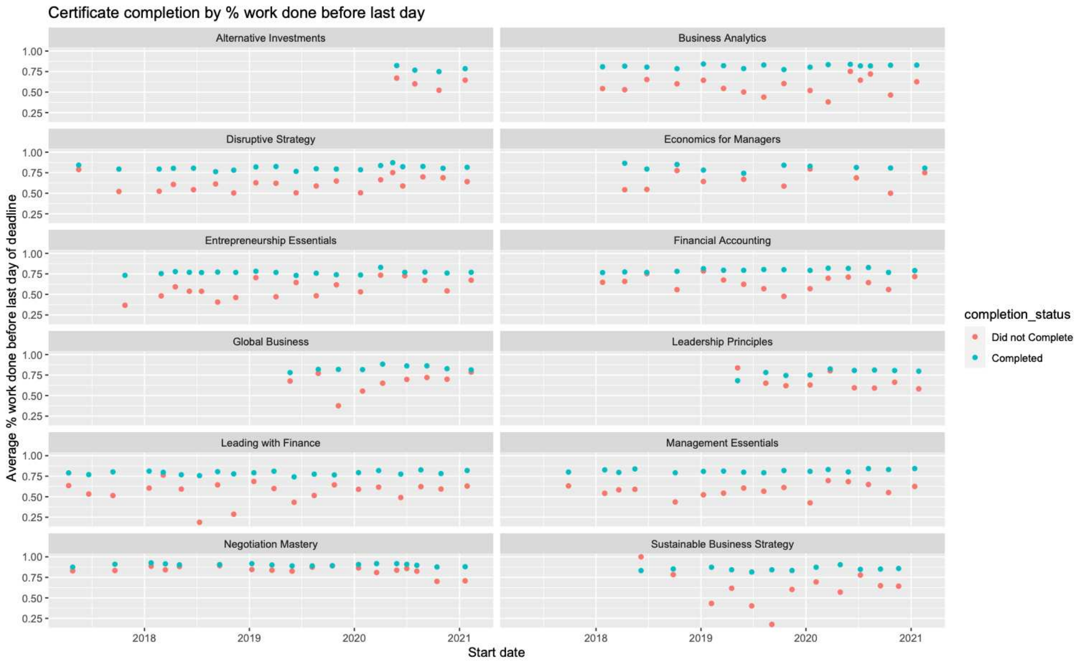


Figure 3: Procrastination by completion status within course and offering

5 WHY THE CONSISTENCY?

The literature has mixed findings on procrastination with some studies showing negative effects for procrastination, while others show neutral or even some positive effects (Kim and Seo, 2015). The strength and consistency of the relationship between procrastination and outcomes across both the credential and certificate programs in this research raises the question of why this would be so consistent across so many different courses, students, and time periods.

The answer may lie in the consistency of the courses' active, social, and case-based pedagogy (Benson & Houtti, 2022). The intent of the pedagogy is to create the type of engagement found in in-person case-based discussions (Ellet, 2007). This teaching method is akin to interactive storytelling, with real-world situations being presented, discussion of course material in relation to those cases, and the opportunity and need to engage with other students around the material. The requirements in these courses to consider, reflect, and interact around the subject matter may not lend itself to fast-forwarding, whether due to active/purposeful procrastination or passive/unwanted work delay.

6 INTERVENTIONS

There appears to be an opportunity to use these procrastination measurements to improve completion rates and grades of students in the platform. One experiment underway is to send out-of-platform email messages to CORe students with messages from previous students with advice on how to do well in CORe. These messages invariably warn against procrastination, and also include information on other recommended techniques like forming peer study groups.

There is also consideration being given to a platform enhancement that would provide students time management information, showing on the dashboard how much time they have spent before the deadline compared to previously successful students, and encouraging them to set goals of completing more work in a module before the critical last day. This is related to work described in (Molenaar *et al.*, 2020) and (Pesonen, Palo-oja and Ng, 2021).

7 CONCLUSION

It is clear that procrastination, or lack thereof, is a key indicator of success or failure in these mostly-asynchronous credential and certificate online business courses. Additional research will be required to better understand why procrastination plays this key role. An important follow-up to this work will be an attempt to put this information into students' hands in an actionable way that improves their learning outcomes and allows them to fully achieve their learning goals.

8 ACKNOWLEDGEMENTS

This work was initially inspired by a visualization similar to Figure 2 done on early data from the CORe program by Jan Hammond (Harvard Business School) and Brian Keegan (University of Colorado at Boulder) in 2015. The hourly data representation was developed with Mohamed Houtti (University of Minnesota) over the summer of 2021.

REFERENCES

Benson, B., & Houtti, M. (2022). Developing Social Interaction Metrics for an Active, Social, and Case-Based Online Learning Platform. In *Social and Emotional Learning—An Inclusive Learning Analytics Perspective*. Springer Nature.

Ellet, W. (2007). *The case study handbook: How to read, discuss, and write persuasively about cases*. Harvard Business School Press.

Kim, K. R., & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, *82*, 26–33. https://doi.org/10.1016/j.paid.2015.02.038

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2015). Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. *Journal of Learning Analytics*, *2*(3), 81–110. https://doi.org/10.18608/jla.2015.23.6

Molenaar, I., Horvers, A., Dijkstra, R., & Baker, R. S. (2020). Personalized visualizations to promote young learners' SRL: The learning path app. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 330–339. https://doi.org/10.1145/3375462.3375465

Nguyen, Q. (2020). Rethinking time-on-task estimation with outlier detection accounting for individual, time, and task differences. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 376–381. https://doi.org/10.1145/3375462.3375538

Online Business Courses & Certifications | HBS Online. (2021). https://online.hbs.edu/courses/

Pesonen, J., Palo-oja, O.-M., & Ng, K. (2021). Personalized Learning Analytics-Based Feedback on a Self-Paced Online Course. *Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge (LAK'21)*, 13–14.

Tutors' deliberations for implementing a Learning Analytics Dashboard in the Educational Context

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

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

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

Anouk Pouwels
Department of Educational Sciences, Utrecht University, The Netherlands
A.pouwels@uu.nl

Marieke van der Schaaf
Utrecht Centre for Research and Development of Health Professions Education, University Medical Centre Utrecht, The Netherlands
M.f.vanderschaaf-5@umcutrecht.nl

Presentation abstract: Designing a Learning Analytics Dashboard (LAD) to provide students with insight and actionable feedback should be done with involvement of relevant stakeholders (e.g.) students, tutors, tutor program coordinators, and study advisors. However, even then effective uptake is not assured, as the implementation into the educational context is a distinct part of the process. This paper describes how interviews and a focus group were held to map an implementation process, and how to support implemention of a LAD into an educational context. Results showed that implementation decisions were not always based on pedagogical principles and were sometimes practical in nature. As more elaborate deliberations could positively affect implementation effectiveness, a roadmap was designed to support the implementation process. In it, tutors answer 'what' and 'why' questions, aimed to deepen the choices at phases of the implementation.

Keywords: Learning Analytics Dashboard, Study behavior, Implementation, Tutor

PROJECT BACKGROUND

Learning analytics can be used to monitor study motivation and progress and to provide just-in-time feedback. This paper presents first implementation findings of the Thermos project, which aims to support Higher Education students by providing insight and actionable feedback on aspects of study motivation and study behavior via a learning analytics dashboard (LAD). In it, Motivation and Engagement (MES) (Martin, 2007) and groupwork skills (GSQ) (Cumming et al., 2015) are self-assessed and visualized. Per construct, actionable feedback is presented (cf. Hattie & Timperly, 2007) (see Figure 1). This dashboard was developed through multiple design iterations, in collaboration with multiple stakeholders (De Vreugd et al., 2021). However, simply designing a LAD is no guarantee for successful uptake (Wise & Vytasek, 2017). In case of Thermos, the most important stakeholders are students, tutors, and tutor program coordinators. While students can access the dashboard, tutors stimulate students to reflect on their study behavior by (e.g.) advocating use of the dashboard. This report focusses on: 1) how tutors shape the dashboard implementation process and the underlying motives, and 2) supporting this process by an implementation roadmap.

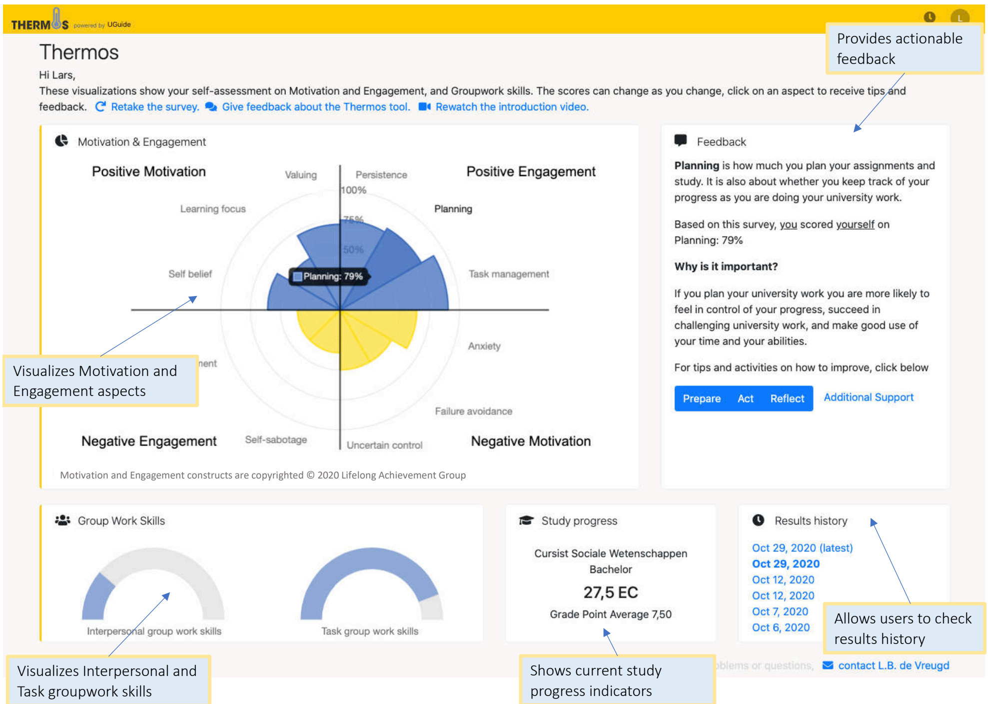


Figure 1. The Thermos-dashboard.

METHODOLOGY - MAPPING THE IMPLEMENTATION PROCESS

In academic year 20/21, the Thermos dashboard was introduced in six study programs from three different faculties at Utrecht University. Approximately 350 students engaged with the dashboard. How it would be introduced to and used by students was determined in collaboration with the tutor program coordinator. Study programs differed with regard to (i.a.) magnitude of the tutor program, student population (alpha or beta), and choices regarding dashboard implementation. The implementation process was mapped and summarized by conducting semi-structured interviews with tutors and tutor program coordinators afterwards. Then, a focus group with the same participants was held to explore overlap and differences in implementation deliberations between tutor programs.

Interviews: methodology and results

Semi-structured interviews were held with a representative of every participating study program. The goal was determining what implementation choices were made ('what') and to explore the underlying motives ('why'). The interview protocol was based on the Align Design Framework (Wise et al., 2016), adjusted and supplemented to match the specific implementation context (Figure 2, phases 1 to 5). For instance, for 'Integration', the question was: 'How was the dashboard integrated in the bachelor program, and why in this way?'. Data was analyzed using descriptive meta-matrixes per tutor program. To investigate overlap and differences, two between-case displays were made – one for the 'what', one for the 'why'. As expected, tutor programs differed in the extent to which Thermos was integrated. Several tutor programs used the dashboard's outcome as the starting point for an existing self-reflection exercise. Other study programs made dashboard use voluntary and introduced the dashboard as a way for students to gain insight into their study behavior. Motives for implementation choices were often practical in nature (e.g., a meeting concerning study reflection was already planned). However, while the main motive appeared practicality, participants could not explain their motive for many implementation choices. Pedagogical principles were often absent in the mentioned reasons for particular choices. Participants also learned of new possibilities for implementation during the interview (e.g. when asked if peer discussion was facilitated, some participants indicated this was not thought of – and therefore not facilitated). More elaborate consideration of possible pedagogical uses of the dashboard could improve the implementation and thereby its usefulness to students. These findings led to the preliminary conclusion that support is needed to make deliberate implementation decisions, and by extension to ensure successful uptake of the dashboard.

Focus group: methodology and results

To thoroughly understand tutors' implementation deliberations and to collectively determine which implementation aspects are most important, a focus group was organized. Tutor program coordinators discussed 1) what implementation aspects they deemed most important, and 2) what future study programs can do to achieve successful implementation. Five tutor program coordinators (one per study program) participated (a tutor of the sixth study program was unable to attend). Input for the focus group were nine statements that described crucial implementation aspects for successful uptake, distilled from the interviews and LA literature (e.g. 'When implementing the Thermos dashboard, it's important to think about connecting the tutor program's goals to Thermos'). Participants first ranked the statements in order of importance for successful implementation. Next, the group was asked to collaboratively come up with suggestions for how to respond to the three highest rated statement. For example, for the statement that it is desirable that students compare outcomes with peers, the group suggested to actively facilitate this comparison in tutor meetings.

The ranking process showed that alignment of dashboard implementation to the tutor program's goals was deemed most important. Suggestions to achieve this revolved around making the connection between the tutor program goals and the dashboard explicit, by writing it down or verbalizing it within the team (e.g. 'a goal of our study program is stimulating self-reflection, the dashboard can help by offering concrete aspects to reflect upon'). Second in importance was the introduction of the dashboard in a manner that motivates students. A suggestion for accomplishing student motivation was by having 'ambassadors' (tutors and/or students) introduce the dashboard and making it extremely clear why dashboard use is beneficial for students. Third in importance was consciously determining when students would use the dashboard. Arguments were that if introduced too early, students may not have enough experience to reliably self-assess their study behavior (e.g. failure avoidance) (see Figure 1). Also, students can feel overwhelmed at the start of their study, so reflecting on study behavior may not fit that stage. However, it was also argued that introducing the dashboard in an early stage of the study program may set a standard towards students - by showing that using this dashboard and self-reflecting are part of being a student.

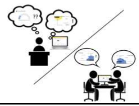
0. Goals	1: First use	2: Communication	3: Student use	4: Reflection & Action	5: Follow-up
					
What is the aim of implementing Thermos?	What study year(s) will use the dashboard? In what semester?	By whom will students be informed about the dashboard?	How will the dashboard be embedded in the study program?	Will students be able to compare results with peers?	How many moments of use are planned?
What is the reason for this aim?	Why at this moment?	Why that person(s)?	Why this form of integration?	Why (not) and why this way?	Why this number of moments?

Figure 2. Phases of the roadmap with example questions per phase.

CONCLUSION – DESIGN OF IMPLEMENTATION ROADMAP

Effectively implementing a learning analytics dashboard into an educational context in a meaningful way may not be a process that always occurs naturally. Results from the interviews and focus group showed the need for implementation support. This could help tutor program coordinators make more deliberate implementation decisions, leading to a more tight-knit integration of the dashboard in the curriculum. Which, in turn, is essential for successful uptake of the dashboard by students. Practically, such support can be offered by asking 'what' and 'why' questions when preparing the implementation and by co-designing the implementation with stakeholders. For Thermos specifically, the interviews and focus group have led to the design of an implementation roadmap for study programs. Figure 2 shows an adaptation of part of the roadmap. It was used in preparation of the current academic year (21/22). The roadmap poses 'what' and 'why' questions, which are supposed to be answered by the tutor coordinator in collaboration with colleagues. This roadmap thereby aims to support conscious and deliberate decision making.

REFERENCES

Cumming, J., Woodcock, C., Cooley, S. J., Holland, M. J., & Burns, V. E. (2015). Development and validation of the groupwork skills questionnaire (GSQ) for higher education. *Assessment & Evaluation in Higher Education*, *40*(7), 988-1001. doi: 10.1080/02602938.2014.957642

De Vreugd, Jansen, van Leeuwen, Sosnovsky, & Van der Schaaf (2021). Supporting Study Behavior with the Thermos Learning Analytics Dashboard. In *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21).*

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. doi: 10.3102/003465430298487

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, 77(2), 413-440. doi: 10.1348/000709906X118036

Wise, A. F., Vytasek, J. M., Hausknecht, S., & Zhao, Y. (2016). Developing Learning Analytics Design Knowledge in the" Middle Space": The Student Tuning Model and Align Design Framework for Learning Analytics Use. *Online Learning*, *20*(2), 155-182.

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of learning analytics* (pp. 151 – 160). New York, NY, USA: SOLAR, Society for Learning Analytics and Research. DOI: 10.18608/hla17.013

Adoption and usage challenges of a Learning Analytics Dashboard for game-based learning: design and implementation implications

Andrea Gauthier, Manolis Mavrikis, Laura Benton, Mina Vasalou
UCL Knowledge Lab, Institute of Education, University College London
[andrea.gauthier, m.mavrikis, l.benton, a.vasalou]@ucl.ac.uk

ABSTRACT: We report challenges associated with the adoption and (mis)use of a teacher-facing learning analytics dashboard (LAD) created for data-informed decision-making in primary school literacy education. We developed a LAD to facilitate teachers' planning of game-based literacy learning activities, identification of children requiring additional support, and self-evaluation of teaching practices. Two teachers had access to the LAD over three months. However, it was not used as expected. One teacher did not access the LAD, whilst the other attempted to use the LAD for summative assessment purposes, rather than formatively, which raised issues for the interpretation of data. This paper makes a unique contribution to the LA community by reflecting on the implications of the challenges we faced in terms of the LAD's failed adoption and misuse for future LAD design and implementation in the classroom.

Keywords: learning analytics dashboard, game-based learning, implementation challenges

1 INTRODUCTION

With increased use of learning technologies in the classroom, comes opportunities to utilise learning analytics (LA) to drive pedagogical change and improve children's learning in a formative manner (Kovanovic et al., 2021). *Navigo* is an adaptive, tablet-based game designed to reinforce literacy learning in primary schools across a range of language categories, e.g., phonics, prefixes/suffixes, grammar. It has 16 mini-game mechanics and 900+ game activities that produce LA on what, when, and how well the children played. Co-design workshops with teachers generated pedagogical aims and initial design concepts for how LA might support data-informed decision-making in literacy teaching through *Navigo* using a LA Dashboard (LAD) (Vezzoli et al., 2020). The LAD design (Figure 1 and https://tinyurl.com/NavigoLAD) catered to five aims: (1) plan learning activities for common gaps in class, using the bar chart identifying "top areas that need work"; (2) decide when to move to next learning objective at a class level, using the line chart; (3) identify lack of student engagement, using the line graph charting progress over time; (4) assess students' strengths and weaknesses and plan individualized support, using the filtered view by student and the tree map; and (5) self-assess teaching practices, by combining information from all views. However, because education often places an emphasis on summative evaluation and less on formative assessment (Shute, 2008), teachers may not necessarily approach a LAD with a formative, decision-making mindset, which could limit the capacity of our LAD to be used as a feedback loop and contribute toward continual development.

2 CASE STUDY METHODOLOGY

We implemented the LAD in one school in England with two teachers who were using *Navigo*. Both teachers taught literacy to children aged 6-7 and used *Navigo* weekly for about 30 minutes. They also assigned *Navigo* activities for children to play at home during the COVID-19 lockdown. Whilst the LAD was only accessible from March-May 2021 (three months total), it displayed analytics collected for *Navigo*'s entire use period (Nov 2020 - May 2021). One teacher, T1, had a special interest in using data for learning and a data analytics background, whilst T2, was a more experienced teacher and very data-literate. T2 was also involved in the LAD's online design workshops. Prior to the LAD's implementation, we held an online teacher training session to show the different LAD functionalities. We purposefully took a hands-off approach to the implementation of the LAD in the classroom, allowing the teachers to use it as they would outside of a research context, apart from a few emails to check if there were any problems and respond to any teacher enquiries. The case study involved log data collection of when/how long the teachers used the LAD, as well as 1:1 interviews in June 2021. As a secondary data source, we include emails from T1 to the researchers and an online discussion (prior to the final interview), which informed us about their initial encounters with the LAD.

Figure 1. (a) *Navigo* LAD displaying language feature frequency and performance, which can be filtered by time, individual student, and category. (b) Hovering shows additional information.

3 CASE STUDY RESULTS

<u>Logged LAD usage</u>. There was **limited use of the LAD**. T1, used it four times during May 2021, for an average of 6.31 (SD=5.76) minutes per session, totaling 25.23 minutes. T2, did not use the LAD at all during the three-month trial. As such, our findings are largely based on T1's data.

<u>Contact point – email communication</u>. An email from T1 revealed that she **approached the LAD through a lens of summative assessment** because she had been asked by her Deputy Headteacher to provide any data or summary of literacy improvements of students using *Navigo*. This was an unexpected use of the LAD that deviated from the intended formative aims. Based on this request, T1 used the LAD to look for "significance", to summatively assess how children were performing and report back. During this process, T1 emailed the researchers to discuss her interpretations about her class's literacy performance in *Navigo*: "If I look at the 'competence level' parameter, am I right in thinking it is a way to assess their improvement? […] I would have imagined that the best way to get significance is over the class as a whole and on the 1st term when children were in lockdown [Nov-Feb] and used it the most. The average competence level on Jan 10th was 7.7 and it went up to 8.65 on Feb 21st before going down along with the amount of usage as the children came out of lockdown." T1 was trying to draw summative conclusions based on the line graph in an *unfiltered* state, that visualized data from all language categories, which, as discussed below, might mislead summative assessment. This email highlighted that **teachers' interpretation of visualizations should be further scaffolded**, e.g., by training *throughout* its implementation, for it to be used well formatively.

Contact point – online discussion. Following this email, we held an online discussion with T1 to review how she might use the LAD toward summative evaluation of children's literacy skills and fulfill the Deputy Headteacher's request. By screen-sharing the LAD, we jointly reflected on the challenges in interpreting "significant" improvement over time based on competence levels in the unfiltered line graph. By exploring the data, we identified that the recent drop in competence (identified in T1's email) may have been due to children tackling new, more difficult content, e.g., adverbs & adjectives, as opposed to easier features that were played during lockdown. We pointed out that it would be **inaccurate to take a snapshot of data from all language features over time**, as features were tackled at different times of the year and with increasing difficulty. As such, with T1's recent introduction of adverbs/adjectives in the classroom, children's competence appeared to plummet, but easier features (mastered but not played recently) were not factored into the visualization for that time point in the line graph. Contrastingly, by filtering for single language features that were practiced consistently over several months, we could see "truer" trends in how competence changed during lockdown, which could use as summatively. This discussion helped T1 understand that, **to use the LAD *summatively*, she would need to consistently assign the same language feature over a longitudinal period**, so that competency changes could be tracked with enough data. She suggested that she would try this going forward. We then reintroduced how T1 might use the LAD visualizations *formatively*, to better students as they progress in their learning. However, whilst this prospect excited her, T1 noted that time in class was a major limiting factor: **she did not have time to use the LAD in a formative way**.

Interviews. T1 reflected that our online discussion did not change how she approached the LAD, used *Navigo*, or taught literacy in the classroom, but reflected that not changing her approach proved disadvantageous for summative evaluation: "We continued looking at the games, changing the games regularly. Which meant that when we were looking at the results [in the LAD] and trying to see if we could pull out any positive data about the use of *Navigo*, we couldn't necessarily because every time I introduced something [new] it was gonna be hard. […] If I want to use *Navigo* to have data at the end of the year and show improvement, then […] I would assign some games and leave the games the same across the whole year, or across a long a period of time." T1 then reflected that she wished she had had earlier **discussions with the researchers about types of strategies** she could have used to ensure the game data and LAD visualisations could have been used for summative assessment, saying "well, I guess I didn't discuss maybe as much as I should […] I didn't discuss very much with you the different kind of strategies of how you can use *Navigo*". Yet, T1 saw how the LAD could be useful to inform her teaching practices, indicating even though she did not use the LAD formatively, it could have had potential to facilitate self-evaluation of her teaching practices and plan future learning activities: "[…] I still want to see how they're doing because if they're not doing well with adverbs and we've been talking of adverbs that means I haven't done a good enough job, or they haven't understood the game. So, we need to talk about it and maybe do another lesson. […] It will be an informative tool to decide what lessons I should do." Despite two email prompts about the LAD from the research team, T2 expressed in her interview that she forgot about using it. She went on to say that "I'm quite data-driven so I would find [the LAD] incredibly useful". Yet, like T1, she also expressed the lack of time in the classroom for this type of data exploration, and that it was difficult enough to find time for the game-based learning activities in the first place, let alone the LAD.

4 IMPLICATIONS FOR LAD DESIGN & IMPLEMENTATION

Our communications with T1 highlighted the issue of **approaching the LAD – which was designed for formative processes – through a summative evaluation lens**, in part due to the focus on summative assessments in schools (Beck & Nunnaley, 2021). The *Navigo* game breadth with 900+ activities combined with limited classroom time meant that children's gameplay was spread out. Thus, there was not enough data on any one language category to make concrete summative evaluations. When the unfiltered data was visualized on the LAD it created a biased and misleading visual representation of students' performance, based on how teachers selected new activities (Pelanek, 2021). Whilst we demonstrated a priori how the LAD was designed to be used for formative processes, we should have (in hindsight) explicitly discussed with teachers how a game with broad content, like *Navigo*, could also be used for summative evaluation, e.g., by limiting content assigned to children. LAD designers might also remove unfiltered data view in the line chart, so that misinterpretations cannot be made.

We also identified that **teachers need specific and contextual experience and practice with LADs to understand how to interpret visualisations and approach them with a formative mindset**. This implies that, in addition to pre-training, LAD providers should provide training over several months during use, to allow teachers to practice and build their skills in between training sessions. LAD designers might also build scaffolds directly into the LAD to guide teachers' formative interpretation of visualisations. For instance, **a LAD might incorporate alerts or advice to facilitate teachers' formative appraisal of data and act upon insights in the classroom**. Previous research has compared LADs which (i) mirror students' interactions and performance, (ii) alert teachers to data that need attention, or (iii) advise teachers about what actions to take next in the classroom (e.g., van Leeuwen et al., 2019). Such studies provide evidence that alerting and advising approaches might help teachers better interpret learning situations during 'live' classroom activities. Whilst we used color-coding to alert teachers to problematic concepts needing to be addressed, integrating *specific advice* on how to deal with these issues might have reduced T1's cognitive load and encouraged its more regular and formative use. In summary, despite interest and involvement in the design of the LAD and high data literacy, the teachers did not have time or skill capacity to use the LAD in meaningful, formative ways.

REFERENCES

Beck, J. S., & Nunnaley, D. (2021). A continuum of data literacy for teaching. *Studies in Educational Evaluation*, *69*(May 2020), 100871.

Kovanovic, V., Mazziotti, C., & Lodge, J. (2021). Learning Analytics for Primary and Secondary Schools. *Journal of Learning Analytics*, *8*(2), 1–5.

Pelanek, R. (2021). Analyzing and Visualizing Learning Data : A System Designer's Perspective. *Journal of Learning Analytics*, *8*(2), 93–104.

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

van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 261–289.

Vezzoli, Y., Mavrikis, M., & Vasalou, A. (2020). Inspiration cards workshops with primary teachers in the early co-design stages of learning analytics. *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK'20)*, 73–82.

Participatory Design of a Writing Analytics Tool: Teachers' Needs and Design Solutions

Tong Li
Arizona State University
tongli4@asu.edu

Sarah D. Creer
University of New Hampshire
Sarah.Creer@unh.edu

Tracy Arner
Arizona State University
tarner@asu.edu

Rod D. Roscoe
Arizona State University
work.roscoe@gmail.com

Laura K. Allen
University of New Hampshire
Laura.Allen@unh.edu

Danielle S. McNamara
Arizona State University
dsmcnamara1@gmail.com

ABSTRACT: Automated writing evaluation (AWE) tools can facilitate teachers' analysis of and feedback on students' writing. However, increasing evidence indicates that writing instructors experience challenges in implementing AWE tools successfully. For this reason, our development of the Writing Analytics Tool (WAT) has employed a participatory approach that includes teachers throughout all stages of design. WAT aims to generate writing analytics that are both useful to teachers and actionable by students. In this report, we discuss our participatory design approach and resulting insights, including (a) teachers' expressed instructional needs and goals, and (b) design solutions that were conceptualized to address both factors.

Keywords: Automated Writing Evaluation, Participatory Design, Writing Analytics Tool

1 INTRODUCTION

Automated writing evaluation (AWE) tools are computer-based systems designed to reduce teachers' workload by providing formative feedback to students on their writing (Wilson et al., 2021). Such systems typically use natural language processing (NLP) to analyze writing on basic features (e.g., grammar, mechanics, and punctuation) and more complex dimensions (e.g., cohesion and syntactic complexity). AWE systems then provide feedback to students that identifies potential targets and strategies for improvement. Studies have reported that AWE can assist teachers in improving students' writing quality and attitudes about writing (e.g., Wilson & Czik, 2016). However, several studies have documented teachers' challenges in using AWE tools to support writing instruction. For example, teachers may experience *increased* workload when the feedback provided by AWE systems is inaccurate, unclear, or otherwise difficult for students to interpret. In such cases, AWE implementation can have *negative* impacts on student writing (Palermo & Thompson, 2018; Wilson et al., 2021). When teachers have concerns about system accuracy or functionality, they may eliminate AWE tools (and related benefits) from their instruction.

For these reasons, our development of the Writing Analytics Tool (WAT) has embraced *participatory design methods* that incorporate teachers' insights and instructional goals throughout the process. WAT is an AWE system that aims to generate writing analytics that reveal meaningful features of students' writing for teachers, students, and researchers. In this paper, we outline the collaborative design process used to identify teachers' needs and then describe potential design solutions that can make writing analytics more accessible to teachers.

2 PARTICIPATORY DESIGN

Participatory design emphasizes a collaboration between designers and teachers wherein the mutual expertise of both is *equally valued*. This process recognizes that design challenges and solutions are context-specific, and context experts are crucial members of the design team (Kuhn & Muller, 1993). Following these principles, we recruited a WAT Teacher Advisory Board (TAB) comprising six secondary writing teachers (i.e., 4 high school; 2 middle school) via their participation in the National Writing Project (NWP). The group possessed an average of 18.7 years of teaching experience (ranging from 9 to 25 years) and all of the teachers had professional experience in supporting writing instruction with technology.

Teachers participated in six recorded focus group sessions (60-90 minutes) conducted over Zoom. Each meeting followed a similar structure, but activities and discussion varied organically based on the topic (see Table 1 for a high-level summary). Discussions from prior sessions informed the topics of later sessions.

Table 1: Focus Group Session Topics

Session	1st	2nd	3rd	4th	5th	6th
Topic(s)	personal introductions; overview of WAT project	information teachers need or desire for evaluation	features and challenges of student writing	core needs for WAT interface design	planning and developing writing assignments	providing feedback and dialog with students

Teachers had multiple opportunities and modalities for expressing their needs, ideas, and goals. First, sessions were preceded by a pre-meeting survey that introduced session topics and promoted initial reflection. Survey responses could be revisited during the session to elicit further discussion. Second, each session included a shared digital document that teachers and developers could contribute to simultaneously (e.g., notes and commentary). Thus, teachers were able to express themselves in their own words, both in speech and in text. Third, each session included a "Writing into the Day" activity conducted within the shared document. Prompts for these exercises linked content from previous sessions to the current session, and prompts were often derived from teachers' own quotes from prior meetings. Fourth, virtual "breakout rooms" enabled focused small group discussions, which were then shared back to the larger group. These discussions focused on identifying "pain points" and teachers' AWE needs for overcoming these challenges.

3 NEEDS AND PROPOSED SOLUTIONS

As needs were uncovered, the design team identified potential solutions that were shared for teacher input and collaboration. Several sets of design needs and potential solutions emerged from the TAB focus groups and iterative dialog. For brevity, three examples are described below.

3.1 Need 1: Emphasizing teachers' roles and agency in AWE implementation

Teachers expressed concern about the accuracy and efficacy of AWE systems. In their view, current AWE systems are not contributing to students' deeper learning or writing skill development. They explained that students who use these tools often focus on the "scores" and make superficial changes rather than meaningfully improving their texts. TAB members argued that AWEs should supplement, not replace, teacher feedback. Specifically, teachers are better equipped to provide personalized and contextualized feedback that is aligned to students' needs, and thus their role and agency in crafting, framing, and delivering feedback must be paramount.

To address these needs, the writing analytic aspects of WAT were (re)conceptualized to emphasize *detecting* and *identifying* features of student writing, rather than *evaluating* or *scoring* such features. Using such information, teachers are free to communicate their own guidance or evaluation. For example (Figure 1), WAT can assess the level of text concreteness or abstraction. This pattern can be shown to teachers and students, and teachers can elaborate in a comment box (e.g., how to modify concreteness depending on instructional goals). Teachers can also provide evaluative framing by categorizing the current feature as positive, in need of improvement, or simply general information.

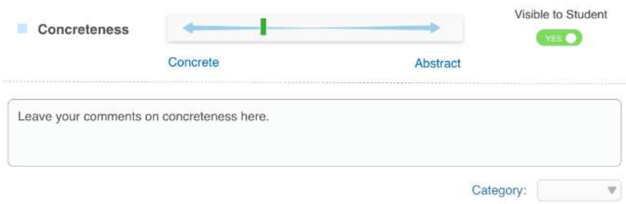
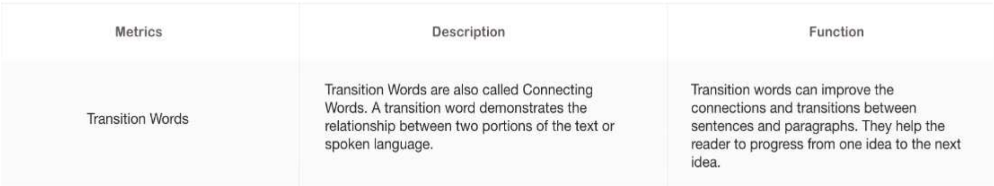


Figure 1: Feedback interface

3.2 Need 2: Supporting students' writing agency

A discussion about "writing as a process" revealed that teachers wanted AWE tools to also reinforce students' own agency in writing and revising. TAB members argued that automated feedback should not govern students' writing nor limit their independence. Resulting design solutions focused on helping students understand the meaning of selected metrics along with strategies for modifying such features of their writing. We thus added a "Library" that explains each metric and how it contributes to communicating with readers (Figure 2). Each metric is accompanied by considerations for revising (not shown in Figure 2). This functionality enables students to adopt a more informed and intentional approach.



Metrics	Description	Function
Transition Words	Transition Words are also called Connecting Words. A transition word demonstrates the relationship between two portions of the text or spoken language.	Transition words can improve the connections and transitions between sentences and paragraphs. They help the reader to progress from one idea to the next idea.

Figure 2: Metric library interface

3.3 Need 3: Supporting teacher-student communication

TAB members agreed that an important element of their instruction was the ability to engage students in dialog about writing. In these interactions, teachers use their wealth of information about students to provide personalized feedback. Thus, teachers wanted AWEs to help them further understand students' writing perceptions, struggles, or intentions, which are not always visible or explicit. We thus developed a "tag and flag" function that enables students to communicate about their writing by annotating segments of their essays. "Tags" allow students to label specific writing elements (e.g., arguments and topic sentences) and "flags" allow students to metacognitively indicate their status (e.g., "I am still working on this"), intent (e.g., "I'm trying to be funny here"), or ask for help (e.g., "Does this sound right?").

Figure 3: Teacher-student communication interface

4 CONCLUSION AND NEXT STEPS

Participatory and collaborative design with teachers revealed that teachers crave AWE systems that will (a) support and not replace their central roles as writing instructors and mentors, (b) provide them with actionable information about students' writing, and (c) facilitate students' writing agency. Our next steps are to complete and test prototypes for emerging solutions, including seeking TAB expertise on whether proposed solutions actually satisfy their needs and expectations. Upcoming focus group sessions with the TAB (Phase 2) are expected to uncover further needs and functionalities for WAT design. In essence, participatory design is a continuous improvement process that never ends.

ACKNOWLEDGEMENTS: This work is supported by the IES foundation (R305A180261).

REFERENCES

Kuhn, S., & Muller, M. J. (1993). Participatory design. *Communications of the ACM*, *36*(6), 24-29.

Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research, 12*(1), 63–108.

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education*, *168*, 104208.

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94-109.

Human-centered learning analytics in interdisciplinary projects: co-designing data collection

Birte Heinemann, Matthias Ehlenz, Ulrik Schroeder
RWTH Aachen University
Heinemann| Ehlenz|Schroeder@cs.rwth-aachen.de

ABSTRACT: Analysis of learning data can provide insights into the learning process, ideas for improving learning environments, and approaches for feedback. In virtual reality, a wide variety of data can be collected. Matching this data with learning objectives and finding indicators and the interdisciplinary path between didactics, the scientific domain of the application and the developers to establish a learning analytics strategy is a challenge. This paper presents two field-tested approaches to implement the cyclical process of learning analytics with interdisciplinary stakeholders with a focus on scalable data collection. The wide range of possible applications, the possible interconnection to other technologies and the large breadth of teaching approaches, require sustainable structures and interdisciplinary user-friendly processes, both for the analysis of the data, but also for the development of the technical infrastructures. These requirements are especially true when Open Science practices and "FAIR Data" principles are followed.

Keywords: Human-centered Learning Analytics, HCLA, xAPI, Learning Analytics, Co-Design

1 INTRODUCTION

Teaching and learning are not limited to the realm of the traditional learning management system (LMS); large multi-touch tablets, distributed learning in remote labs and virtual and augmented reality (VR and AR) bring new opportunities and challenges. Learning analytics (LA) is one approach to improving learning technologies in many ways. In addition, other developments can be observed: Open Educational Resources and Open Science, which stand for active exchange in teaching and research. This raises the overarching question of how these processes can be supported and how the exchange in science can be optimized and promoted. In addition to these goals, research for education in VR/AR builds on a variety of learning models, teaching and learning methods and approaches. Educational technology brings together different disciplines and is a highly interdisciplinary field of research.

This paper describes different approaches and fundamentals to integrate learning analytics in VR/AR and shows different requirements for an implementation process. It highlights the co-design of the data collection step and the creation of sustainable infrastructures with interdisciplinary stakeholders. Based on the results of the analysis, tried and tested approaches and experiences in interdisciplinary projects are described and needs for sustainable and open developments are explained.

2 STATE OF THE ART AND REQUIREMENTS

Among others, (Prieto-Alvarez et al., 2018) report on the co-design of learning analytics tools. They show a process model that helps to involve the different stakeholders, especially learners. The overview about current approaches to connect design thinking, human-centered learning analytics and learning analytics mentions, that it is also interesting to involve the stakeholders for research in privacy and learning analytics. Complementary to the involvement of learners and teachers, the interoperability of systems and privacy by design are important points when talking about privacy and the development of learning analytics (Hoel et al., 2017).

In the described/underlying project, we are building an infrastructure for learning analytics, which should be scalable, transferable, sustainable and expandable. (Kraker et al., 2011) and (Murray-Rust, 2008) laid important foundations for the topics of "Open Science and FAIR data", which we incorporate in our developments. They argue that being able to share data and aggregate different data sets can eliminate measurement effects, help to get larger samples, and thus make more valid statements about learning. After research on data formats, we have chosen Experience API (xAPI) as the data standard for our projects, as we also want to merge multimodal data sources, technologies and projects (Ehlenz et al., 2020). In VR for example, we are able to collect gaze data, gestures and logging of interactions, which we connect to Moodle data and data from real experiences in a learning laboratory (Pfeiffer et al., 2020). (Mu et al., 2020) show a wide range of data sources, which are in complex relationships with learning indicators, which are interesting for research projects, for giving feedback to learners and teachers. The uniform semantics in a standardized data format saves resources and facilitate the analyses. The possibility of recording multi-agent statements is an important point for scalability.

In order to find the interesting indicators for laboratory-based learning in engineering education and to map or prepare the analysis with the xAPI statements, we have conducted various workshops. The workshops and the lessons learned are described in the following chapter.

3 CASE STUDY CO-DESIGN WITH DEVELOPERS, EDUCATORS AND RESEARCHERS

At first, getting familiar with the idea of xAPI statements was easy for the interdisciplinary team of educators, researchers, and developers – and the computer science students who connected with the project through research-based teaching. But it became difficult in the practical implementation because different specialized didactics, perspectives on learning, technical vocabulary and ideas have collided and need to be formalized in xAPI statements, which have to fit the technical infrastructure, the possibilities of the institutions, the research questions and the research setting. The lack of databases of well-defined dictionaries and the lack of public and well-documented community standards, see (Ehlenz et al., 2020), lead to our efforts to push forward the process and development of data collection with co-design and with this push the exchange, comparison, and reproducibility of research results. Many of the methods described in (Prieto-Alvarez et al., 2018) have also been applied in our project (personas, focus groups and prototyping). We additionally used storyboards to visualize the interesting scenarios and applications of learning analytics and to familiarize some partners with the human-centered design process before the described workshop. All participants knew the basic structure of xAPI statements from previous meetings.

The workshop(s) described are based on shared project work and represent a practical view on co-design, which followed an iterative design process. The workshops lasted 1.5 hours, with a group work phase of one hour. In the beginning and the end were plenary phases, first to recall prior meetings, second to share the results after the group working phase. Seventeen participants were divided into three groups for three learning scenarios and worked on possible learning indicators and “we” tried to collect vocabulary for the statements as well as connections to the indicators. Among the participants were the addressed stakeholders: engineers, education specialists, teachers and developers.

The first workshop with the goal to create a well-defined database for the vocabulary was based on knowledge mapping and mind/concept maps (Prieto-Alvarez et al., 2018). The participants formed groups themselves - learning scenario-specific, they wanted to work on their own use case. Fig. 1 shows an excerpt of the results of one group. Later workshops proved helpful to vary these groups from time to time and to form groups according to their perspective, e.g. all education experts in one group. In the first workshop, we distinguished between scenario-driven and question-driven approaches (left or right branch in fig. 1). The participants could choose whether it is easier for them to identify indicators when they start from the learning objectives or from the learning actions. Both pathways addressed different stakeholder groups and the work was partly done in parallel. Especially the connection of the perspectives was interesting and lead to some discussions. One lesson learned is that the creative way of working together was difficult for them at first. The participants wished they had learned more about the methods of design thinking and co-creation beforehand. In a subsequent workshop, with nearly the same group, we therefore adopted the visualization and prepared a timeline of learning activities (fig. 2). Participants could assign learning objectives, research questions and refined learning actions at each point, leading to measurement methods, data collection (xAPI vocabulary and definitions) and indicators. The stronger order through the timeline helps to start and have discussions between stakeholder groups. Overall, the second approach was more successful. In future iterations, we would share the timeline with the participants in advance, because most wished time to prepare better. According to the global restrictions, we look forward to test this concept in face to face workshops. Compared to the previously used methods of 1:1 interviews and questionnaires, we can see a stronger involvement in learning analytics in the project and a broader understanding of our own methods and research interests.

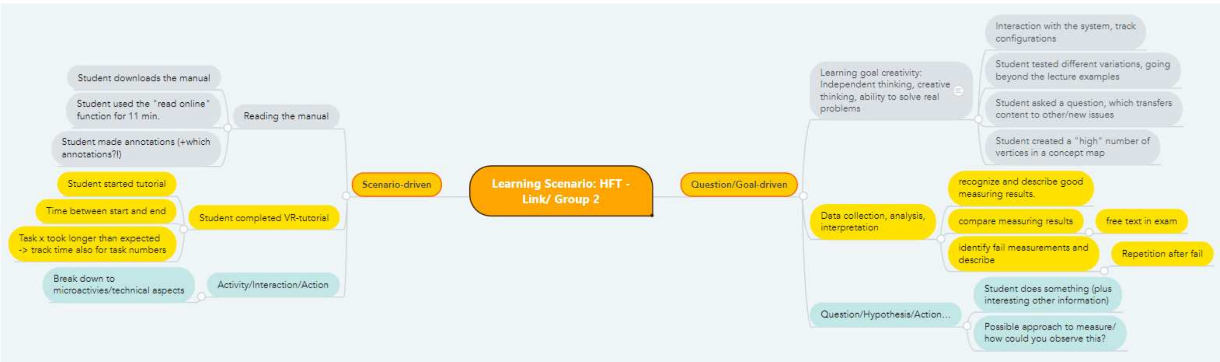


Figure 1: Excerpt of results of the first workshop - finding indicators analyzing activities (left) or connecting to objectives and questions (right).

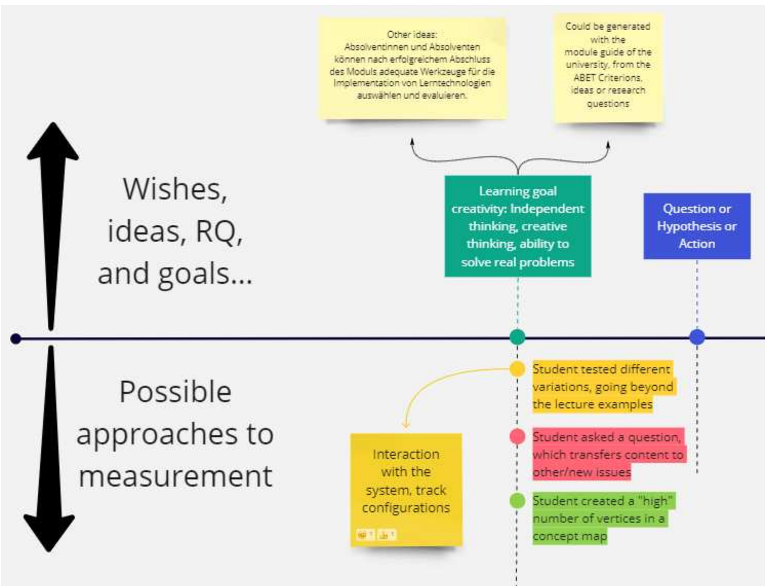


Figure 2: Timeline of the learning activities. In the upper part wishes, ideas and goals for the learner as well as research questions of the project, connected to specific activities. Below are possible approaches to measure learning and indicators with color-coding. The color-coding - showing how easy it is to be implemented - was done after the workshop by the LA experts in order not to restrict the creativity of the participants.

4 CONCLUSION AND OUTLOOK

In research on education, a multitude of methods, theories, technologies and tools collide. To further augment these challenges, interdisciplinary research projects increase this heterogeneity by integrating styles and methods of operation from these domains. Networked and cross-disciplinary work provides challenges for a systemized approach to sense-making. The foundation for this is laid in the first connecting step: data collection. Collecting sufficient data in this multi-modal context requires a common layer of scientific communication, or to be precise: specification, standardization and meta-data, a challenge also described by (Son & Cho, 2017). We need to consider at least two perspectives. First, from a technological standpoint, it requires data formats and specifications. The second perspective we discuss in this report is interdisciplinary acceptance, it is the collaborative interdisciplinary development of learning analytics, human-centered learning analytics. A requirement for effective usage beyond the obvious is a common understanding of the possibilities of LA in our fellow researchers from other domains. Human-centered learning analytics provides principles to assist the process of building an appropriate toolbox for these challenges. We tried to integrate all involved parties early in the process of learning analytics, even before the first data set is collected, by using design-thinking methods. This paper shares our experiences in constructive workshops to integrate all perspectives of an international, cross-domain research project, which remains a difficult task. In our case it was more efficient to use temporal sequences to discuss with various stakeholders and to create a joint data collection. In the future, we aim to reiterate this method, digital and face to face, and -virtually by itself- create a meta-data repository for xAPI definitions with ever-increasing acceptance from a wide range of fields.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the funding by the Federal Ministry of Education and Research (BMBF) for the joint project Open Digital Lab 4 you (grant no. 16DHB2114)

REFERENCES

Ehlenz, M., Heinemann, B., Leonhardt, T., Röpke, R., Lukarov, V., & Schroeder, U. (2020). Eine forschungspraktische Perspektive auf xAPI-Registries. In R. Zender, D. Ifenthaler, T. Leonhardt, & C. Schumacher (Hrsg.), *DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.* (S. 331–336). Gesellschaft für Informatik e.V.

Hoel, T., Griffiths, D., & Chen, W. (2017). The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 243–252. https://doi.org/10.1145/3027385.3027414

Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. *International Journal for Technology Enhanced Learning*, *3*(6). https://doi.org/10.1504/IJTEL.2011.045454

Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors (Switzerland)*, *20*(23), 1–27. Scopus. https://doi.org/10.3390/s20236856

Murray-Rust, P. (2008). Open Data in Science. *Serials Review*, *34*(1), 52–64. https://doi.org/10.1080/00987913.2008.10765152

Pfeiffer, A., Heinemann, B., Uckelmann, D., & Schroeder, U. (2020). Learning Analytics in Engineering Education: How to Support Laboratory-based Learning with Learning Analytics. *ICERI2020 Proceedings*, 2484–2492. https://doi.org/10.21125/iceri.2020.0589

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. D. (2018). Co-designing learning analytics tools with learners. In J. M. Lodge, J. C. Horvath, & L. Corrin (Hrsg.), *Learning Analytics in the Classroom* (S. 93–110). Routledge.

Son, J., & Cho, Y. (2017). A Study on Capturing Learning Data from Virtual and Mixed Reality Contents Through Data Collection API. *Workshop Proceedings of the 25th International Conference on Computers in Education*, 8.

Amplifying the Signal: Exploring the Usefulness of Natural Language Processing to Inform Student Success Decision-Making

Marilee Bresciani Ludvik, Henry Anderson, Elizabeth Clinton Powers, Pete Smith, Catherine Robert, Amanda Olsen, Mitch Strahlman, Nasima Subedi, Shiming Zhang, Danielle Klein, & Molly Albart
University of Texas at Arlington

Stephen Schellenberg, Sandy Kahn, Nina Potter, Maureen Guarcello, & Rey Monzon
San Diego State University

ABSTRACT: As COVID-19 recovery efforts continue, universities are faced with thousands of students whose on-campus experience was delayed by stay-at-home orders and other pandemic-related precautions and obstacles. This presentation outlines research conducted during the pandemic, which demonstrated early indications of equity gaps intensifying among various student identities. Responding to students' needs, and working to ensure their paths to healthy, equitable, and ultimately academically successful college experiences is a time-sensitive and often indistinctive undertaking. With a deluge of available data, even sophisticated data analysis models struggle to cut through multitudinous variables and the noise that might otherwise broadcast a clear signal of students' needs. The first-person voices of students, specifically those responding to prompts soliciting first year experience descriptions, provides timely and actionable data, but are unwieldy to collect and time-consuming to analyze. The study observed the viability of using open-ended student responses to inform just-in-time interventions. Using various NLP analysis related to amplifying students' first-person voices during and after COVID-19 remote learning took place, open-ended response results from students enrolled in a first-year university seminar at an urban Hispanic Serving Institution prior to COVID and during COVID were compared for usefulness in informing just-in-time interventions to decrease academic achievement gaps.

Keywords: student success, natural language processing, equity gaps, COVID-19, remote learning

1 BACKGROUND

Research indicates that understanding students' experiences in school can be useful to inform decisions that improve student success, and can also close equity gaps (Stephens et al., 2014). Often, students share elucidating information when they meet with student affairs professionals and faculty one-on-one (Moran, 2001). Qualitative data reveal insights into students' thoughts and feelings about their various in- and out-of-class experiences (Yazedjian et al., 2008). These data provide a rich understanding about current students' needs, and often those findings can also be used as a blueprint to support future students.

Gathering students' first-person voices has equipped institutional leaders with helpful information about how policy and practice may enhance student success for all, while also informing changes that can close equity gaps (Maunder et al., 2012). Of recent concern is understanding how the pandemic has impacted students' experiences, particularly as the pandemic has shown to exacerbate equity gaps (Duckworth, et al., 2021; Levine & Van Pelt, 2021).

1.1 Problem

While first-person student voices can be useful in understanding how to improve first time in college (FTIC) experiences, they can take a significant amount of time to collect and analyze. Thus, the institutional ability to apply what they learn from student narratives to inform decision-making in a timely manner can be challenging. During the time it takes to responsibly collect, analyze, interpret, and use those data, institutions may unknowingly be perpetuating equity gaps.

1.2 Significance

Identifying a way to collect, analyze, interpret, and apply first-person voice to expedited decision-making for first-year, first-time college (FTIC) students prior to the end of their first term could prove beneficial to preventing equity gaps that often materialize in first-term institutional academic achievement measures such as cumulative GPA and term-to-term persistence. This research team posited that discovering efficiencies in data collection and analysis may reduce the time it takes to make decisions and iterate on policy and practice in between semesters, thus addressing equity gaps faster and with more personalized interventions.

2 IMPLEMENTATION

2.1 Data Collection

Data collection took place at a large, urban, public HSI research intensive institution, with the administration of one open-ended question at the end of a FTIC university seminar, as part of a course assignment measuring pre- and post-intrapersonal competency cultivation gains. The question invited students (N=2,050) to “briefly describe your experiences this semester.” The first-term, one-unit credit/no-credit course was designed to cultivate sense of belonging, positive future self, metacognitive awareness, and other intrapersonal competencies known to significantly influence student success (NAS, 2017; 2018; Zelazo Blair & Willoughby, 2016).

2.2 Analysis

To understand whether first-person student voice could easily be collected prior to the end of the first semester and used to inform just-in-time decisions that could take place prior to the second semester, the research team set out to analyze the end-of-term, open-ended survey question. Comparing two Natural Language Processing (NLP) analytic methods – Linguistic Inquiry and Word Count (LIWC) and TextRank – using a set of pre-COVID (Fall 2019) and during-COVID (Fall 2020) FTIC student data of those enrolled in a first-year university seminar. The team sought to understand pre-COVID and during-COVID FTIC experiences in a manner that could improve the design of the FTIC experience and support between-term contacts with students. The intent of this exploratory analysis was to assess the feasibility of whether LIWC and TextRank could also be used by student success practitioners on weekly journal collections to inform just-in-time intervention decisions during the first term.

3 METHODOLOGY

3.1 TextRank

Exploratory and qualitative analysis was performed using the TextRank algorithm, resulting in a list of highly representative sentences (rather than words) from each set of item responses. TextRank assigns a score to each sentence with a high score indicating a sentence that is "important," i.e., it is similar to a number of other "important" sentences. The resulting scores serve as proxies for *representativeness;* sentences with the highest scores serve as summaries for the overall trends.

3.2 Linguistic Inquiry and Word Count (LIWC)

Other aspects of the text were analyzed using the 2015 version of *Linguistic Inquiry and Word Count* (*LIWC)* tool (Pennebaker, 2015). This research tool has been used in computational analysis of texts across a variety of domains over the past 15 years. LIWC reports on a range of measures for each text, primarily including high-level psychometric measures such as *authenticity* and *clout*, and heavily researched linguistic features such as *number of words per sentence*, *number of words greater than six letters*, number *of pronouns,* and *number of sentences per paragraph*.

4 FINDINGS

4.1 Overall Student Sentiment

Results demonstrated that students reported *positive* experiences during Fall 2019 semester, including plans to take more courses during future semesters, they also expressed concerns about achievement, work, and time. Contrarily, students reported overwhelmingly *negative* experiences over the Fall 2020 semester. These experiences were most often complaints about the online modality and the lack of interaction with peers as a result.

4.2 Pre-COVID Responses

In Fall 2019, students overwhelmingly reported positive experiences during the semester. Specifically, TextRank analysis revealed a consistent set of themes among the top-ranked sentences. Some students reported plans to take more courses in the future, citing the unexpected ease of the previous semester. Students also demonstrated slightly higher concern about *work* and *time* than would be expected in natural speech, and much higher emphasis on *achievement* than would be expected in natural speech, compared to Pennebaker et al. (2015)'s published baselines.

Using a standard LIWC analysis, emotional tone is measured on a normalized 1-100 scale with the strength of the sentiment indicated by scores further from the midpoint. Approximately 10 percent of responses demonstrated *extremely* negative emotions on this scale, and close to 60 percent showed *extremely* positive emotion. The remainder of responses fell into moderate positive and negative ranges. The distribution was validated by separate measures of the percentage of negative and positive words used in responses compared to the LIWC baseline measures.

4.3 During-COVID Responses

TextRank analysis of student responses during COVID revealed a consistent set of themes among the top-ranked sentences. Themes included: students overwhelmingly reporting negative experiences directly relating to online instruction modality, and about the difficulty of meeting other students. Contrary to the LIWC sentiment findings pre-COVID, student emotional tone outcomes showed 40 percent of responses demonstrated a negative tone, with most scores falling below 25, while 30 percent showed a positive tone. Analysis did not show an increased level of concern in the *health*, *social,* or *friend* categories, though it did reveal a heightened degree of concern about work, which was mirrored in the percentage of words included in the *achievement* and *time* categories.

5 LIMITATIONS AND RECOMMENDATIONS

Through this successful proof of concept, lessons learned using NLP methods to analyze student data, range from providing greater writing prompt clarity, to the distribution of writing assignments throughout the semester. Responses from the different methods of NLP analysis yielded data informing in-between term interventions for the FTIC cohort. Limitations included the inability to draw conclusions for specific identities because of the large sample size needed for NLP analysis. The team noted the difficulty of extracting additional rich text from the learning management system, and the inherent bias of those students whose responses were analyzed, as well as the implicit bias of the research team. Future aspirations include the introduction of learning management system data and predictive modeling to create a more robust dataset, and to potentially triangulate findings with the rich quantitative data collected on intrapersonal competencies and other learning management system engagement, before carefully considering intervention strategies.

REFERENCES

Duckworth, A. L., Kautz, T., Defnet, A., Satlof-Bedrick, E., Talamas, S., Lira, B., & Steinberg, L. (2021). Students attending school remotely suffer socially, emotionally, and academically. *Educational Researcher*. https://doi.org/10.3102/0013189X211031551

Levine, A., & Pelt, S. V. (2021, August 26). Opinion: 5 ways higher ed will be upended. *The Chronicle of Higher Education*.

Maunder, R. E., Cunliffe, M., Galvin, J. et al. (2013). Listening to student voices: student researchers exploring undergraduate experiences of university transition. *Higher Education 66*, 139–152 (2013). https://doi.org/10.1007/s10734-012-9595-3

Moran, C. D. (2001). Purpose in Life, Student Development, and Well-Being: Recommendations for student affairs practitioners, *NASPA Journal*, *38*(3), 269-279, DOI: 10.2202/1949-6605.1147

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin. doi:10.15781/T29G6Z

Stephens, N. M., Hamedani, M. G., & Destin, M. (2014). Closing the social-class achievement gap: A difference-education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science*, *25*(4), 943–953. https://doi.org/10.1177/0956797613518349

Zelazo, P.D., Blair, C.B., & Willoughby, M.T. (2016). *Executive function: Implications for education* (NCER 2017-2000) National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

A Framework for Equitable Educational R&D

Author(s): Mark Rose Lewis
Learn-RD
mark@learn-rd.com

ABSTRACT: How do learning analytics (LA) and academic analytics (AA) lead to substantial increases in student learning/success and reductions in the persistent equity gaps that plague education? The short answer is that LA and AA can't do it alone and that any progress in this space requires both technical and social innovation. In this poster, we present a framework for combining LA and AA with social research, evaluation, and design practices as well as integrated learning engineering platforms into a broader practice of Equitable Educational R&D. We highlight some of the shortcomings of LA, AA, and current learning engineering platforms that inspire this work and provide a case study of how this framework has been implemented along with key lessons learned.

Keywords: Learning Analytics, Equity, Developmental Evaluation

1 INTRODUCTION

The need for profound educational change is clear. Students continue to struggle to learn and thrive in deeply inequitable institutions and systems. LA and AA are promising, but still new, fields that have the potential to drive significant improvements in student learning and equity. That said, a number of limitations have characterized their application to date: a focus on small-scale, exploratory studies; the relatively infrequent inclusion of teachers/learners as co-designers; a tendency towards reductionist "single number" outputs; a failure to be integrated with theory building; serious concerns about equity in development and use; and uncertain paths to real-world impact (Buckingham Shum et al., 2019; Campbell et al., 2007; Dawson et al., 2019; Ferguson, 2012; Selwyn, 2019). In this paper, we propose a framework for Equitable Educational R&D (EER&D) that integrates LA and AA with social practices from developmental evaluation, equitable evaluation, and action research to build shared knowledge, effective tools, and user-centered analytics in order to drive substantial improvements in student learning, success, and equity.

In this framework (see Figure 1), improvements in learning, success, and equity are driven by shared research, evaluation, and design practices that result in co-designed, equity-focused LA and AA that highlight potential paths to equity and student success; equitable and effective tools, interventions, practices, and policies; and an actionable and accessible understanding of student learning. The products of this shared R&D allow teachers, support staff, administrators, and policymakers to know where students are at academically and socially-emotionally and what works and when for diverse students. It also gives them access to tools, practices, programs, and policies that work. This work is enabled by rich information that comes from integrated learning engineering platforms that capture holistic, naturalistic learning data from diverse students from a wide range of sources. It is also enabled by the voices of students, teachers, and community members gathered through qualitative inquiry and co-design. At every step of the way, a focus is placed on equity – not only on producing more equitable outcomes, but ensuring that data is representative and fairly processed and ensuring voice and genuine participation in the process.

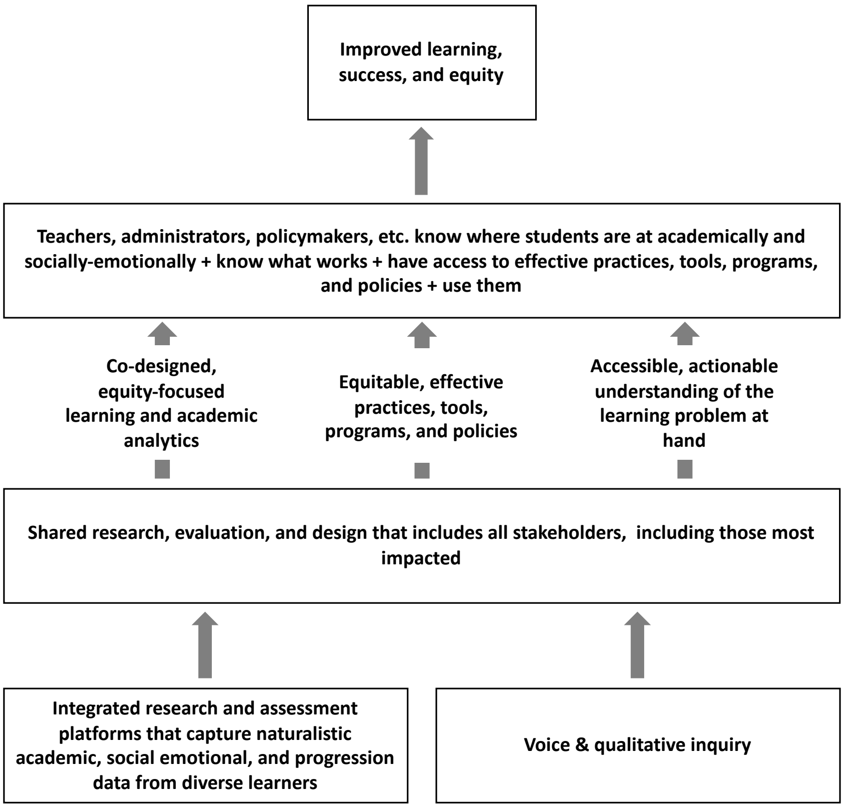


Figure 1: A theory of change for EER&D.

2 CASE STUDY – RETHINKING COLLEGE COURSE PLACEMENT

We used this framework to design and test a new course placement process at a large suburban community college. Most community college students in the United States must take a series of placement exams before registering for classes. The results of these exams determine if a student can take standard college-level courses or must take developmental (historically call "remedial") classes that do not count toward degree completion and can add months, or even years, to a student's degree plan. This results in an impactful and highly inequitable process, with significantly higher percentages of BIPOC students placing into developmental courses than do white students.

This work began by assembling a representative team of instructors, advisors, administrators, admissions staff, IT personnel, and an educational psychologist/developmental evaluator. The next step involved building a platform that integrated academic and SEL learning data from the college's learning management system, assessment systems, and survey platform as well as academic progress data from the college's student information system. These data were transformed into a set of prototype user-centered views of student learning, progress, and equity that revealed large inequities in course placement, course-level learning, course success, course completion, and degree completion associated with the existing placement process. The collaborative research team next engaged in a wide range of stakeholder interviews to get qualitative information to drive understanding and support the design of a new process – these interviews revealed large emotional impacts on students and strong faculty cultural assumptions that needed to be accounted for during design, implementation, and testing. We then used the platform to generate a set of what-if analyses (through regression models and Monte Carlo simulation) that simulated the impact of different policies on student success, learning, and equity gaps. Finally, we used the platform to analyze and share the results of a randomized control trial (RCT) of a new policy – co-designing custom views of the results and future scenarios for different stakeholder groups.

RCT results showed a significant reduction in course placement equity gaps (50% in reading) and significant increases in college-level writing and math completion within one year (26% and 60%, respectively). We presented these results dozens of times, resulting in a shared college decision to implement this new policy. We continue to use this platform to track the impact of this new policy and new programs that have been designed since on academic progress and student learning at the institutional-, departmental-, and course-level – looking at everything from degree completion to the quality of student argumentation.

Some of the key lessons learned from this work are outlined in Table 1.

Table 1: Key Lessons Learned

Integrated learning engineering platforms are critical to this work. An integrated learning engineering platform gave the research team the ability to rapidly test hypotheses about academic and social-emotional learning and student success and to easily track the effects of experimental conditions in an RCT.
Culture is powerful, but can be changed. The largest barrier to change in this work was cultural expectations among instructors, but these shifted as the work progressed and student voices and quantitative results were analyzed by the team. Participation in research, not just the results of the research, has had an impact on student success.
Research and evaluation expertise is critical to supporting this type of work. Change would not have been possible without patient, embedded research support.
Equity must be centered at all stages. Equity must be highlighted through the inclusion of diverse voices, representative data in learning models, a continuous focus on equitable outcomes, and co-design with those impacted by tools, practices, and policies.

REFERENCES

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

Campbell, J.P., Deblois, P.B., & Oblinger, D.G. (2007). Academic Analytics: A New Tool for a New Era. *EDUCAUSE Review*, *July/August*, 41–57.

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.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5–6), 304–317.

Selwyn, N. (2019). What's the Problem with Learning Analytics? *Journal of Learning Analytics*, *6*(3), 11-19-11–19. https://doi.org/10.18608/jla.2019.63.3

Team engagement with learning analytics dashboards

Andrew Zamecnik[1], Vitomir Kovanović[1], Georg Grossmann[1], Srećko Joksimović[1], Abelardo Pardo[1]
[1]University of South Australia
andrew.zamecnik@mymail.unisa.edu.au, vitomir.kovanovic@unisa.edu.au, georg.grossmann@unisa.edu.au, srecko.joksimovic@unisa.edu.au, abelardo.pardo@unisa.edu.au

ABSTRACT: This poster describes a natural experiment in a team-based learning environment, which involves a mixed-methods approach to understand how teams use a learning analytics dashboard (LAD) and how it impacts their collaborative learning. The aim is to understand team member interactions with the LAD by collecting their trace data and the influence on their learning processes, such as team learning behaviours. The results from the LAD trace data show that team members and leaders have regular and distinctive engagement behaviours with the LAD's features, such as the networking visualizations. The study concludes by recommending the designs of collaborative LADs to personalize to team members roles and exploring how the trace data can be used to analyze learning processes.

Keywords: collaboration, learning analytics, K-12, dashboard

1 INTRODUCTION

There are current literature gaps that explore how teams in collaborative learning interact with a learning analytics dashboard (LADs). The importance of these interactions will allow a further understanding of how K-12 students use them to regulate their learning and enable opportunities to personalize their learning (Bodily & Verbert, 2017). Many studies show that LADs empower learners by helping them monitor and reflect on their learning progress and outcome through feedback (Jivet et al., 2021; Verbert et al., 2020). However, most of the research has mainly centered around individual learning, as opposed to team-based learning. Studies that have collected log data based on the student's interaction with the dashboard drawing connections with their learning behaviours is also limited. Nevertheless, literature utilizing LADs in teams is growing, but more research is required to understand how teams and their roles engage with the dashboard and the impacts on their collaborative learning. Therefore, the study offers two complementary goals: 1) develop a LAD specifically for a team-based learning environment, which aims to inform teams about their collaborative learning progress using various visualizations and; 2) collect the trace data from teams to establish a relationship with their collaborative learning behaviours and academic performance using quantitative approaches.

2 METHODOLOGY

The study involves eight teams working towards their Certificate III in a K-12 learning environment using the Challenge-based learning platform (http://challenge.curtin.edu.au) for 20 weeks. The challenge platform is designed to support students' learning in individual and team-based contexts, focusing on open-ended problem-based learning. As such, team members must upload documents and edit their work collaboratively to complete a task on the platform, which is cross-checked by the instructor, who provides feedback and determines if the document meets the learning outcomes required (i.e. problem-solving abilities). Teams must complete 91 artefacts categorized into 26 major tasks for their main problem-solving project. As such, we built a student-facing LAD using the LATUX workflow methodology by Martinez-Maldonado et al. (2015). The LAD was introduced to teams in Week 6 and allowed teams to view their collaborative processes over time through visual depictions (See Figure 1). Moreover, given many tasks, teams used the LAD to help monitor and regulate their learning by observing team member engagement and collaborations with each artefact. The dashboard allows teams to keep track of their tasks (i.e. including edits made), team member interactions and their learning outcome progression related to problem-solving over time informed by learning analytics visual illustrations. The main research aims are to understand how teams use the LAD for collaboration and the impacts on their learning.

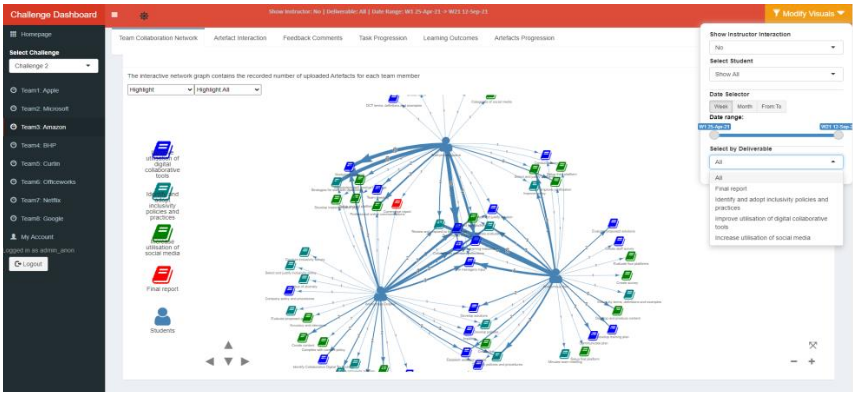


Figure 1: Example of LAD Team Collaboration Network

To analyze the data, we measured team members' interactions with the dashboard's features (i.e. collaborative network, artefact interaction, feedback comments artefact progression, task progression and learning outcomes) to understand the differences between team roles.

3 RESULTS AND DISCUSSION

Table 1: Total clicks (CL) and the proportion of clicks (%) with the dashboard features by each team leader (TL) and team member (TM) per team

Team Role	Collaborative Network		Artefact Interaction		Feedback Comments		Artefact Progression		Task Progression		Learning Outcomes	
	CL	%	CL	%	CL	%	CL	%	CL	%	CL	%
TM	43	48	7	7	6	7	19	26	5	6	5	5
TL	25	61	1	2	2	3	11	26	2	3	2	4

Table 1 shows a trend where on average, the team leader had fewer overall interactions than team members with each feature of the LAD. However, team leaders had 13% more engagement with the collaborative network feature than team members in terms of interaction proportion. The preliminary findings indicate that team leaders tend to find the collaborative network more engaging and useful than other features of the LAD. Overall, the surprising and unexpected finding is that team members are more engaged with the LAD than team leaders. Team roles may exhibit particular collaborative learning strategies when monitoring their teams progress. However, further research is required to understand why team roles display different behaviours with the LAD.

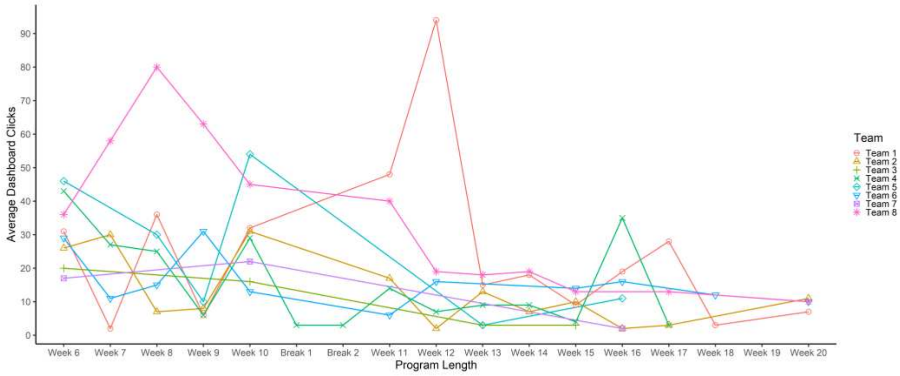


Figure 2: Average number of clicks spanning Week 6 and Week 20 for each team

Figure 2 shows that teams had more interactions with the LAD between Week 6 and Week 13. However, from Week 13, the number of interactions falls over time. We postulate that the higher interactions since the LAD's introduction are due to teams familiarizing themselves with the tool. As a result, they spend less time on the LAD, as they have learned to navigate in fewer interactions.

4 CONCLUSION

The study shows that the implemented LAD engages most teams in a K-12 team-based learning environment. There is value in collecting trace data logs from dashboards the same way as trace data logs are collected from learning management systems (LMS) that researchers and educators use to quantitatively support students learning, design better systems and provide personalization. The trace data gathered from LADs offers opportunities to analyze team learning processes, which are currently a challenging issue to address in collaborative learning research.

REFERENCES

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 309–318.

Jivet, I., Wong, J., Scheffel, M., Valle Torre, M., Specht, M., & Drachsler, H. (2021). Quantum of Choice: How learners' feedback monitoring decisions, goals and self-regulated learning skills are related. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 416–427.

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2015). LATUX: An Iterative Workflow for Designing, Validating and Deploying Learning Analytics Visualisations. *Journal of Learning Analytics*, *2*(3), 9–39.

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.

Using Gameplay Data to Investigate Students' Problem-Solving Behaviors in Zoombinis

Tongxi Liu
University of Florida
tongxiliu@ufl.edu

ABSTRACT: One of the most critical intellectual abilities in today's K-12 education is problem-solving skills. In game-based learning environments, it is essential to analyze students' problem-solving behaviors so that instructors and researchers can provide timely feedback and relevant interventions. However, it is difficult to comprehend how students solve problems in games because game-based learning is always implicit and complicated. This study aims to investigate students' problem-solving behaviors in a specific puzzle-based game context. Specifically, we used the Continuous Hidden Markov Model (CHMM) to identify students' problem-solving stages, as well as sequence analysis to examine their problem-solving strategies. According to the findings, CHMM performed well to determine students' problem-solving stages, and sequence analysis efficiently identified five problem-solving strategies. By applying the integrated techniques, we automatically categorized students' real-time problem-solving behaviors and examined frequent strategies they applied, which can be used to better understand students' problem-solving processes and provide real-time interventions for future studies.

Keywords: Hidden Markov Model, Sequence analysis, Problem-solving behaviors

1 INTRODUCTION

Problem-solving is related to many cognitive abilities, such as reasoning and analysis (İncebacak & Ersoy, 2016). It requires individuals to adapt prior knowledge and experience to new problems and design correct solutions (Lishinski et al., 2016). Given the ever- changing education and career landscape, problem-solving ability is a crucial competency for students in the twenty-first century (Khoiriyah & Husamah, 2018). Game-based learning environments have been shown to increase students' engagement and achievement (Plass et al., 2015), in which the tasks are constructed with complex problems. It enables students to engage in a variety of problem-solving scenarios. Since the epidemic of COVID-19, game-based learning has been viewed as an efficient setting that allows most students to acquire multiple skills by solving ill-structured problems (Wati, 2020). However, analyzing students' problem-solving behaviors in game-based learning environments remains difficult. Students' problem-solving in games is dynamic (Csapó & Molnár, 2017) and implicit (Anderson, 2012), making it challenging to uncover their problem-solving stages and strategies. Surveys and interviews lack abilities to determine real-time stages in students' problem-solving processes, and therefore the feedback is lagged. Furthermore, manually labeling students' problem-solving behaviors would be extremely time-consuming, especially on a large scale. This study applied data mining approaches to investigate students' problem-solving behaviors in a game-based learning environment. Specifically, we used CHMM to identify students' problem-solving stages, followed by sequence mining to gain further insights into problem-solving strategies they employed.

2 METHODOLOGY

In this study, the learning setting is a puzzle game situated in problem-solving scenarios called **Zoombinis** (Asbell-Clarke et al., 2021). It includes 12 puzzles in which students bring characters to safety by applying efficient strategies (Rowe et al., 2017). Data was collected from 158 students in grades 3-8 (88 males, 70 females) who played with a specific puzzle called **Pizza Pass**. In this puzzle, students must find the combination of pizza (and at higher level ice cream) toppings to unlock the characters' path. There were 11014 attempts generated from students' gameplay data in total. CHMM modeled students' each attempt as one specific problem-solving stages, while sequence mining identified frequent patterns among these attempts. All the data were processed with Python codes.

3 RESULTS AND DISCUSSION

Table 1 shows several examples illustrating how to identify students' problem-solving stages by CHMM. Each attempt was classified as the corresponding stage *i* where its log-likelihood was the largest (λ_T : Trial and Error; λ_S : Systematic Testing; λ_I : Implement Solution; λ_G : Generalize Solution). To validate the accuracy of the developed model, 10-fold cross-validation was applied. CHMM had 93.53% agreement with human labels, performed an ROC/AUC score of 0.76.

Table 1: Examples of the identification results.

Human labels	CHMM	Log-likelihood			
		λ_T	λ_S	λ_I	λ_G
Trial and Error	Trial and Error	-3.34	-6.75	-5.34	-14.74
Systematic Testing	Implement solution	-3.57	-2.09	-2.05	-3.31
Implement solution	Implement solution	-3.40	-1.78	-1.74	-2.84
Generalize Solution	Generalize Solution	-4.72	-2.28	-2.77	-1.90

According to CHMM, students' gameplay attempts can be categorized as different problem-solving stages, which helps us better understand students' problem-solving processes and locate the struggling moments when students were stuck in a certain stage. Besides, CHMM suggests that the transition between two stages is likely to occur when the log-likelihoods are close. In summary, CHMM not only efficiently labels students' problem-solving stages but also indicates the probabilities of transitions happening.

Table 2: Patterns of problem-solving strategies.

Strategy	Patterns	Number of students
Testing one	Select 1→Deliver→Select 3*→Deliver	158
Additive	Select 1→Deliver→Select 1→Select 2→Deliver**	127
Replacing	Select 1→Select 5→Deliver→Select 1→Select 4→Deliver	93
Winnowing	Select 1→ Select 2→Select 3→Deliver→Select 1→Select 2→Deliver	65
Subtracting	Select 1→ Select 3→Select 5→Deliver→Select 2→Select 4→Deliver	52

* Select pizza topping 3.

** Deliver pizza toppings 1 and 2.

Sequence mining techniques discovered five problem-solving strategies: test pizza toppings one by one (*Testing one*); combine one untested and all other correctly tested pizza toppings (*Additive*); replace one pizza topping while keep others remained (*Replacing*); remove selected pizza toppings one by one (*Winnowing*); select the complement of previous pizza toppings (*Subtracting*). Table 2 shows the patterns and the number of occurrences of each strategy. Based on the findings, *Testing one* is the most commonly used strategy, which may efficiently help students get out of loops and facilitate their systematic problem-solving processes.

4 CONCLUSION

By applying CHMM and sequence analysis, we obtained a thorough depiction of students' problem-solving behaviors in **Zoombinis**. This integrated method can be utilized to analyze students' implicit problem-solving stages and discover effective problem-solving strategies in game-based learning environments. Revealing students' hidden problem-solving processes through their gameplay data can help with the development of exemplary learning protocol and the implementation of timely scaffolds, which is critical to game-based learning assessment and enhancement.

REFERENCES

Anderson, J. R. (2012). Tracking problem solving by multivariate pattern analysis and Hidden Markov Model algorithms. *Neuropsychologia*, *50*(4), 487-498.

Asbell-Clarke, J., Rowe, E., Almeda, V., Edwards, T., Bardar, E., Gasca, S., Baker, R.S. & Scruggs, R., (2021). The development of students' computational thinking practices in elementary- and middle-school classes using the learning game, Zoombinis. *Computers in Human Behavior*, *115*, 106587.

Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology*, *8*, 2022.

İncebacak, B. B., & Ersoy, E. (2016). Problem solving skills of secondary school students. *China-USA Business Review*, *15*(6), 275-285.

Khoiriyah, A. J., & Husamah, H. (2018). Problem-based learning: Creative thinking skills, problem-solving skills, and learning outcome of seventh grade students. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *4*(2), 151-160.

Lishinski, A., Yadav, A., Enbody, R., & Good, J. (2016, February). The influence of problem solving abilities on students' performance on different assessment tasks in CS1. In Proceedings of the 47th ACM technical symposium on computing science education (pp. 329-334).

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258-283.

Rowe, E., Asbell-Clarke, J., Cunningham, K., & Gasca, S. (2017, October). Assessing implicit computational thinking in zoombinis gameplay: Pizza pass, fleens & bubblewonder abyss. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play* (pp. 195-200).

Wati, I. F. (2020, December). Digital Game-Based Learning as A Solution to Fun Learning Challenges During the Covid-19 Pandemic. In *1st International Conference on Information Technology and Education (ICITE 2020)* (pp. 202-210). Atlantis Press.

Encoding students reading characteristics to improve low academic performance predictive models

Erwin D. Lopez Z., Tsubasa Minematsu, Taniguchi Yuta, Fumiya Okubo, Atsushi Shimada
Graduate School of Information Science and Electrical Engineering, Kyushu University
lopez.zapata.erwin.242@s.kyushu-u.ac.jp, atsushi@limu.ait.kyushu-u.ac.jp

ABSTRACT: The emergence of digital textbook reading systems and their ability of recording reader interactions has opened the possibility of modeling students reading characteristics. Based on the modeling approaches of previous works, this poster integrated an unsupervised learning model and a low academic performance predictive model. Our results suggest that this new model is able to encode the students' reading characteristics and better understand the relationship between them and low academic performance, achieving better predictions.

Keywords: Digital textbook, educational data encoding, academic performance prediction

1 INTRODUCTION

The Bookroll application is an e-book reading system that provides students with different types of online interactions. Moreover, all of these interactions are recorded in a database, leaving open the possibility to analyze these data (Flanagan & Ogata, 2017). As a result, different works have carried out compelling analyses identifying students' clusters and tendencies (Akçapinar et al., 2020; Yang et al., 2019) by processing these data with unsupervised machine learning (UML) models, while other works have attempted to use the Bookroll data to predict the students' low academic performance (Cheng et al., 2021; Okubo et al., 2018) with the use of supervised machine learning (SML) models.

In this context, the present work aims to explore whether or not a low academic performance predictive model can benefit from both UML and SML models. The intuitive idea is that non-labeled data used by UML models preserve the general reading characteristics while labeled data used by SML models record the relationship between these characteristics and academic performance.

2 METHOD

In the first stage, we trained a UML model to learn how to encode the reading characteristics. We then defined two SML models, one as a baseline and another to be integrated with the UML model. Finally, we trained these SML models and compared their validation results. We used two different datasets to take advantage of the larger amount of data of unlabeled datasets.

2.1 First Stage: Unsupervised model

We created a dataset containing 9,885,286 logs of unlabeled data collected from 3124 university students enrolled in 41 different courses from the period of April 2019 to April 2021. These data were converted into 3679 samples, each one containing all the actions performed in the system by a student in a course. In contrast to previous works, we considered all the possible interactions that a student could perform in the reading system (Table 1). As a result, each sample was represented as a 37-dimensional vector. Finally, we split this dataset into train data (67%) and validation data (33%).

We carried out several exploratory experiments for finding the optimal unsupervised models that could encode the information of our dataset with the support of the optimization framework Optuna (Akiba et al., 2019). The considered models consisted of clusterization models, PCA, Deep Autoencoders, Sparse Autoencoders, Variational Autoencoders (VAE), and β-VAE (Kingma & Welling, 2013; Higgins et al., 2017) while the evaluation measure was the reconstruction MSE. As a result, the selected model was the β-VAE (β =0.8) with an 8-dimensional latent space shown in Figure 1a.

Table 1: Summary of the considered features.

Type of features	# Features	Examples of the considered features
Device stamps	3	DESKTOP, TABLET, MOBILE
Frequent interactions	8	OPEN, NEXT, PAGE_JUMP, CLOSE_TIMEOUT, READ_TIME
Unfrequent interactions	26	ADD_MARKER, ADD_MEMO, SEARCH, GET_IT, LINK_CLICK

2.2 Second stage: Supervised models comparison

The dataset was obtained from the train data of the LAK19 Data Challenge workshop. This dataset contains 1,914,680 logs corresponding to 1347 labeled samples collected from the period of April 2018 to August 2018. We split this dataset into train data (67%) and validation data (33%).

Finally, we defined the two SML models, each consisting of a Deep Neural Network with 3 layers and one-dimensional output. The first (Figure 1b) represents our baseline since its architecture is similar to previous works. The second, as shown in Figure 1c, was connected with the encoder part of the UML model of section 2.1 and should learn to map the encodings to a low score probability.

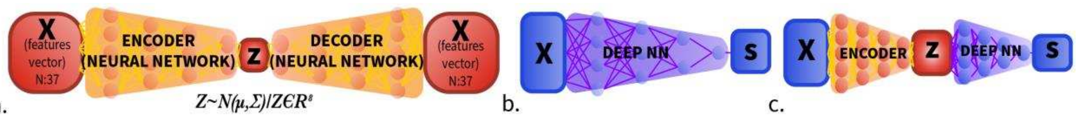
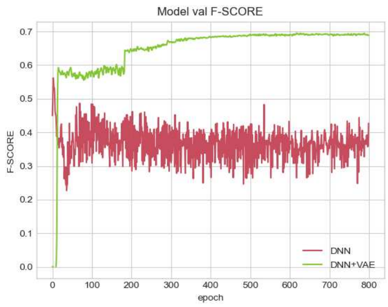


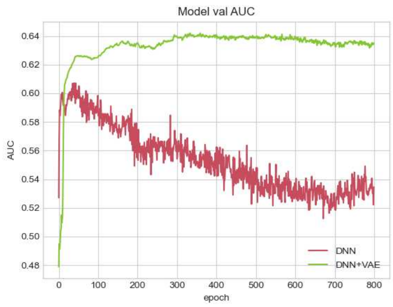
Figure 1: a. Unsupervised β-VAE model. b. First supervised model. c. Second supervised model.

3 RESULTS AND DISCUSSION

As shown in Figure 2, the baseline model exhibited an unstable and declining validation performance indicating a tendency in overfitting to the training dataset. On the other hand, the proposed model not only exhibited a more stable validation performance but also outperformed the baseline model. Nevertheless, it should be noted that due to the probabilistic nature of the β-VAE used to encode the data, our model had the chance of assigning low score and non-low score predictions to different points of the same probabilistic distribution of a single student. In summary, while the observed performance improvement should not ve overestimated, we point out the capability of our model to generalize the prediction results to the validation dataset and its stability, qualities that suggest that the unsupervised model part could encode the students' reading characteristics allowing the supervised model part to better interpret the relations present in the labeled data.



a.



b.

Figure 2: a. Comparison of F-score validation results. b. Comparison of AUC validation results.

Finally, the proposed model opens the possibility to conduct different studies. For example, with a similar approach to previous works that used UML models, researchers could identify the existent shades of the spectrum between two different study approaches instead of only clusters. Also, they could study the relationships between these shades and students' low score probability by exploring the meaning of the latent space axes from the proposed model UML part and then the importance that its SML part attaches to each of these axes when making a prediction.

ACKNOWLEDGMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grant Number JP18H04125, Japan.

REFERENCES

Flanagan, B., Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. In International Conference on Computers in Education (ICCE2017) (pp.333-338).

Akçapınar, G., Mei-Rong A. C., Majumdar, R., Flanagan, B. & Ogata H. (2020) Exploring student approaches to learning through sequence analysis of reading logs. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20).

Yang, C., Flanagan, B., Akçapınar, G. & Ogata, H. (2019) Investigating Subpopulation of Students in Digital Textbook Reading Logs by Clustering. In International Conference on Learning Analytics and Knowledge (LAK '19).

Chen, C., Yang, S.J., Weng, J., Ogata, H., & Su, C. (2021). Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers. Australasian Journal of Educational Technology, 37(4), 130-144.

Okubo, F., Yamashita, T., Shimada, A., Taniguchi, Y., & Konomi, S. (2018). On the prediction of students' quiz score by recurrent neural network. CEUR Workshop Proceedings, 2163.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In International Conference on Knowledge Discovery & Data Mining (KDD '19).

Kingma, D.P., & Welling, M. (2014). Auto-Encoding Variational Bayes. CoRR, abs/1312.6114.

Higgins, I., Matthey, L., Pal, A., Burgess, C.P., Glorot, X., Botvinick, M.M., Mohamed, S., & Lerchner, A. (2017). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework (ICLR2017).

What are we measuring?: A topic modeling framework to map professionalism aspects to responses in a Situational Judgment Test

Cole Walsh[1], Alexander MacIntosh[1], Okan Bulut[2], Jinnie Shin[3]
[1]Altus Assessments Inc., [2]University of Alberta, [3]University of Florida
cwalsh@altusassessments.com, amacintosh@altusassessments.com, bulut@ualberta.ca, jinnie.shin@coe.ufl.edu

ABSTRACT: Situational Judgment Tests (SJTs) are widely used as high-stakes assessments to measure decision-making capabilities and professionalism in various contexts. In recent years, constructed-response SJTs have gained attention for allowing the examinee to provide nuanced responses that draw on their own experiences, opening new avenues for measuring professionalism. Understanding the examinee's competency in different aspects of professionalism is of high priority for educators and professional organizations. However, there is currently no way to systematically characterize responses into specific aspects of professionalism such as collaboration or empathy. In this study, we present the development of an unsupervised topic modeling framework for interpreting examinees' responses to a high-stakes SJT administered nationally to health professions schools in Canada and the United States. The approach involves mapping specific professionalism aspects to unsupervised topics derived from constructed responses. Preliminary results show that aspect definitions and responses can be mapped to the same topics and used to facilitate deeper understanding of examinees' professionalism.

Keywords: Assessment, Situational Judgment Test, Topic modeling, Constructed-response

1 INTRODUCTION

Situational Judgment Tests (SJTs) are increasingly adopted in high-stakes assessments and evaluations. In SJTs, examinees are shown hypothetical real-life scenarios and assessed on their interpretation of the scenarios and their suggested courses of action (Corstjens *et al.*, 2017). SJTs are particularly useful for assessing professionalism and interpersonal skills encompassing multi-dimensional and overlapping aspects including collaboration, motivation, and ethics (Mahon *et al.*, 2013). These aspects are frequently identified as priorities in Health Professions Education and numerous professional governing bodies emphasize the importance of multifaceted roles for healthcare professionals (Yaszay *et al.*, 2009; Frank & Langer, 2003).

While SJTs traditionally use a closed-response format (e.g., multiple choice), constructed-response SJTs have gained popularity in recent years. SJTs with constructed-response items allow examinees to draw from their experiences and explain in their own words what type of action they would take for each hypothetical scenario. Constructed-response SJTs also allow multiple aspects of professionalism to be measured simultaneously while placing minimal restrictions on how examinees can respond (i.e., examinees are not constrained to speak to any specific aspect of professionalism). This freedom poses a challenge for evaluating examinees' responses, hence, we need a way to reliably and systematically characterize specific aspects of interest (e.g., collaboration). The need for a systematic approach to characterize responses becomes even more critical with the exponential growth of data and information gathered from constructed-response SJTs. In this study, we present a natural language processing (NLP) framework using topic modeling approaches that disambiguates the underlying structure of examinees' responses in SJTs. We use this framework to map specific professionalism aspects to unsupervised topics derived from responses to a nationally administered SJT, *Casper* (Dore *et al*., 2017). Casper contains 12 sections — four written prompts and eight video-based dilemmas — where each examinee has five minutes to respond to three questions related to the prompt or dilemma. Casper is typically required by medical schools in Canada and the United States as a measure of non-cognitive skills. Approximately 160,000 examinees completed Casper in the 2021-2022 admissions cycle.

2 ANALYSIS FRAMEWORK

We analyzed 1,876 written responses to a single Casper scenario where examinees were presented with a real-life scenario that involved an aggressive player on a school volleyball team. Examinees filled the role of a teammate. We investigated the examinees' responses to the prompt: *"When building a team, what is the most important quality that every team member should have? Explain your reasoning."* The questions were designed to measure certain facets of interpersonal skills, such as *collaboration* and *problem-solving* aspects of professionalism.

We first conducted standard text preprocessing (*expanding contractions, removing stopwords and punctuation, and lemmatization)* to remove variational word forms from responses and increase the accuracy and performance of our topic models. Next, we used a TF-IDF vectorization with non-negative matrix factorization (NMF) on examinees' responses to disambiguate the underlying topics. The resulting topic model was then applied to the definitions of ten aspects of professionalism developed alongside Casper content. This procedure allowed us to characterize the alignment between predefined professionalism aspects and the examinees' responses. Figure 1 provides a conceptual representation of our modeling framework.

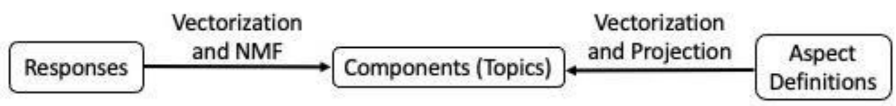


Figure 1: Summary of modeling framework used to analyze Casper responses and CANMEDS aspects.

3 RESULTS

Figure 2 provides an example alignment between predefined professionalism aspects and two of the topics generated from the examinees' responses. Each topic was represented by a set of representative keywords based on their topic-word distributions. For Topic 1, 'respect' weighted the highest, followed by other tokens, such as 'teammate' and 'mutual'. Another topic (Topic 2) had contributing tokens like 'goal', 'common', 'achieve', and 'reach'. Among the ten predefined aspects of professionalism, 'Empathy', 'Equity', 'Professionalism', and 'Ethics' corresponded to the content of Topic 1. Similarly, the 'Collaboration' aspect was most strongly mapped to Topic 2.

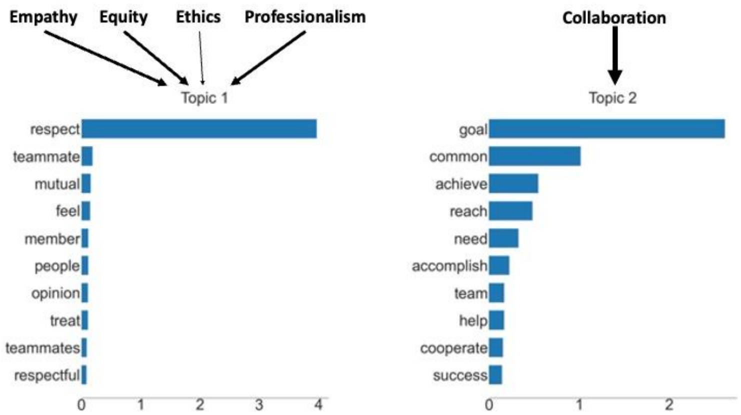


Figure 2: Example of alignment between professionalism aspects and two topics. Topic 1 encompasses multiple professionalism aspects, while Topic 2 largely captures a single aspect. Arrows between aspects and topics are proportional to the model coefficients (aspects with small coefficients have been dropped for clarity). The x-axis measures the token weights for each topic.

4 DISCUSSION AND FUTURE WORK

Our analysis introduced a novel element to the topic modeling process wherein aspects of professionalism, such as empathy and collaboration, were mapped to topics derived from constructed responses to an SJT. Future work will examine whether the methods used here can also motivate other constructs to characterize responses beyond the fixed set of professionalism aspects considered in this study. We also plan to expand on the professionalism aspect definitions used here and develop a corpus that represents multiple health professions' education priorities (e.g., Medicine, Physician Assistants, Nursing). The next stage of this project will also include further validation studies to evaluate the quality of generated topics as they relate to human-labelled data and examinees' scores on Casper. This work will ultimately inform inclusive evaluation practices to better understand how learners interact with constructed-response professionalism SJTs like Casper.

REFERENCES

Corstjens, J., Lievens, F., & Krumm, S. (2017). Situational judgment tests for selection. The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention, 226-246.

Dore, K. L., Reiter, H. I., Kreuger, S., & Norman, G. R. (2017). CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. *Advances in Health Sciences Education*, *22*(2), 327-336.

Frank, J. R., & Langer, B. (2003). Collaboration, communication, management, and advocacy: teaching surgeons new skills through the CanMEDS Project. World journal of surgery, 27(8), 972-978.

Mahon, K. E., Henderson, M. K., & Kirch, D. G. (2013). Selecting tomorrow's physicians: the key to the future health care workforce. Academic Medicine, 88(12), 1806-1811.

Yaszay, B., Kubiak, E., Agel, J., & Hanel, D. P. (2009). ACGME core competencies: where are we?. *Orthopedics (Online)*, *32*(3), 171.

Assessing the Integration of United Nations Sustainable Development Goals in a University General Education Curriculum

Chi-Un Lei *^, Xinyi Liang *, Chun Yin Cham *, Xue Qian *, Xiaodong Hu *
* The University of Hong Kong
^ University of Illinois Urbana-Champaign
* culei@hku.hk, cindylxy@hku.hk, tcycham@hku.hk,
qianxue@connect.hku.hk, u3586792@connect.hku.hk
^ clei7@illinois.edu

ABSTRACT: Higher education plays an essential role in achieving United Nations Sustainable Development Goals (SDGs). However, there are only scattered studies on monitoring how universities holistically promote SDGs through their curriculum. The main purpose of this study is to investigate the connection of existing general education courses in a university to SDG education. In particular, we want to know how general education courses can be classified according to SDGs. In this poster paper, we use machine learning approaches to tag the 167 general education courses in a university with SDGs, then analyze the results based on visualizations. Our training dataset comes from the OSDG public community dataset which the community had verified. Meanwhile, the learning outcomes and descriptions of general education courses had been used for the classification. We use the multinomial logistic regression algorithm as the algorithm and for the classification. Examples of calculated SDG probability of courses and the overall curriculum were used to illustrate the proposed approach's functions.

Keywords: Sustainable Development Goals, Classification, Curriculum Analysis

1 INTRODUCTION

1.1 Background

In 2015, United Nations established 17 sustainable development goals (SDGs) on sustainable economic growth and social development. Higher education plays an essential role in achieving SDGs. Several studies (Goodall, 2019) have been conducted for monitoring how universities achieve SDGs via campus infrastructure development (Omazic, 2021) and research (Bautista-Puig, 2021). However, there are only scattered studies on monitoring how universities holistically promote SDGs through their curriculum. It could be challenging to match SDGs with courses in practice. Course teachers may have difficulties understanding all 17 SDGs. Meanwhile, it is time-consuming for SDG consultants to analyze every piece of course content in hundreds of courses.

1.2 Objectives and Research Questions

The main purpose of this study is to investigate the connection of general education (GE) courses in a university to SDG education. By categorizing the extent of coverage of various SDGs in the courses, it would then be possible to adjust courses to ensure a well-rounded curriculum is provided to students. Through the investigation, teachers can be more informed to develop GE courses with consideration to SDGs. Students can also be benefited by selecting relevant GE courses for studying based on their SDG interests. The research question is: How can GE courses be classified through machine learning algorithms according to SDGs?

2 RESEARCH METHOD

This study tags SDGs for 167 GE courses offered at the University of Hong Kong between 2020-2021. We use the crawler method to extract the course learning outcomes and course descriptions from the GE course portal as our research dataset. Meanwhile, our training dataset comes from the OSDG Community dataset (OSDG, 2021). The dataset is based on publicly available documents, including reports, policy documents and publication abstracts. These documents are mainly from United Nations and often already have SDG labels associated with them. The OSDG community extracted records from these documents. Currently, there are around 32000 records of text comprised of 3 to 6 sentences. More than 1000 community volunteers then validated the records on the relevance to SDGs. The dataset only includes SDGs between 1- 15 because SDGs 16 and 17 are overarching goals that might pop up in almost all kinds of texts.

In the machine learning process, we first need to pre-process data, convert words of each document to numeric feature vectors, and use feature vectors and labels, for the predictive model. After pre-processing, we use frequency-inverse document frequency (TF-IDF) to do the feature extraction and the KBest algorithm to select features. The third step is classification. We use the multinomial logistic regression algorithm and use the Scikit-learn package for the classification.

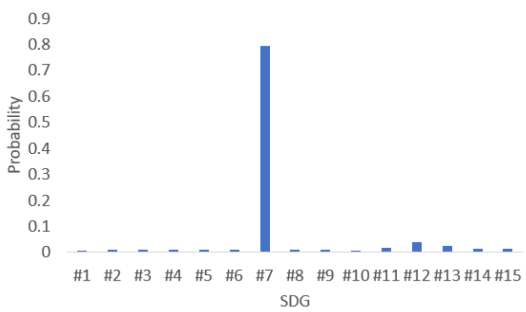


Figure 1: Calculated probability of CCST9016

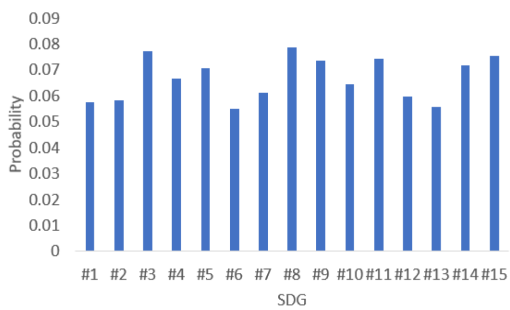


Figure 2: Calculated probability of CCCH9025

3 RESULTS

3.1 Overall

Figures 1 and 2 show the tagging results of two GE courses. For CCST9016 (Energy: Its Evolution and Environmental Impacts) in Figure 1, the calculated probability of this course being tagged to SDG 7 (affordable and clean energy) is about 0.8. So, this course should be tagged to SDG 7. On the other hand, For CCCH9025 (Humanity and Nature in Chinese Thought) in Figure 2, the probability of this course is relatively average for each SDG. Therefore, we believe it is not relevant to any SDGs (SDGs 1-15). Figure 3 shows the distribution of the average SDG calculated probability of courses through the raw classification, which approximately indicates what SDG content do students learn through a GE course. Among courses, SDG 3 (Good Health and Well-being), SDG 8 (Decent Work and Economic

Growth), SDG 9 (Industry, Innovation and Infrastructure), and SDG 11 (Sustainable Cities and communities) are the most mentioned SDGs in courses.

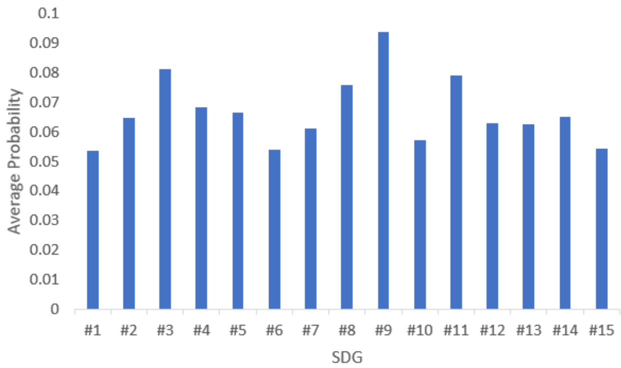


Figure 3: Average SDG calculated probability of courses through the raw classification

3.2 Course Selection

Students should study at least one general course in each area cluster (Area of Inquiry, AoI). The analysis of SDGs can help them determine which courses in each AoI to choose if they want to study in-depth on an SDG. For example, if they are interested in SDG 3 (Good Health and Well Being), the recommended course selection can be CCHU9022 Journey into Madness (0.63), CCST9078 Health Literacy: Things to Know Before Consulting Dr. Google (0.36), CCGL9053 Suicide: Risks, Research, and Realities (0.24) and CCCH9039 Curing the Chinese: Medicine and Society in Modern China (0.20).

4 CONCLUSION

The study has demonstrated how SDG can be tagged via machine learning based on the public data verified by the community. Currently, we are i) exploring how courses can be clustered, such that we can identify mini-SDG-curriculum or concentration for students, ii) defining a more relevant test set for more accurate classification, and iii) comparing the difference between teacher classification, expert classification and algorithm classification.

REFERENCES

Bautista-Puig, N., Aleixo, A. M., Leal, S., Azeiteiro, U., & Costas, R. (2021). Unveiling the Research Landscape of Sustainable Development Goals and Their Inclusion in Higher Education Institutions and Research Centers: Major Trends in 2000–2017. *Frontiers in Sustainability, 2*, 12.

Goodall, M., & Moore, E. (2019). Integrating the Sustainable Development Goals into teaching, research, operations, and service: A case report of Yale University. *Sustainability: The Journal of Record, 12*(2), 93-96.

Omazic, A., & Zunk, B. M. (2021). Semi-Systematic Literature Review on Sustainability and Sustainable Development in Higher Education Institutions. *Sustainability, 13*(14), 7683.

OSDG, UNDP IICPSD SDG AI Lab, & PPMI. (2021). OSDG Community Dataset (OSDG-CD) (2021.09) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5550238

Understanding College Students' Self-Regulated Learning Using Process Mining

Linyu Yu, Matthew L. Bernacki, Jeffrey A. Greene, Robert D. Plumley, Rebekah F. Duke, Rebekah Freed, Kathleen M. Gates, Abigail T. Panter
The University of North Carolina at Chapel Hill
Emails: linyu@live.unc.edu; mlb@unc.edu; jagreene@email.unc.edu

ABSTRACT: **Poster**. Self-Regulated Learning (SRL) has been increasingly viewed in terms of events that learners execute rather than traits that learners possess. These events reflect a process that unfolds over time in an ordered event sequence, as described in SRL models. We used process mining techniques to analyze undergraduates' verbalizations of their learning events when completing a lesson drawn from an introductory biology class. Sequences of events within successful learners' processes aligned closely to theorized models of SRL processes. Successful and less successful learners differed in how they approached their reading, planned their learning, and engaged in assessment strategies. Successful learners engaged in more complex cyclical patterns of SRL events than less successful learners. These findings provide a basis for the design of scaffolds to support students' adoption of productive SRL processes in college STEM course contexts.

Keywords: Self-Regulated Learning, Process Mining, Think-aloud

1 SELF-REGULATED LEARNING AND PROCESS MINING

Self-Regulated learning (SRL) theories (e.g. Winne & Hadwin, 1998) describe a loosely sequenced, cyclical process comprising the cognitive, metacognitive, motivational, affective, and contextual factors that occur during learning as students pursue valued goals (Schunk & Greene, 2018). Increasingly, SRL has been viewed in terms of actions rather than dispositions, thus It is critical to examine the sequence of these events and the cognitive and metacognitive processes they reflect.

Process mining (PM) is an analytical method that can be applied to sequential SRL event data and involves process discovery, conformance checking, and model enhancement (Van Der Aalast, 2012). Educational researchers have been using it to study SRL, given one assumption of PM is that there is a latent model in the event data (Bannert et al., 2014), and this latent model can be aligned with SRL theory. Our aim in this study was to understand the SRL processes undertaken by more and less successful students, so that we might determine whether the theorized sequence of processes in SRL frameworks (e.g., Winne & Hadwin, 1998) was present in the data and more commonly undertaken by those who performed well in a complex task compared to those who did not.

2 CURRENT STUDY

2.1 Participants and Method

Undergraduate students (N=49) from a large public university in the Southeastern U.S. were recruited from their introductory biology course and completed one lesson from the course in the laboratory. All course assignments involved preparatory reading while answering ungraded guided reading questions and then the completion of a graded pre-lecture homework. We selected 10 *Successful* learners who scored 5 out 5 in their homework and 10 *Less Successful* learners who scored 3 or less out of 5. During the study, students were asked to think aloud, and then trained researchers coded their verbalizations using micro codes from a codebook (Table 1). These micro codes were nested in macro codes. Then, coded events were analyzed using Fuzzy Miner algorithm (Günther & Van Der Aalst, 2007) in ProM 6.10. Node cutoff was set to 0 and Edge cutoff set to .2.

Table 1: Selected TAP Macro and Micro codes (Greene & Azevedo, 2009)

Macro:	Task Codes	Monitoring	Domain-General Strategies	Assessment Strategies
Micro:	Recycling	Content Evaluation	Taking notes	Ruling out answers
	Task definition	Judgment of Correctness	Self-testing	Matching

2.2 Results and Discussion

Descriptively, Less Successful students engaged in more codable TAP strategies overall, and Successful students completed the task in less time overall (Table 2). Successful students also allocated less time on Task Codes and Monitoring, and more time on Reading. In summary, the frequency and duration of SRL events differed by group, but these events alone did not explain achievement difference.

Table 2: Occurrences and Durations of TAP Macros and Reading of Two Groups

TAP Macro	Successful students (n=10)		Less Successful students (n=10)	
	Occurrences	Duration (unit: sec)	Occurrences	Duration (unit: sec)
Task Codes	89 (15.98%)	619.08 (9.34%)	143 (17.90%)	1,138.59 (15.20%)
Monitoring	156 (28.01%)	451.97 (6.82%)	245 (30.66%)	763.38 (10.19%)
Domain-General	143 (25.67%)	1,614.46 (24.34%)	211 (26.41%)	1,931.96 (25.79%)
Assessment	41 (7.36%)	645.24 (9.73%)	65 (8.14%)	688.46 (9.19%)
Reading	128 (22.98%)	3,301.05 (49.78%)	135 (16.90%)	2,968.38 (39.63%)
Total	557	6,631.81	799	7,490.77

2.2.1 Process Mining Results

Process models for each group in Figure 1 illustrate some similarities across groups: reading showed the highest significance and Assessment strategies showed the lowest (i.e., number inside nodes range from 1.0 [high] to 0.0 [low], Günther & Van Der Aalst, 2007). A bidirectional relationship between reading and monitoring indicated both groups monitored while reading, but only Successful learners showed connections to the domain-general strategies that reflect active reading (i.e., re-reading, taking notes, forming new conclusions during reading) and have been shown to promote learning. In contrast, Less Successful Learners engaged in more instrumental reading, reflected by the successive process from Assessment strategies to Reading. This was likely induced by the need to complete guided reading questions, or to answer graded homework questions.

A second notable difference between groups was that Task Codes – reflecting learners' initial and ongoing consideration of the goal of the task and resources provided to pursue it – were less significant to the process model of Successful Learners but were bidirectionally connected with their monitoring events. This may reflect Winne and Hadwin's assertion that students frequently metacognitively monitor their task engagement in light of the conditions imposed by the task and standards for task engagement the learner sets, and choose their strategies accordingly. Thereafter, students may evaluate the influence of that strategy on progress towards the task goal and adapt accordingly. No such connections were observed for less successful learners. A third difference involved Assessment strategies, which were unconnected and less significant in the Successful Learner model compared to the Less Successful Learner model. Reading and assessment (i.e., homework completion) were distinct activities, whereas Less Successful learners iteratively returned to reading during homework, potentially to seek, confirm, or revise answers. Finally, the overall model for Successful learners more fully reflects the complexity proposed by SRL theorists, whereas the Less Successful student process model lacks connections between events that reflect essential assumptions of SRL frameworks (e.g., monitoring and control events; Winne & Hadwin, 1998).

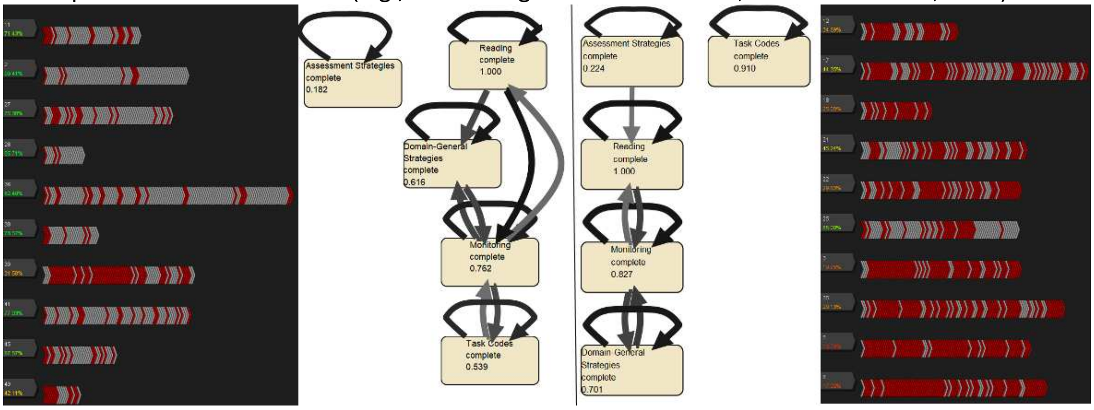


Figure 1: Successful (L) and Less Successful (R) learners' SRL sequences and trace conformances

We conducted conformance checking to gauge the models' *fitness* – the percentage of actual event sequences observed in logs that could be captured successfully by the model – similar to the notion of explained variance. The trace conformance (silver in Figure 2, dark panels) of Successful learners was much higher (M=70.22%, SD=18.94%) than Less Successful Learners (M=32.26%, SD=15.74%), suggesting Successful learners followed similar SRL models, whereas Less Successful learners had more idiosyncratic learning processes among their SRL event sets.

2.3 Conclusion, Limitation, and Future Directions

Even with a small sample and a coarse macro-level SRL taxonomy, process mining produced evidence that corroborated theoretical assumptions in ways that contributed beyond analyses of the frequency and duration of SRL events. The process models have potential implications for college STEM instructors to support less successful students. SRL processes can be mined and included in dashboards for instructors and students, and can inform prediction models that identify students whose learning process misaligns to SRL models, and who might benefit from additional support.

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. https://doi.org/10.1007/s11409-013-9107-6

Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, *34*(1), 18–29. https://doi.org/10.1016/j.cedpsych.2008.05.006

Günther, C. W., & Van Der Aalst, W. M. (2007). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. *International Conference on Business Process Management*, 328–343.

Schunk, D. H., & Greene, J. A. (2018). *Handbook of Self-Regulation of Learning and Performance* (2nd ed.). Routledge. https://doi.org/10.4324/9781315697048

van der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management of Information Systems, 3*(2). pp. 7:1–7:17

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

How do students perceive algorithmic grouping in higher education?

Author(s): Clara Schumacher and Jakub Kuzilek
Humboldt-Universität zu Berlin, clara.schumacher@hu-berlin.de

ABSTRACT: Collaboration in virtual teams is considered crucial in the 21st century. Student characteristics impact learning as well as collaboration. To facilitate grouping considering students' characteristics algorithmic approaches might be supportive. This proposal reports research findings from a qualitative study investigating students' perceptions of algorithmic group formation for a project assignment. Findings indicate that students perceive the approach resulted in fair grouping, a good fit of skills among members, and was beneficial as groups were more diverse.

Keywords: collaboration, group formation, algorithm, higher education

1 INTRODUCTION

A crucial skill in the 21st century is collaboration in virtual teams. During the pandemic situation virtual facilitation of collaboration for learning and teaching in higher education is increasingly relevant. Students' characteristics impact learning processes and outcomes (Nakayama et al., 2014) as well as their collaboration (e.g., dealing with learning tasks, peer interaction and task contribution) (Hadwin et al., 2017). Hence, group formation considering students' characteristics might be beneficial for collaborative learning and outcomes. For example, heterogeneity of extraversion and conscientiousness was found to have a positive impact on group processes and outcomes (Bellhäuser et al., 2018). Grouping students heterogeneously based on subject knowledge, motivation and demographics had a positive effect on emotions but not significantly on performance (Sun & Chiarandini, 2021). Technical solutions using algorithms are beneficial when considering multiple characteristics in larger cohorts to reduce teacher workload for grouping and finding optimal solutions.

Besides the impact that group formation might have on collaborative processes and outcomes it needs to be investigated how students perceive being grouped based on algorithms. Thus, this study used an algorithmic approach for forming groups based on self-reported characteristics and investigated students' perceptions using qualitative semi-structured guided group interviews to answer the following research questions: *How do students perceive being grouped by an algorithm? How do students evaluate the fit of their group members compared to the task requirements?*

2 METHODS

2.1 Participants and design

The study was conducted in a computer science master's course on social media and collaborative systems at a European university. Of the *N* = 44 students enrolled in the course, n_1 = 23 students participated voluntarily. The groups worked virtually on a programming task in which they could choose from a set of user stories to develop microservices targeting the coordination needs of NGO members. The project assignment lasted for two months in January and February 2021 including weekly online course sessions for presenting project progress, asking questions, and receiving feedback. Before the project assignment started participating students answered a questionnaire assessing their personality traits, learning and achievement goals, learning styles, preferences for group work, course-related self-efficacy beliefs, prior knowledge, interest in the course topic, and demographic data. The applied algorithm assigned the students into five groups (with 3-6 students per group) based on the self-reported data. The remaining students (n_2 = 21) in the course were randomly assigned to another five groups. After the project was finished students grouped by the algorithm participated in a qualitative interview study.

2.2 Grouping algorithm

The initially developed approach of Kardan and Sadeghi (2016) has been adapted. The algorithm processes students' self-reported data and combines them with teacher defined requirements on heterogeneity and homogeneity of the groups to be formed (Figure 1). First, the algorithm computes the so-called compatibility measure, which combines multiple attributes in one characteristic ranging from 0 to 1. The compatibility is calculated for each student pair. The resulting compatibility matrix, containing all student pairs' compatibilities, is used as the input to the binary integer programming optimization solver, which produces students' optimal grouping.

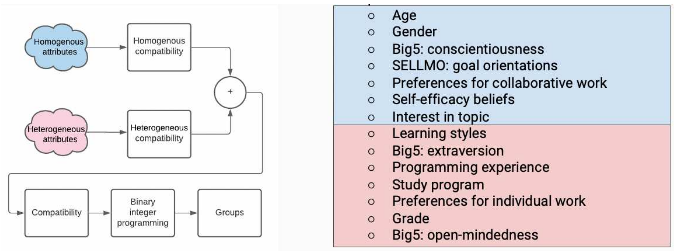


Figure 1: Algorithm workflow and student self-reported attributes for automatic grouping

2.3 Instruments

To investigate participants' perceptions of the grouping and the group work, qualitative semi structured guided group interviews were conducted after the project assignment. The interviews were conducted online and took between 30 and 45 minutes. Questions asked participants about the grouping process, the fairness of this approach compared to traditional approaches, work allocation among group members, available skills of the group members and how they met the task requirements plus if they perceived a learning gain regarding subject-related or collaborative skills. Interviews were transcribed and analyzed using qualitative contents analysis.

3 RESULTS

Regarding students' perceptions of being grouped by an algorithm (RQ1), results indicate that students liked to be grouped in more diverse teams with unknown peers that have different knowledge increasing their opportunities to learn new things. Some also mentioned that they would have preferred working with friends. Fairness was similar or better than random grouping given that questionnaires were answered honestly. It was further mentioned that they perceived that "someone took care to obtain a good match" (group 5). However, students asked for more insights into the algorithm's assumptions. In addition, participants stated that particularly in the online semester and when not knowing their peers the algorithmic grouping was efficient but would also be beneficial for large on-site courses and first year students.

Investigating the fit of the group members' skills to perform the task, four groups stated that their prior knowledge was sufficiently diverse to complement each other for achieving the task goal. Only one group had the impression of having less prior knowledge compared to others (group 4). Smaller groups of three stated to have less coordination tasks (group 1) but also perceived limitations of what they could achieve (group 2). In contrast, larger groups of five had difficulties to allocate responsibilities and faced higher coordination needs (group 3 and group 4). The size of four members seemed to be the optimum of coordination effort and sufficient capacities for implementing the programming task (group 5).

4 DISCUSSION

Using a qualitative approach enabled us to gain insights into students' perceptions of algorithmic grouping and their group work. Results indicated that a good fit was experienced when members had diverse task-related prior knowledge and experiences. In addition, a group size of four seemed to be suitable for the project assignment especially in virtual collaboration. Upcoming analyses will investigate students' digital collaboration process as indicated by their trace data and relate this to their reported self-regulation and the group performance in the project assignment. However, this study has limitations as the sample is rather small and from one institution only. Furthermore, as a larger cohort would facilitate the group formation this is currently investigated in a larger project. It should also be examined how to include students' preferences of working together with friends to the group formation. This approach might be a support for facilitators of larger courses, digital formats and first-year students for finding a suitable group and to avoid negative group experiences as reported by our participants.

REFERENCES

Bellhäuser, H., Konert, J., Müller, A., & Röpke, R. (2018). Who is the perfect match? Effects of algorithmic learning group formation using personality traits. *i-com*, *17*(1), 65-77.

Hadwin, A. F., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 83-106). Routledge.

Kardan, A. A., & Sadeghi, H. (2016). An efficacious dynamic mathematical modelling approach for creation of best collaborative groups. *Mathematical and Computer Modelling of Dynamical Systems*, *22*(1), 39-53. https://doi.org/10.1080/13873954.2015.1086382

Nakayama, M., Mutsuura, K., & Yamamoto, H. (2014). Impact of learner's characteristics and learning behaviour on learning performance during a fully online course. *The Electronic Journal of e-Learning*, *12*(4), 394-408.

Sun, Z., & Chiarandini, M. (2021). An exact algorithm for group formation to promote collaborative learning. In *LAK 21: 11th International Lerning Analytics and Knowledge Conference*. ACM.

Towards a Learning Analytics Metadata Model

Ian Wolff
University of Magdeburg
ian.wolff@ovgu.de

David Broneske
DZHW Hannover
broneske@dzhw.eu

Veit Köppen
ZLB
veit.koeppen@zlb.de

ABSTRACT: In learning analytics research, methods for data collection and the produced data are manyfold. To ensure the research data's findability, accessibility, interoperability, and reusability structured information along the research data life cycle is required. This paper outlines the development of LAMM – Learning Analytics Metadata Model – a metadata model designed to describe the generation of learning analytics research data. We develop this model in close collaboration with learning analytics researchers in the field of evaluation of collaborative programming scenarios. The main challenge is to provide all relevant information and important search criteria for reuse of data within a collaborative working space as well as for third-party scientists. Therefore, metadata capture information about observed learner, software environments, where learning takes place, as well as methods and measurement instruments used for data collection. Such metadata make the research data FAIR and enable a sustainable research data management for learning analytics.

Keywords: Research Data Management, Metadata, Open Science

1 INTRODUCTION

Learning analytics is a data-driven science and once data are generated a sustainable research data management is required that handles data creation and processing transparently and enables traceability and replication. Until now, there is no professional, established culture in learning analytics for storing and reusing accumulated data. Data management of learning analytics data comes with several challenges that hinder the development of a learning analytics open science culture and prevent researchers from sharing their data so far (Biernacka & Pinkwart, 2021). Nevertheless, data sharing is useful to give third parties the opportunity to use secondary analyses to make thematic and methodological comparisons with existing data while working on new research questions and contexts without generating data in the same context due to the complex and time-consuming data gathering processes (Pasquetto et al., 2017). One component of a sustainable research data management is structured data about the data, called metadata. They describe the structure of objects, and provide important administrative information on rights and property rights. Therefore, we look from a library and information science perspective on generated data in a learning analytics context to enhance data reusability for researchers within a collaborative working space and for third-party scientists.

2 COMPONENTS OF A LEARNING ANALYTICS METADATA MODEL

In (Wolff et al., 2021), we present a literature review on research methods and techniques used in learning analytics science of collaborative programming scenarios to deduce requirements for a learning analytics metadata model. From gathered data, we derive a model of objects of all research components involved in the learning analytics, cf. Figure 1. Furthermore, we develop and enhance

LAMM – Learning Analytics Metadata Model – on collaborative base in our project with fellow learning analytics researchers from a collaborative programming scenario context from three Universities within the DiP-iT[1] project. We consider existing standards, like Dublin Core, DataCite, and RADAR, but extend these by domain-specific metadata, to improve data reusability by providing information about data provenance.

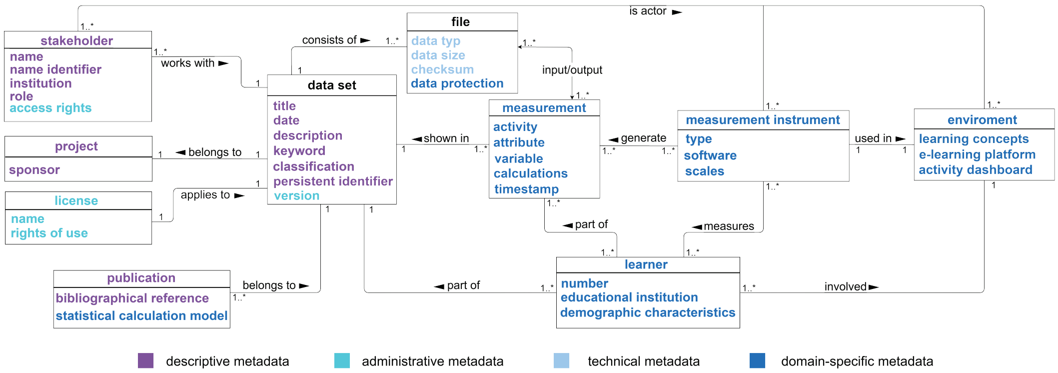


Figure 1: Model of objects with domain-specific characteristics for learning analytics

Our primary goal is to capture all relevant criteria along the learning analytics research data lifecycle (Wissik & Ďurčo, 2015) to make data comprehensible and reusable for third-party scientists. Therefore, we take the FAIR principles for "Findability, Accessibility, Interoperability, and Reusability" into consideration, which act as a guideline to enhance reusability of data holdings by using adequate metadata (Wolff et al., 2021). In the following, we categorize the metadata from our model of objects in Figure 1 according to their specific characteristics that make the corresponding research data FAIR and meeting researchers needs on research data management (Baca, 2008).

The first type is the **descriptive metadata**, which accomplishes researchers needs by increasing their data findability and identify and describe the dataset by giving related information about the resource. These are elementary entities like *title*, *keywords*, *description,* and related *persons* as well as the *subject* area of the data set. Furthermore, important for repository usage is a mapping of the data with an established *classification,* like the Dewey Decimal Classification (DDC) (Dewey, 1989), and with related *publications* with their *bibliographic references*.
Researchers also require easy access to their data, what we ensure with **administrative metadata.** It is of importance for learning analytics data, according to privacy and ethical issues (Pardo & Siemens, 2014). We use this type for managing and administering a dataset and ensure accessibility, mostly represented by the entity *license*, which manages to what extent the research data can be reused and accessed.
A data set can also consist of multiple files with file attributes like *name*, *type,* and *size* and can be enriched with a *checksum* or *persistent identifier*, which also ensures findability. We categorize such entities as **technical metadata**.
The last type is the **domain-specific metadata**, which is not a classic metadata type, but important for reusability and interoperability of a data set. It defines learning analytics characteristics along the

[1] https://www.dip-it.ovgu.de/

research data lifecycle. Here, the *learner* and the *environment* used for the analysis of the *activity* measurement and the *measurement instrument* are defined, as well as the *calculations* or *computation* used for analysis of the learning process. These entities allow third-party learning analytics researchers to understand the data provenance in detail, to reuse the data, e.g., for a new study.

3 CONCLUSION AND OUTLOOK

Research data management and data reusability is an emerging topic for all sciences. Currently, we are developing LAMM – Learning Analytics Metadata Model – (Wolff et al., 2022) for research data management. This basis builds already established metadata models, enriched with discipline-specific metadata to include data provenance and learning process measurement in detail. Therefore, the FAIR principles are adapted to meet the community standards and researchers needs. The metadata model will then be implemented in a research data repository and tested with different kinds of data from the universities in our project, and furthermore discussed with peers and continuously enhanced.

Acknowledgements

Funding: This work was supported by the German Federal Ministry of Education and Research [grant number 16DHB 3008].

REFERENCES

Baca, M. (2008). *Introduction to metadata* (2nd ed.). Getty Research Institute.

Biernacka, K., & Pinkwart, N. (2021). Opportunities for Adopting Open Research Data in Learning Analytics. In L. Tomei et al. (Eds.), *Advances in Educational Technologies and Instructional Design.* (pp. 29–60). IGI Global. https://doi.org/10.4018/978-1-7998-7103-3.ch002

Dewey, M. (1989). *Dewey decimal classification and relative index. 2*, Schedules 000-599. OCLC Library.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal, 16*, Article 8. https://doi.org/10.5334/dsj-2017-008

Wissik, T., & Ďurčo, M. (2015). Research Data Workflows: From Research Data Lifecycle Models to Institutional Solutions, *Linköping Electronic Conference Proceedings, 180*, 94-107. https://ep.liu.se/ecp/123/ecp15123.pdf

Wolff, I., Broneske, D., & Köppen, V. (2021). A First Metadata Schema for Learning Analytics Research Data Management. *O-Bib. Das Offene Bibliotheksjournal, 8*(4), 1–20. https://doi.org/10.5282/o-bib/5735

Wolff, I.; Broneske, D.; Köppen, V. (2021). FAIR Research Data Management for Learning Analytics. *Proceeding of the DELFI Workshop*, Germany, 158–163. https://repositorium.hs-ruhrwest.de/frontdoor/index/index/docId/733

Wolff, I., Broneske, D., & Köppen, V. (2022). *LAMM – Learning Analytics Metadata Model* (Version 0.1)[Data set]. https://doi.org/10.24352/ub.ovgu-2022-013

Exploring the use of GPT-3 as a tool for evaluating text-based collaborative discourse

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

Authors: Bradford Mott, James C. Lester
North Carolina State University
bwmott@ncsu.edu, lester@ncsu.edu

ABSTRACT: Natural language processing (NLP) models have previously been used to classify and summarize student collaborative actions in various online and computerized learning environments. However, due to limitations related to insufficient or inappropriate training data, these models are limited in their applications and impact. In this study, we explore how a new model, GPT-3, summarizes student chat in a computer-supported collaborative learning environment. With only a sentence explaining the context of the learning environment and two training examples, GPT-3 was able to effectively extract and summarize student conversations (properly attributing states such as frustration and confusion), reliably synthesize statements not present in the source text, and effectively ignore extraneous noise in the student chat. We discuss how this summarization could be used to support teachers understanding of student collaboration in computer supported collaborative learning environments.

Keywords: Natural Language Processing, Collaborative Learning, Discourse

1 INTRODUCTION

To support collaboration, instructors need to understand information about social interaction among students as they engage in computer-supported-collaborative learning, game-based learning, and other forms of instruction that focus on supporting group work. Researchers have used natural language processing (NLP) to analyze collaborative actions which are captured as speech or text (Blikstein & Worsley, 2016). However, most of these models require extensive training, and are mainly applied post-hoc, meaning they cannot be easily used for real-time orchestration. Even when sufficient previous data exists, these types of models tend to “fail non-gracefully” (Roschelle et al., 2020) when the context of the data changes slightly. Some researchers have used pre-trained NLP models to accomplish this task. However, the success of pre models has been limited because these models are often trained in a specific context, such as Wikipedia or Twitter data that doesn’t generalize well to student generated text (Phillips et al., 2021). A new model, GPT-3, has the potential to change this. GPT-3 was trained on a corpus of 410 billion tokens drawn from a common, largely indiscriminate crawl of the internet and is ten times large than any previous NLP model (Brown et al., 2020). This paper explores the potential of GPT-3 as a tool to summarize student chat in a computer-supported collaborative learning environment using only two training examples.

2 METHODS

The data analyzed in this study was collected from four 12-to-14-year-old students participating in a collaborative game-based learning environment designed to teach ecosystems concepts. In the learning environment, students are on a field trip to a fictional island in the Philippines and conduct investigations on why tilapia in the local fisheries are sick. They gather evidence, and then work together to determine the cause of the sickness. One of the major features of the game is a virtual whiteboard where students can drag-and-drop their collected evidence, organizing it by topic and potential explanations. Student discussion is supported by an in-game chat feature where students generate explanations, discuss evidence, and engage in group inquiry.

GPT-3 has been shown to be successful at NLP classification, summarization, and completion tasks with no or few training examples (Brown et al., 2020). To understand how it might be utilized in our game-based learning environment, GPT-3 was presented with the follow prompt:

> *Summarize the following conversations between four students, (Eagle520, Jeepney520, Sun520 and Turtle520) and their tutor (wizard520) who are playing an educational video game.*

We then primed GPT-3 by manually summarizing the first 10 lines of code in two chunks. This method could theoretically be implemented in practice, with teachers summarizing the first several lines of chat at the beginning of the class and allowing GPT-3 to continue during the implementation.

After being presented with these two training examples, the text of student chat from our data was split into groups of approximately 500 *tokens* to conform with the length limitations of GPT-3, or approximately 300-400 words. GPT-3 was asked to summarize these chunks. A total of 18 summaries were created over the 702 lines of chat.

3 RESULTS

Overall, GPT-3 was effective at summarizing chat conversations and was robust against previously unseen and unusual language. Additionally, GPT-3 could correctly attribute abstract concepts (e.g., spamming the chat, confusion, frustration) to specific students, even when the exact words of the summary did not appear in the chat. Because GPT-3 is text-bound, it does not have a robust understanding of context, and can make inaccurate inferences (see examples). This limited understanding of context means GPT-3 cannot always discern between content that is considered important in the specific context of a collaborative educational game and everyday conversation.

Example 1: *"Summary: Eagle520 is spamming the chat. Sun520 and Jeepney520 are upset with Eagle's behavior. Wizard520 asks the students to talk more respectfully as a team."* In this example, over half of the chat lines were the student's avatar of Eagle typing "REEEEE" into the chat. Both Jeepney and Sun express frustration with this, though Sun expresses this abstractly, saying: "can we kick this kid." The word "spam" appears within the text ("he kind of spam tho"), however, the phrase "spamming the chat" was generated by GPT-3.

Example 2: *"Summary: Wizard520 welcomes the students to the board. Wizard520 says they can get started by dragging and dropping notes from their notebook into the different columns. Sun and*

Jeepney are confused about how to do this." Despite not having been explicitly told about the format of the whiteboard, GPT-3 properly abstracts the concepts of notes dragged into columns on a board. Additionally, it correctly attributed an abstract state to Jeepney and Sun using a word not present in the text (confusion). Jeepney states: "We need help," while Sun states "I DONT SEE ANY COLLUMS."

Example 3: *"Summary: Wizard520 tells the students that they can't vote for their own notes. Eagle is having a mental breakdown. Sun thinks Turtle is a secret agent who is secretly not a turtle."* In this example, we see again how GPT-3's limited understanding of context means it is not capable of discerning between content-focused and social conversations. It nearly directly quotes eagle ("I WILL HAVE A MENTAL BREAKDOWN") and does not understand this is a hyperbolic expression of frustration. Additionally, it does not recognize the social nature of Sun and Turtle's conversation.

4 DISCUSSION

This paper explores the potential of GPT-3 as an aid in formative assessment of students in technology-enhanced learning environments. With only two training examples, GPT-3 was able to accurately summarize student chat in a collaborative game-based learning environment. It was able to accurately attribute abstract states to students, such as frustration and confusion, and behaviors such as "spamming the chat." Because of its limited understanding of the context of the student chat, GPT-3 does not discriminate between certain topics and does not always recognize hyperbolic statements. However, in the case of Example 3, this limitation could potentially prove invaluable in understanding the nature of collaboration. This is because it is possible that conversations that appear unproductive are instrumental in helping students regulate negative emotions (Author, 2011). Rather than labeling student actions along a spectrum of productive and unproductive behaviors, the model provides a simplistic description of the social situation. In doing so, the model allows the teacher to make inferences about the nature of student collaboration. The limitations of GPT-3 in summarizing the text in this study were relatively obvious and easily understood. If these limitations were properly expressed to teachers, this model could represent a substantial asset to teachers. Being able to view summaries of student conversation in real time would allow teachers to better allocate their attention to frustrated and confused students and gain a insight into students' progress in technology-enhanced learning activities. By placing the teacher as an intermediary between the model inference and pedagogical interventions, we could protect against the model failing non-gracefully (Roschelle et al., 2020). Overall, GPT-3 has the potential to markedly increase the accuracy and utility of NLP models in real-time analysis of educational data.

REFERENCES

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *ArXiv*.

Phillips, T., Saleh, A., Glazewski, K. D., Hmelo-silver, C. E., Lee, S., Mott, B., & Lester, J. C. (2021). Comparing Natural Language Processing Methods for Text Classification of Small Educational Data. *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge*.

Roschelle, J., Lester, J., & Fusco, J. (2020). *AI and the Future of Learning: Expert Panel Report Suggested Citation Acknowledgements* (Issue November). https://circls.org/reports/ai-report.

Controlling for Speededness in PANCE Examinees' Responses Using Change-Point Analysis

Sanaz Nazari
University of Florida
sanaznazari@ufl.edu

John B. Weir, Joshua Goodman
National Commission on Certification of Physician Assistants
johnw@nccpa.net, joshuag@nccpa.net

ABSTRACT: Commonly, time limits are a necessary part of every exam, and they may introduce an unintended influence such as speededness on the item and ability parameter estimates as they have not been accounted for while modeling the latent ability. The change-point analysis (CPA) method can be used to obtain more accurate parameters by detecting speeded examinees, the location of change-point, removing speeded responses, and reestimating parameters. In the current study, several examinees were detected as speeded across five sections of a 250-item exam, and two main patterns were observed. In addition, speededness was further investigated using response times (RTs) per item and two patterns were observed for examinees with a decrease in performance after the estimated change-point. Recommendations for practitioners, limitations, and future research were discussed in the conclusion section.

Keywords: Speededness, Change-point Analysis, False Discovery Rate

1 INTRODUCTION

Lu and Sireci (2007) defined test speededness as "the situation where the time limits on tests do not allow substantial numbers of examinees to fully consider all test items" (p. 29), so, the validity and reliability of results are in question. Shao et al. (2016) proposed a model based on change-point analysis (CPA) to separate speeded and non-speeded examinees. Compared to previous models, this model has the advantage of locating the change-point, which is so informative to determine test length. For example, if many examinees are speeded toward the end of the exam, this shows that the time limits are not compatible with the time required for each question. Therefore, regarding tests with strict time limits, we may change the number of items or test length. In cases that test duration is more flexible, we can recalculate the time allocated to each item. The second important advantage of this method is improving the ability and item estimates by removing the speeded responses and recalculating the parameters. In other words, we obtain a more accurate estimate of examinees' abilities which is very critical for deciding the mastery of examinees in a criterion-referenced test. This study aims at using CPA to elaborate more on its pros and cons. So, the objective of this study was to apply CPA to real data to evaluate change points pattern, test duration, and ability estimates precision. Following Shao et al.'s (2016) method to detect speededness, the research question is: Given examinees' response vectors, how does CPA detect speededness among examinees across all five sections of the Physician Assistant National Certifying Examination (PANCE)?

2 METHOD

Responses for 10,266 examinees who took PANCE in 2020 were collected and 38 responses were removed from the population due to different timing for examinees with disability. Among five parallel forms of the exam, form three with a sample size of 2,471 was selected for the current study. Examinees typically take PANCE upon graduation from PA programs at the National Commission on Certification of Physician Assistants (NCCPA) to be able to practice medicine. PANCE is administered at one session of five hours and 45 minutes, that is, five blocks of 60 questions with 60 minutes in addition to 45 minutes of break time. It consists of five parallel forms and each exam form contains 300 multiple-choice items across five sections. Out of 60 items per section, 10 items are experimental and are not scored. Regarding reliability-internal consistency of items, NCCPA reported a Cronbach's Alpha of 0.85 for form three in PANCE 2020. In our sample, the calculated Cronbach's Alpha was 0.83 which aligned with the NCCPA report. PANCE items are multiple-choice questions, but for the current study, items were dichotomously coded as correct and incorrect, and the Rasch version (Rasch, 1993) of Shao et al.'s (2016) formula was used to model the change-points. Rasch model was selected as the PANCE items, by design, mostly vary in difficulty, and discrimination parameter estimates are so small. R statistical software was used to do the analysis.

3 RESULT

To ensure a linear test, each section of the test with 50 operational items was considered as a separate test and the degree of speededness was compared among five sections of form three. Examinees could see any random order of section sequences; therefore, to make sure comparability of results across all sections, we checked the average difficulty of items per section to examine how items were distributed. The average difficulty across all five sections in form three was about 0.06. Out of 2,471 examinees in form three, 109, 86, 119, 105, and 134 were detected as speeded across five sections. CPA method detects significant changes in examinees' ability estimates before and after the estimated change-point. For that reason, across all speeded responses, we observed two main patterns: the first showed examinees performed well until the estimated change-point, and after that their performance significantly decreased. This pattern was compatible with what we expected about examinees answering items toward the end of the exam more quickly because of the imposed time limits. The second pattern was exactly the opposite. Examinees scored lower on the first part and then scored significantly higher on the second part after the estimated change-point. Across all sections, the proportions of first and second patterns were 56.4% and 43.6%, respectively. As shown in Figure 1, in section one, examinee 3352's performance dropped significantly after item 37 (estimated change-point), and examinee 266, scored higher after item 30.

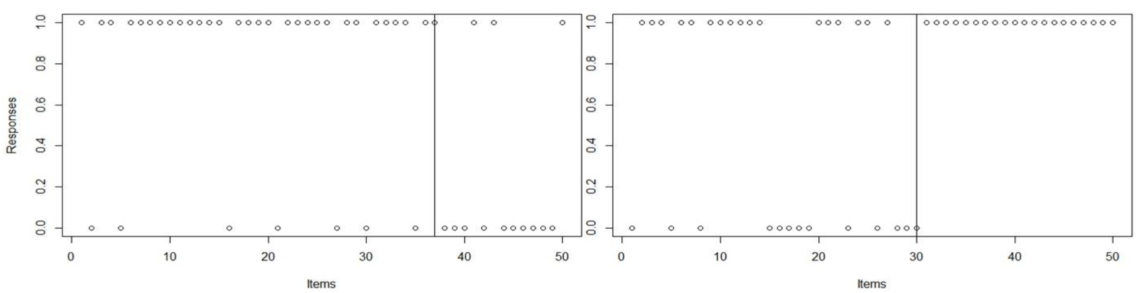


Figure 2: lower performance (left) vs. higher performance (right) after the estimated change-point.

We had access to examinees' response times (RTs) per item, and among several examinees with a decrease in performance after the change-point, two main patterns were observed. It seems that as examinees spent either much longer time or much shorter time than average on an item, the probability of getting an item incorrect, increased.

4 DISCUSSION AND CONCLUSION

By utilizing CPA, several examinees per section were detected as speeded. In their empirical research, Shao et al. (2016) did not report any low-to-high trends; this study is the first to show multiple observed patterns. This may be due to applying CPA to a different test within a different test setting. Consequently, this study helps to generalize CPA applications to multiple tests, and it extends the method, as well. The result of this study certainly shows that there is a need to improve CPA to detect multiple change-points across items. Possibly, a few change-points are needed to cover the warmup effect at the beginning (low-to-high pattern) of the test or other incidents during the exam. Other change-points may be needed to address speededness, exhaustion, or carelessness compatible with the decrease in person abilities after the change-point.

Researchers and practitioners can apply this method to detect the proportion of speeded examinees (high-to-low pattern) and remove the speeded items and recalculate the person abilities. However, if the proportion is large, they can use correction methods such as multidimensional IRT models to model speededness as another factor. In addition, if the number of examinees who show speeded behavior is not overly large, from an operational perspective, it is helpful to know when responses from those examinees deviate from typical test-taking behavior, because those aberrant responses may adversely affect estimates of item difficulty and person ability. With strong evidence that responses change after a given point and likely capture something other than evidence of person ability, one might elect to exclude those responses *before* calibration.

Limitation and future research. We did not investigate change points locations greatly, so further research on this would inform other underlying reasons (e.g., warm-up effects) of different patterns. We may also consider the Expected a Posteriori (EAP) estimator with prior known ability estimates instead of MLE. Kern and Choe (2021), used joint EAP in a computerized adaptive test along with response times and item responses to control differential speededness. They showed that by using joint EAP, ability estimates and speededness were less biased.

REFERENCES

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*(4), 29-37. https://doi.org/10.1111/j.1745-3992.2007.00106.x

Shao, C., Li, J. & Cheng, Y. (2016). Detection of Test Speededness Using Change-Point Analysis. *Psychometrika* **81,** 1118–1141 (2016). https://doi.org/10.1007/s11336-015-9476-7

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. MESA Press, 5835 S. Kimbark Ave., Chicago, IL 60637; e-mail: MESA@ uchicago. edu; web address: www. rasch. org; tele.

Kern, J. L., & Choe, E. (2021). Using a Response Time–Based Expected A Posteriori Estimator to Control for Differential Speededness in Computerized Adaptive Test. *Applied Psychological Measurement*, 01466216211014601. https://doi.org/10.1177%2F01466216211014601

Learner characteristics and dashboard feedback. Is comparison with peers effective for all learners?

Anja Hawlitschek
Otto-von-Guericke-Universität Magdeburg, Germany
anja.hawlitschek@ovgu.de

Galina Rudolf & Sebastian Zug
TU Bergakademie Freiberg, Germany
Galina.Rudolf@informatik.tu-freiberg.de
Sebastian.Zug@informatik.tu-freiberg.de

ABSTRACT (Poster): Comparison with peers is a widely used reference frame in dashboards. Indicators like students' class rank or students' performance in comparison with the class average are supposed to increase learners' engagement. However, learner who have a mastery goal orientation or are classified as low performers, might perceive such comparison features as frustrating or pressure. A decrease in motivation and learning could be the result. The consideration of such undesired effects is still at an early stage in learning analytics research. Our research wants to contribute to close this research gap. We examined the moderating effects of prior knowledge and goal orientations on the effects of providing comparison feedback versus content-oriented feedback on learning outcomes and motivation. Our findings indicate that prior knowledge indeed is a moderator and therefore has to be considered for tailored dashboard feedback. However, this is a preliminary result, as data from only one of two student cohorts is analyzed yet.

Keywords: learner characteristics, motivation, comparison feedback, dashboard

1 INTRODUCTION

The design of dashboard feedback is relevant for effects on learner. The reference frame of the feedback might yield undesired motivational or affective responses (Howell, Roberts, & Mancini (2018). Feedback with comparison features, which is very common in dashboards, might be suboptimal for those learners who are especially in need of instructional feedback and guidance – learners with low prior knowledge or low learning performance in comparison with the average. Comparison with peers, especially with competitive features like leaderboards, might be perceived as pressure and trigger stress or discouragement (Roberts, Howell, & Seaman, 2017). Results of studies provide mixed results on the effects of comparison features in dashboards so far – negative (e.g., Hanus & Fox, 2015) as well as positive (e.g., Aljohani et al., 2019) or no effects (e.g., Howell, Roberts, & Mancini, 2018). Studies indicate that effects are moderated and mediated by learners' cognitive (Teasley, 2017) and motivational characteristics (Schumacher & Ifenthaler, 2018). Comparison feedback in dashboards is typically outcome feedback which gives information about learners' performance but not on strategies to perform better (Aljohani et al., 2019; Lim et al., 2019). Students who have low prior knowledge and therefore perform lower on (initial) tasks might need content-related feedback. For them, competitive features might result in extraneous cognitive load and thus hinder performance (Kalyuga, 2017) or lower motivation. Furthermore, competitive features might only motivate students with a focus on performance goals. For students with a focus on mastery goals the comparison with peers might reduce motivation and performance (Schumacher & Ifenthaler, 2018).

The aim of the current study is to provide answers to the following question: Are there moderating effects of motivational predispositions and cognitive characteristics of learners on the impact of dashboard feedback on learners' motivation and learning performance?

2 STUDY

We carried out a one-factor experimental design in a course on learning programming in C. 27 students were randomly assigned to one of the two conditions – competitive or content-oriented dashboard feedback (see Table 1). Students got dashboard feedback only once, at the start of the semester, on their prior knowledge. Therefore, they had to solve C programming tasks. We measured their achievement goal orientation with approach items from the AGQ-R (Elliot & Murayama, 2008) in the first exercise. At the end of the semester they had to indicate their intrinsic motivation (Isen & Reeve, 2005) and solve a test with C programming tasks.

Table 1: Description of the two conditions

condition	**Description**	**students get …**
1 (N = 14)	content feedback	… their points in tasks, indication of difficulties in the tasks and hints about knowledge gaps that need to be filled = instructional guidance
2 (N = 13)	competitive feedback	… their points in tasks and the average, ranking in anonymized leaderboard = no instructional guidance concerning knowledge gaps and potential difficulties in tasks

3 RESULTS

Results of regressions indicate that prior knowledge moderates the effect of the two kinds of dashboard feedback on learning outcome in the way expected (see Table 2).

Table 2: Results of regressions of prior knowledge on learning outcome; note: *$p \leq .05$; **$p \leq .00$

		beta	T	p
learning outcome	Condition	.45	1.82	.08
(adjusted $R^2 = .63$; R^2 change = .09*)	prior knowledge**	.82	6.6	.00
	condition x prior knowledge*	-.58	-2.39	.02

We did not find any significant difference concerning intrinsic motivation or moderating effects of achievement goal orientation.

4 DISCUSSION

Our findings indicate that prior knowledge moderate the effects of the two types of dashboard feedback. For students with low prior knowledge, a competitive comparison feedback was counterproductive. However, students with higher prior knowledge achieved better learning outcomes in the competitive condition. A possible explanation for the result is that their good results in the comparison provided an incentive to elaborate further. Students with lower prior knowledge on the other hand benefit from content feedback, whereas students with higher prior knowledge did not. This result can be explained with reference to expertise-reversal effect (Kalyuga, 2007). However, limitations arise from our small N. For getting final results, we still have to integrate data from a second course where we implemented the same research design into our analysis.

From our results, we conclude that comparison features have to be considered carefully with regard to learner characteristics. More research considering moderating effects of cognitive and motivational learner characteristics is necessary to tailor dashboards to learners' needs.

REFERENCES

Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basheri, M., & Aslam, M. A. (2019). An integrated framework for course adapted student learning analytics dashboard, *Computers in Human Behavior*, 92, 2019, 679-690, https://doi.org/10.1016/j.chb.2018.03.035.

Elliot, A. J. & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, 100(3), 613-628

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80, 152-161.

Howell, J., Roberts, L. D., & Mancini, V. O. (2018). Learning analytics messages: Impact of grade, sender, comparative information and message style on student affect and academic resilience, *Computers in Human Behavior*, 89, 8-15. doi: 10.1016/j.chb.2018.07.021

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, 29, 295–323. doi:10.1007/s11031-006-9019-8

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19, 509–539. DOI 10.1007/s10648-007-9054-3

Lim, L.-A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2021). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course, *Learning and Instruction*, 72. https://doi.org/10.1016/j.learninstruc.2019.04.003.

Roberts, L.D., Howell, J.A. & Seaman, K. (2017). Give Me a Customizable Dashboard: Personalized Learning Analytics Dashboards in Higher Education. *Tech Know Learn,* 22, 317–333. https://doi.org/10.1007/s10758-017-9316-1

Schumacher, C. & Ifenthaler, D. (2018). The importance of students' motivational dispositions for designing learning analytics. *Journal of Computing in Higher Education*, 30, 599–619. 10.1007/s12528-018-9188-y.

Teasley, S. D. (2017). Student facing dashboards: One size fits all? *Tech Know Learn*, 22, 377–384. DOI 10.1007/s10758-017-9314-3

EdOptimize – An Open-Source K-12 Learning Analytics Platform

Tirth Shah
Playpower Labs
tirth.shah@playpowerlabs.com

Nirmal Patel
Playpower Labs
nirmal@playpowerlabs.com

Derek Lomas
Delft University of Technology
j.d.lomas@tudelft.nl

Aditya Sharma
Playpower Labs
aditya@playpowerlabs.com

ABSTRACT: Technology aided learning is becoming increasingly popular. In some of the countries, online learning has taken over for traditional classroom-based learning. With this, educational data is being generated in vast amounts. Knowing the potential of this data, many education stakeholders have turned to evidence-based decision making to improve the learning outcomes of the students. EdOptimize platform provides extensive actionable insights for a range of stakeholders through a suite of 3 data dashboards, each one intended for a certain type of stakeholder. We have designed a conceptual model and data architecture that can generalize across many different edtech implementation scenarios. Our source code is available at https://github.com/PlaypowerLabs/EdOptimize

Keywords: learning analytics, digital learning, dashboards, assessment data, curriculum-analytics, platform-analytics, implementation-analytics

Interactive Demo Link - https://youtu.be/NYvO3iOg3Kg

Comparing the Accuracy of Automatic Scoring Solutions for a Text Comprehension Diagramming Intervention

Héctor J. Pijeira-Díaz
Maastricht University
h.pijeiradiaz@maastrichtuniversity.nl

Jasper Snel
Brightlands Institute for Smart Society Heerlen
j.snel@maastrichtuniversity.nl

Sophia Braumann
Utrecht University
s.e.e.braumann@uu.nl

Janneke van de Pol
Utrecht University
J.E.vandePol@uu.nl

Tamara van Gog
Utrecht University
T.vanGog@uu.nl

Anique de Bruin
Maastricht University
anique.debruin@maastrichtuniversity.nl

ABSTRACT: Students typically have great difficulty monitoring their comprehension of textual materials. Completing diagrams about causal relations in expository texts has been a successful intervention to enhance the accuracy of students' reading comprehension judgments (i.e., monitoring accuracy), although there is still room for improvement. Such judgments play a role in crucial self-regulated learning decisions that students make such as allocating time and effort, selecting content for restudy, and/or consulting additional sources. The automated scoring of students' diagram content can provide a basis for strengthening the diagramming intervention with individual and simultaneous feedback to a high number of students. Leveraging an existing human-coded (correct and incorrect) dataset of 6000+ diagram answers (completed in Dutch by 700+ secondary students), we compared different automatic scoring solutions in terms of classification accuracy. Four computational linguistic models for Dutch were identified and tested in combination with four popular machine learning classification algorithms. The best solution reached 81% accuracy (i.e., four out of five answers matched the human coding). Depending on the accuracy required for different applications, these results could be used for fully- or semi-automated scorings of students' answers to generative activities used in reading comprehension interventions.

Keywords: computational linguistics, automatic scoring, reading comprehension, diagramming, monitoring judgments

1 INTRODUCTION

As students progress through the educational system, they switch from an early goal of learning to read, to increasingly reading to learn. In the process of reading to acquire knowledge, students monitor their comprehension, which is an important determinant of their study decisions (Thiede et al., 2009). For example, they spend more time restudying those texts that they think they understand less well. Study decisions, in turn, are likely to influence their exam scores.

The accuracy of students' monitoring of their reading comprehension is determined by relating how students think they will score on a test about a certain text, to their actual test scores. Unfortunately, meta-analytic results show an average accuracy below 0.30 (Goodman and Kruskal's gamma) (Prinz et al., 2020). Consequently, educational researchers have developed interventions (e.g., generating keywords, summaries, or diagrams) to enhance students' monitoring accuracy. In particular, completing pre-structured diagrams about causal relations in expository texts have proved effective in raising monitoring accuracy over 0.55 (van de Pol et al., 2019). Nonetheless, the numbers indicate that, despite the progress, there is room for improvement.

Concurrently, advances in computational linguistics offer promising opportunities for the automatic scoring of the diagrams produced by the students in those interventions. The interest in automation is motivated by the real-time, individual feedback possibilities it affords towards increasing monitoring accuracy. Leveraging an existing dataset of 6000+ diagram answers in Dutch, including human scores of whether answers are correct or not, we evaluated the accuracy of automatic scoring as compared to human scoring. The dataset answers were produced by 700+ Dutch-speaking secondary education students in a series of diagram completion interventions to enhance monitoring accuracy (van de Pol et al., 2019; van Loon et al., 2014). The automatic scoring capitalizes on computational linguistics models for text representation (Mikolov et al., 2013) in combination with machine learning classification algorithms. Our aims are therefore to identify available computational linguistics models for representing Dutch text and to compare their performance in terms of classification accuracy. Four such models and four classifiers are tested in this study.

2 METHODOLOGY

First, a literature review was conducted to identify the latest developments in computational linguistic models for text representation in Dutch. Then, four popular classification algorithms in machine learning, namely 1) logistic regression, 2) support vector machines (SVM), 3) random forests, and 4) neural networks; were used in combination with the linguistic models.

The content of each diagram answer in the existing dataset was human coded as correct or incorrect with satisfactory interrater reliability. For the automatic scoring, since mathematical computer algorithms operate with numbers instead of texts, the first step was to represent the text of each diagram answer as a multidimensional numerical vector (300 dimensions) using the identified computational linguistic models. Then, the classification algorithms were trained with 90% of the data, leaving the remaining 10% for evaluation of the classification performance.

3 RESULTS

Four available state-of-the-art computational linguistics models were identified, namely 1) "spaCy medium", 2) "spaCy large", 3) "FastText", and 4) "ConceptNet Numberbatch" (ConNum). Table 1 shows the classification accuracies of the automatic scoring of the texts obtained for all the combinations of the four Dutch models and the four classification algorithms. The size of the model is also included as an indication of their complexity. The best accuracy (81%) was obtained for the "spaCy medium" model in combination with a neural network classifier. Thus, the best performing combination for automatic scoring gave, in slightly over 4 out of 5 cases, the same score as a human did. Remarkably, it also means that the simplest model in terms of size (i.e., "spaCy medium") offered the best automatic scoring accuracy.

Table 1. Automatic scoring accuracy by classification algorithm and Dutch language model.

Dutch Model	Size	Classification Algorithm			
		Logistic Regression	SVM	Random Forests	Neural Networks
FastText	7GB	79%	77%	70%	81%
spaCy medium	44MB	80%	79%	70%	80%
spaCy large	545MB	77%	79%	70%	70%
ConNum	223MB	76%	80%	71%	70%

4 EDUCATIONAL SIGNIFICANCE

The educational significance of this work is threefold. First, our automated scoring solution enables the development and testing of automated individualized feedback interventions to further improve students' monitoring accuracy. Second, it can alleviate teachers' workload in scoring students' text comprehension diagrams. Third, it can be integrated into educational technology applications such as intelligent tutoring systems, especially those focusing on assisting reading comprehension.

REFERENCES

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations - Workshop Track Proceedings*.

Prinz, A., Golke, S., & Wittwer, J. (2020). To what extent do situation-model-approach interventions improve relative metacomprehension accuracy? Meta-analytic insights. *Educational Psychology Review*, *32*(4), 917–949. https://doi.org/10.1007/s10648-020-09558-6

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85–106). Routledge. https://doi.org/10.4324/9780203876428

van de Pol, J., de Bruin, A. B. H., van Loon, M. H., & van Gog, T. (2019). Students' and teachers' monitoring and regulation of students' text comprehension: Effects of comprehension cue availability. *Contemporary Educational Psychology*, *56*, 236–249. https://doi.org/10.1016/j.cedpsych.2019.02.001

van Loon, M. H., de Bruin, A. B. H., van Gog, T., van Merriënboer, J. J. G., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica*, *151*(September), 143–154. https://doi.org/10.1016/j.actpsy.2014.06.007

Creating an openly accessible dataset of learning dialogue

Rebecca Ferguson
Institute of Educational Technology, The Open University
rebecca.ferguson@open.ac.uk

Kirsty Kitto
Connected Intelligence Centre, University of Technology Sydney
kirsty.kitto@uts.edu.au

Catherine A. Manly
City University of New York Graduate Center
cmanly@gc.cuny.edu

Oleksandra Poquet
Centre for Research and Interdisciplinarity, University of Paris & INSERM U1284
sasha.poquet@cri-paris.org

ABSTRACT: Content analysis (CA) is a widely used method in the learning sciences, and so has become a well-accepted practice in the domain of learning analytics (LA). Increasingly, we see datasets coded with CA used as labelled datasets to drive machine learning. However, the scarcity of widely shareable datasets means that research groups around the world work independently to code text using CA, with few attempts made to compare results across groups. A risk is emerging that different groups using the same constructs are coding them in different ways, leading to results that will not prove replicable. In this poster, we report on the development of an openly accessible database containing the discussion associated with an international online course on learning analytics, which ran for four weeks on the Slack platform. Participants were aware that their postings would form part of the database, and that any personally identifiable information would be pseudonymised. The database will be shared via GitHub and the SoLAR website to support the development of replicable work on content analysis of learning and teaching dialogue.

Keywords: content analysis, dataset, open access, replication

1 CHALLENGES TO REPLICATION IN LEARNING ANALYTICS

How do we know that the approaches emerging in learning analytics (LA) are valid? Replication is key to validating the theories and models developed using quantitative approaches. Many fields have long established procedures that support reproducibility. For example, data science research communities organize competitions to provide the best solution for core challenges. Specified baseline datasets are analysed using different approaches with the results compared. The released datasets help ensure reproducible research results, and a better understanding of state-of-the-art solutions. These datasets also support new entrants to the field, who can access data, examine how it has been labelled and then work to develop their own sophisticated analytical methods.

Although LA has attempted to develop similar approaches, shareable datasets are uncommon. Some exceptions include a *Journal of Learning Analytics* special section (Dietze et al, 2016) with information about four open datasets; the LAK data challenge (Drachsler et al, 2014); the Pittsburgh datashop[1]; and the release of MOOC data from Stanford[2]. Yet, beyond this, the public release of data is uncommon in LA, and what does get released fails to cover the broad range of learning activities that LA strives to model. Nonetheless, groups continue collecting, cleaning, and exploring learning data, making implicit decisions about how to process and analyse it (Buckingham Shum & Luckin, 2019). Since many such decisions that influence the results are not well documented, different research groups may plausibly make different decisions when cleaning and labelling similar datasets. This lack of openness is a problem, as theoretical constructs adopted in quantitative analysis in LA, particularly those developed in educational research, were derived from rigorous qualitative approaches. They are highly contextual and grounded by rich descriptions of the situation in which they were developed. Our ongoing inability to document the contextual data, in addition to the lack of open and shareable datasets, creates further ambiguity around details relevant to supervised and unsupervised approaches applied towards content analysis to capture particular theoretical lenses.

Reasons as to why shareable datasets are rare and limited in scope are important. Sets of clickstream, discourse, and engagement-pattern data are difficult for research groups to access and cannot be shared without breaching the privacy of individual learners. Yet, the question of replication persists - the ontological and epistemological differences between research methods are frequently overlooked.

2 CREATING AN OPENLY ACCESSIBLE LA DATASET

This poster reports on an effort to create a shareable open dataset that can be used as a baseline in our research community. We seek to generate a discussion about how this initial effort might be scaled. To generate our dataset, we created a short online course about the past, present, and future of learning analytics. We sought ethics approval from the University of South Australia to collect data that will become publicly available, generated through the interactions of the learners. According to the course objectives, its participants were 1) to identify ways in which learning analytics has developed over the past decade; 2) to identify significant challenges for learning analytics in the next five years; and 3) to discuss how work at their institution aligns with the challenges for learning analytics. The curriculum included short open-ended tasks released weekly, that required the participants to engage with short videos by experts in Learning Analytics and provide reflective responses. Each week built on the content from the previous week. The course was developed by four researchers, two – taking on explicit instructional roles in the course, and another two – taking on roles of participants to support emerging discussions.

The invitation to the course with explicit consent information was distributed to the participants of the Learning Analytics Summer Institute 2021. Participants were informed that if they register, they would be able to engage with the themes around learning analytics for four weeks on a closed, specially dedicated Slack channel. They were also informed that any text they share with each other

[1] https://pslcdatashop.web.cmu.edu/

[2] https://datastage.stanford.edu/StanfordMoocPosts/

will form the public dataset, and only their personal names will be replaced by pseudonyms. Fifty-four participants signed up for the course and were added to the private Slack channel. The Slack channel was changed to the public discussion within the LASI'21 community five days into the course, for technical reasons. Once the course was completed, the team manually collected discussion data from the private channel (Figure 1) and downloaded Slack channel data from the public discussion (19 json files of participant activity representative of 19 individual days). We have also recorded screenshots of the discussions to capture the interface of the course.

Such a pilot activity demonstrates how LA researchers can join forces to facilitate data collection to build a dataset that can facilitate replication efforts, particularly though not limited to, automated content analysis. We report statistics to describe the scope of the collected data.

Event_ID	User	TS	Date	Message	Thread	Msg_ID	Likes	Replies	Pinned	Type
1	Winnie The Pooh	2:28AM	3.09.2021	Hello there!You have signed up for the course	1	1	0	0	1	Post
2	Cinderella	6:08PM	3.09.2021	I'll start the Activity 1 introductions. I'm a profess	2	1	2	0	0	Post
3	Little Red Riding Hood	NA	NA	NA	2	1	0	0	0	Like
4	Ariel	NA	NA	NA	2	1	0	0	0	Like
5	Little Red Riding Hood	2:44AM	6.09.2021	I'm a doctoral candidate at a public research ur	3	1	1	0	0	Post
6	Winnie The Pooh	NA			3	1	0	0	0	Like
7	Ariel	4:35AM	6.09.2021	Hi, I am a doctoral candidate in a research univ	4	1	0	2	0	Post
8	Winnie The Pooh	NA		Hi Ariel, am curious about your area of focus wi	4	2	0	0	0	Reply
9	Ariel	NA		Hello Winnie, right now we are designing dashb	4	3	0	0	0	Reply
10	Richard the Lion Heart	5:44AM	6.09.2021	Activity 1 - Introduction of myself: I'm a universit	5	1	0	0	0	Post
11	Richard the Lion Heart	6:15AM	6.09.2021	Activity 2 - LAK12: There were much fewer subı	6	1	0	0	0	Post
12	Phoenix bird	10:36AM	6.09.2021	Hello, I am Phoenix bird, Doctoral Student in In	7	1	0	2	0	Post
13	Winnie The Pooh	NA		Hi Phoenix, what aspects of multimodal learning	7	2	0	0	0	Reply

Figure 1. A Screenshot of the dataset constructed from the first five days of the course

The slack channel generated discussion data from 13 participants and 4 course designers. The dataset contains a total of 99 text messages, which comprised a total of 32 individual threads. The mean number of messages per participant was 5.8, with a median of 3. The highest number of messages was 18 – produced by one of the participants, closely followed by one of the instructors with 16 messages. The minimum message length was 52 characters, a median of 997 characters. The longest participant response was 4466 (for illustration purposes, the text on page two of this proposal is about 3000 characters).

Although our dataset is small for large-scale text analysis, this pilot demonstrates that the LA community can focus on generating ethical and shareable data – generating a resource that can grow over time and support replication efforts in automation of text, particularly around theoretical constructs. We hope to engage LA researchers in a conversation around which of the fields and what format will be most appropriate, to provide sufficient meta-data and document the dataset for release and public use by the LA community, via the website of the Society for Learning Analytics and Research. We also hope to spur discussions around future data collection efforts and a public store.

REFERENCES

Buckingham Shum, S. & Luckin, R. (2019). Learning Analytics and AI: Politics, Pedagogy and Practices. *British Journal of Educational Technology*.

Dietze, S., George, S., Davide, T., & Drachsler, H. (2016). Datasets for learning analytics. *Journal of Learning Analytics* 3(2), 307-311.

Drachsler, H., Dietze, S., Herder, E., d'Aquin, M., & Taibi, D. (2014). The learning analytics & knowledge (LAK) data challenge 2014. In *LAK14* (pp. 289-290). ACM.

How early can we provide feedback?: Predicting success using data on learning behavior and supervised machine learning

Jakob Barth
University of Goettingen
jakob.barth@stud.uni-goettingen.de

Florian Berens
University of Goettingen
florian.berens@uni-goettingen.de

Sebastian Hobert
University of Goettingen
sebastian.hobert@uni-goettingen.de

ABSTRACT: The project aims to predict students' learning success as early as possible during a course using supervised machine learning algorithms. Therefore, cross-sectional data collected during a statistics lecture at the University of Goettingen in the spring semester 2020 is used. The data shows students' learning progress on the course's online learning platform at different points during the semester and is linked with the grade of the final exam at the end of the semester. The findings will be used to provide feedback to students on their predicted learning success based on their previous learning behavior. Preliminary findings using KNN and Classification Tree Algorithms show that the time to predict learning success most accurately, according to the performance of the algorithms, is located at the end of the semester. However, to provide feedback, an earlier point in time is necessary. So, according to first well-performing algorithms, the preferred point in time to give first feedback is around week six of twelve. Further research must show whether this really is the optimal time to provide feedback.

Keywords: supervised machine learning, learning success, feedback, learning behavior, predicting learning success.

THEORY AND BACKGROUND

Various studies examined the success of students learning behavior by using machine learning methods based on data collected during the learning process. Some of those studies use the ASSISTments data frame (Liu & Tan, 2020; Makhlouf & Mine, 2020; Almeda & Baker, 2020). ASSISTments is an online learning platform used for mathematics education (Liu & Tan, 2020, p. 19). This learning data is linked with a variable measuring learning success. Learning success is defined as pursuing a STEM career after graduating from college, and the aim of the studies is to predict these students pursuing STEM careers (Liu; Tan 2020, p. 19). Another way of using machine learning methods in the context of students learning behavior is the detection of students at risk by predicting their exam result. Several studies were conducted in the past few years using data from online learning tools to predict students' performance at university exams. One example is a recent study from Halit Karalar et al. who were able to predict students at risk of failing a course at a Turkish university using an ensemble learning model approach (Karalar et al., 2021).

The planned project pursues a similar goal. The research aims to predict learning success at different points in the semester. The data frame described in the following chapter contains learning data and relates to the learning success measured as the grade in the final exam of the statistics course. The data was collected in the context of a statistics course in the spring semester of 2020 at where an online learning tool was used to support all aspects of learning. In this learning tool, each week of the course, students are provided videos accompanied by questions relating to the lecture. Thus, overall, almost all the students´ learning behavior is represented in this tool. The benefit of the project shall be that students in the following statistics courses should get automatic feedback via the online learning platform on their predicted learning success based on previous learning behavior. This can be seen as a form of formative feedback, defined by Valerie Shute: "[…] as information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning " (Shute, 2008, p. 154). Also, empirical studies show the influence of feedback on learning. For example, John Hattie conducted a meta-study on student achievement in higher education. One of the main findings is that feedback is among the strongest influences on students learning (Hattie, 2015).

This analysis aims to evaluate the best point in time to predict success. By this point in time, one should be able to predict with adequate accuracy while students still have time to adjust their learning behavior according to the given feedback to pass the exam successfully.

DATA

Anonymized data was collected within the online learning platform of the course. This data contains many aspects of students' learning behavior while participating in the statistics course. Thus, for each of the 12 weeks of the course, and additionally for the exam preparation phase of 10 days, cross-sectional data is available and can provide information on learning behavior. The data set contains the measures of how many questions the student answered (in total and distinct), how many of these questions the student answered correctly (in total and distinct), how many minutes the student spent working on questions in total, how many videos the student watched (in total and distinct), and how many days the student has been active on the learning platform. For each case, the learning behavior is connected with the achievement in the final exam of the course. This outcome variable was measured as a percentage of the total score to be achieved. For the analyses, it was recoded into a binary variable. All cases that reached a share of at least 50 percent were defined as passed, all cases with a lower score were defined as failed. The number of cases in the weekly data sets increases from 219 in the second to 247 in the twelfth week, because some students started to use the online tool at a later point of the course.

PRELIMINARY FINDINGS

For the study different algorithms like KNN, Classification Trees, Neural Networks and Support-Vector Machines shall be used. The preliminary findings are based on the performed (K-nearest neighbors) KNN and Classification Trees Algorithms within an implemented threefold cross-validation. Based on these preliminary findings, the predictive power increases in time with a few exceptions. These differ between KNN and Classification Trees. Since other accuracy measures cannot be interpreted very meaningfully due to the skewed distribution of the data, the F-score was mainly used here to evaluate the performance of the respective algorithm.

As a preliminary finding, already halfway through the course, a good prediction can be made. Quite good F-scores (55.3) are reached at week six by both algorithms. The best performance according to the highest F-score in KNN as well as in Classification Trees Algorithm is reached at week 11 of the lecture (Table 1). Because the course lasts for 12 weeks in total (plus the days for exam preparation), this would be too late to inform the students valuably. Therefore, an earlier forecast is preferred. Furthermore, more research must be done on both algorithms and the employment of different algorithms to detect the valuable points in time for predicting the learning success. The preferred point in time would be as early as possible.

Table 1: Preliminary Findings

Week	KNN F-Score	Classification Tree F-Score	Number of Cases
Week 2.5	48.6	54.1	219
Week 3	47.1	61.5	220
Week 4	43.9	60	228
Week 5	48.8	65.6	231
Week 6	55.3	55.3	232
Week 7	56.5	50	233
Week 8	59	47.8	234
Week 9	51.1	63.3	237
Week 10	56.6	66.7	241
Week 11	68	67.7	243
Week 12	56.6	61.8	247

REFERENCES

Almeda, M. V. Q., & Baker, R. S. (2020). Predicting Student Participation in STEM Careers: The Role of Affect and Engagement during Middle School. *Journal of Educational Data Mining, 12(2),* 33-47. https://doi.org/10.5281/zenodo.4008054

Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship of Teaching and Learning in Psychology, 1(1)*, 79–91. http://dx.doi.org/10.1037/stl0000021

Karalar, H., Kapucu, C. & Gürüler, H. (2021). Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education 18 (63)*, 1-18. https://doi.org/10.1186/s41239-021-00300-y

Liu, R., & Tan, A. (2020). Towards Interpretable Automated Machine Learning for STEM Career Prediction. *Journal of Educational Data Mining, 12(2),* 19-32. https://doi.org/10.5281/zenodo.4008073

Makhlouf, J., & Mine, T. (2020). Analysis of Click-Stream Data to Predict STEM Careers from Student Usage of an Intelligent Tutoring System. *Journal of Educational Data Mining, 12(2),* 1-18. https://doi.org/10.5281/zenodo.4008050

Shute, Valerie J. (2008). Focus on formative feedback. *Review of Educational Research 78(1),* 153–89. https://doi.org/10.3102/0034654307313795

Analyzing Problem-Solving Using Process Data and Transitional Networks: The Case of Colombia and Singapore

Daniela Castellanos-Reyes, Katherine Chartier, and Jennifer Richardson
Purdue University, Indiana University
casteld@purdue.edu, kcharti@iu.edu, jennrich@purdue.edu

ABSTRACT: Large scale assessment data has many affordances but relying on test scores alone may not provide a full picture of complex thinking and reasoning. Emerging research has used process data to examine group problem solving. Process data captures individual, team, or country data, as well as important relationships and sequence patterns. We used process data to explore individual problem solving using extreme case analysis. Comparing processes data from a high (Singapore) versus low (Colombia) performing country allowed us to better understand problem solving proficiency. Findings revealed that high-performing learners took more steps to solve a knowledge acquisition task compared to low performing learners based on the geodesic distance of the networks. Furthermore, we found that low performing learners revisited more steps than high performing. These findings suggest that more steps taken (i.e., more clicks) and varying actions instead of revisiting during problem exploration could be related to superior knowledge acquisition scores. Although using process data to understand problem solving is an emerging and underdeveloped area of research, it demonstrates the potential to be a valuable method to use in large scale analysis to better understand problem solving and more reliably interpret scores. This is a poster submission.

Keywords: Process data, problem solving, network analysis, ERGM, large scale assessment.

1 INTRODUCTION

Computer-based testing has been used in large-scale assessments for youth (e.g., PISA). Process data, contained in log files, provides information about response processes, enhancing our understanding of complex cognitive processes essential to assessing problem-solving (He et al., 2019). In the PISA problem-solving large-scale assessment, OECD (2014) reported that some countries' performance (e.g., Korea, Singapore) was stronger than expected on knowledge acquisition tasks but weaker in knowledge application. Conversely, some other countries (e.g., Brazil, Colombia) are particularly stronger in knowledge application but weaker on knowledge acquisition tasks. The OECD (2014) also reported that this phenomenon seems to overlap with related historical and geographical groupings (i.e., Asian and Latin America). Although we know that certain countries are consistent higher performers, we do not know what behaviors make them successful. Examining such behaviors would allow us to make practical recommendations to other countries that want to improve their problem-solving proficiency. To date, only Zhu et al. (2016) investigated individual problem-solving through sequence patterns of computer-based assessment items using transitional networks. They used students' actions to represent nodes and directed links to sequentially connect actions (Zhu et al., 2016). The purpose of this research is to investigate the problem-solving patterns of high performing (HP) learners in contrast to the problem-solving patterns of low performing (LP) learners in knowledge acquisition tasks, specifically, *to what extent do specific behavior patterns account for the overall problem-solving process?*

2 METHOD

2.1 Context and Data Preparation

We used extreme case analysis to compare the problem-solving proficiency in knowledge application tasks of high (i.e., Singapore, n = 469) versus low (n = 752) performing countries. We used the PISA's Climate Control item which asks students to manipulate a new air conditioner that has no instructions. We use transitional networks obtained from learners' interaction within the assessment platform. Learners' actions within the learning platform (e.g., click option A) are represented by nodes and learners' decisions to interact with other elements of the platform are represented by ties (e.g., option A, then option B). Weights were assigned to each edge by counting the frequency in which each action occurred.

2.2 Analytical Approach

We used Exponential Random Graph Models (ERGM) as an analytic strategy to analyze problem-solving transitional networks. ERGM allows researchers to examine different forms of substructures in a network (Robins, 2015), in the case of problem-solving, ERGM are useful for pattern recognition of students' interactions. Two models were built to explore problem-solving in Singapore and Colombia respectively. Table 1 summarizes the effects used to model the behaviors and the results per country.

Table 1: ERGM Effects Definition and Results.

Effect Name	Contextual Definition	Singapore	Colombia
Edges	The likelihood of observing an action (clicking from A to B) in the transition network of a problem acquisition item (A -> B).	-7.481 (.016***)	-7.587 (.016***)
Reciprocity (mutual)	The likelihood of observing the action of clicking on option A, then clicking in option B, and revisiting option A in the transition network of a problem acquisition item (A -> B -> A).	2.089 (.103***)	3.040 (.009***)
Node Covariate (nodecov)	The likelihood of observing more actions (cicks from A to B) based on clicks on the same option.	.003 (.000***)	.001 (.000***)
Out degree two (odegree(2))	The likelihood of observing actions starting from clicks that have outdegree of two (i.e., that two actions started from clicking the same option).	-1.004 (065***)	-1.176 (.068***)

3 RESULTS AND IMPLICATIONS

The ERGM estimates serve as predictors of the presence of actions in the transition network. All estimates in both models are statistically significant. This is likely due to network size. The parameter *edges* is negative in both models indicating that both networks are sparse which means that the transition network has fewer actions within the platform than possible. Low density was expected given that the possibilities of unique actions is exponential. The likelihood of observing actions starting from clicks that have outdegree of two (i.e., that two actions started from clicking the same option) is very low suggesting the need to simplify the transition network by removing actions that occurred few times. Both *reciprocity* and *node covariate* were positive estimates however node covariate was close to zero. An estimate close to zero suggests that the frequency of actions as node's covariate may not be a good predictor of problem-solving behavior. The parameter *reciprocity* was positive for both HP and LP networks. However, the estimate was larger for the LP network suggesting that learners in this group revisited clicks more than the HP learners. This suggests that low estimates of the *mutual* parameter could be a good predictor of problem-solving performance and better represents exploratory behavior than scores alone. These findings agreed with Greiff et al. (2015) who used log files to analyze climate control from the same data set (e.g., PISA 2012). They analyzed optimal problem exploration (e.g., vary-one-thing-at-a-time, VOTAT) and found VOTAT strategy was strongly related to performance on the climate control item as well as on overall problem-solving score. Process data demonstrates the potential to describe underlying problem-solving behaviors and predict scores. This is especially important when using large-scale assessment data as a common limitation is an overreliance on test scores (e.g., learner products) may not fully describe learners' thinking processes (Bergner & von Davier, 2019).

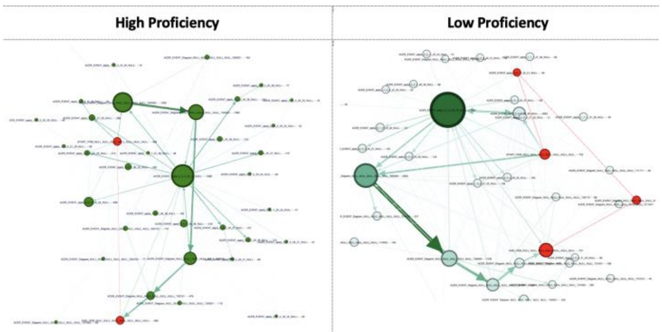


Figure 2: Problem solving transition network visualization

REFERENCES

Bergner, Y., & von Davier, A. A. (2019). Process data in NAEP: Past, present, and future. *Journal of Educational and Behavioral Statistics*, *44*(6), 706–732.

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers and Education*, *91*, 92–105.

He, Q., Borgonovi, F., & Paccagnella, M. (2019). *Using process data to understand adults' problem-solving behaviour in the Programme for the International Assessment of Adult Competencies (PIAAC): Identifying generalised patterns across multiple tasks with sequence mining*.

OECD. (2014). PISA 2012 technical report programme for international student assessment.

Robins, G. (2015). Exponential Random Graph Models for Social Networks. In J. Scott & P. Carrington (Eds.), The SAGE Handbook of Social Network Analysis (pp. 484–500).

Zhu, M., Shu, Z., & von Davier, A. A. (2016). Using networks to visualize and analyze process data for educational assessment. *Journal of Educational Measurement*, *53*(2), 190–211.

Early Prediction of Exam Performance with Gamified Learning App

Julian Langenhagen
Goethe University Frankfurt
langenhagen@econ.uni-frankfurt.de

ABSTRACT: In prior learning analytics research, the implementation of early prediction algorithms has been primarily evaluated in settings with extensive data sets from learning management systems. This study is the first to test whether a gamified learning app can be used as an early predictor of student performance. In addition, this work addresses two underexplored areas of learning analytics: portability and data privacy. The data in this study come from four semesters of a generally identical course. However, due to the outbreak of COVID-19, the face-to-face course was converted into an online lecture after two semesters. This circumstance allows to investigate whether a predictive model from one semester can be transferred to another semester without accuracy loss, especially considering a changed teaching method. Due to compliance with strict data protection regulations, the scope of the data collected is comparatively small. Therefore, the results will also provide information on whether it is possible to implement an accurate early warning system with limited data sets. This poster introduces the research setting and preliminary results to discuss further steps, including the implementation of the early warning system in practice.

Keywords: predictive analytics, gamification, early warning system, portability, data privacy

1 INTRODUCTION

Predictive analytics is one of the main subdomains of learning analytics (Chen et al., 2020). In the majority of studies in this field, extensive data sets from learning management systems (LMS) used in Massive Open Online Courses (MOOC) are examined to predict either the exact score of a student or whether the student will pass or fail (Conijn et al., 2016; Gašević et al., 2016). Mainly due to the limited availability of relevant data, corresponding studies on face-to-face events are scarce (Van Goidsenhoven et al., 2020). To contribute to this research gap, I use data from a gamified learning app that was developed for an undergraduate accounting course at a large European university. This study aims to test whether the usage data of this app can be used to predict the students' final exam outcomes. As the data set covers four consecutive semesters, the study also contributes to the question of portability of a once-determined predictive model to future semesters. By dividing the semesters into subsections, I can also test whether the app can be used as an early warning system (EWS) to detect students who are likely to fail the course. The general content and structure of the lecture did not change over the course of the four semesters. However, due to the outbreak of COVID-19, the face-to-face lecture was changed to an online lecture after two semesters. This characteristic of the research setting allows to investigate whether such a change has an impact on portability. This study also sheds light on the topic of data privacy in the context of predictive learning analytics. On the one hand, the sample size of four semesters with about 600 students per semester is large. On the other hand, however, the scope of the collected data is limited due to strict data protection regulations. Due to those regulations, I was obliged to ask the students for their student ID in the app on a voluntary basis and I need this ID to connect the data points from the two data sources (app and exam). As not every student provided their ID, I was only able to connect the data of the two data sources for a limited number of students (see Table 1). Therefore, the present study can also shed light on the question of whether a functioning EWS can be implemented even with such a limited amount of data.

Table 1: Data Sources and Sample Sizes.

Data Source	SS19	WS19	SS20	WS20
App	559	595	447	546
Exam	575	648	616	644
App+Exam	230	243	190	190
Teaching	face-to-face		online	

2 RESEARCH STRATEGY

Since most prior studies refer to data from an LMS or similar sources, there is not yet a set of established performance measures for the use of gamified learning apps. This study contributes to this research gap by developing and testing novel performance measures tailored to the app used in this research setting. In total, nine performance measures were developed that address different aspects of app usage, covering quantitative (number of total questions answered) and qualitative (share of correct answers) dimensions as well as temporal aspects (number of active days). Various algorithms have been used in past studies to predict exam performance and different algorithms have led to the best result in the individual settings (López-Zambrano et al., 2021). Therefore, the first research question of this project answered in this poster is: Which classification algorithm achieves the best prediction results in the current setting? In the first step, I split the usage data per semester into four equal parts. The first part contains the usage data for the first quarter of the semester, the second part contains the data for the first half, and so forth. In comparable studies the data was split into weeks (Van Goidsenhoven et al., 2020). This procedure would not be reasonable in this setting as every semester has a different number of weeks and the results would not be comparable. With the split into four equal subparts, the development of the predictive power of the algorithms can be examined over time. In the second step, I conducted a receiver operating characteristic (ROC) curve analysis for every algorithm/timeslot combination and calculated the corresponding area under the ROC curve (AUC) as a measure of discrimination. We report the AUC as this measure is considered as an informative measure, especially for cases with class imbalances (Ling et al., 2003). This applies to this setting since the average failure rate in the exam over all four semesters taken into account is 30.45%. I used the academic status (pass/fail) as the binary dependent variable and the performance measures discussed before as independent variables. I used the data to test three different classification models: Logistic Regression, K-Nearest-Neighbors (KNN), and Random Forest. All three belong to the most used algorithms in predictive learning analytics and are therefore in line with prior research (López-Zambrano et al., 2021). To obtain preliminary results, I test the algorithms with the same dataset they were trained with, e.g., the prediction for the outcome of summer semester 2019 is trained and tested with the usage data from (a part of) this semesters' students.

3 PRELIMINARY RESULTS AND NEXT STEPS

Table 2 shows that the Random Forest algorithm achieves the best results in this research setting. The predictive power tends to increase with more data in every semester for every algorithm. Still, only the Random Forest algorithm achieves “outstanding” discrimination scores (according to Hosmer & Lemeshow, 2000, p. 162). The results of this analysis show that the algorithm can explain the students' learning outcome ex-post with high accuracy. However, as discussed before, the AUCs in Table 2 result from models that were trained with the same data they were tested on. This approach is in line with the majority of prior research (Conijn et al., 2016; Gašević et al., 2016). However, this does not simulate a real EWS as its purpose is to predict the students' outcome ex-ante. Therefore, future work in this project will evaluate how the optimal algorithm will perform in a practical test and whether the performance differs across semesters to examine the portability of the early warning system. If the test is successful, the early warning system could be implemented in the app, providing students and instructors with additional feedback during the semester.

Table 2: Preliminary Results (AUC).

	Logistic Regression				Random Forest				KNN			
	SS19	WS19	SS20	WS20	SS19	WS19	SS20	WS20	SS19	WS19	SS20	WS20
25%	0.60	0.67	0.63	0.63	0.92	0.96	0.86	0.92	0.72	0.72	0.71	0.71
50%	0.69	0.59	0.65	0.65	0.95	0.97	0.88	0.92	0.74	0.68	0.74	0.64
75%	0.69	0.63	0.67	0.66	0.97	0.99	0.91	0.97	0.79	0.76	0.76	0.69
100%	0.75	0.71	0.75	0.66	0.99	0.99	0.94	0.98	0.81	0.80	0.81	0.81

REFERENCES

Chen, G., Rolim, V., Mello, R. F., & Gašević, D. (2020). Let’s shine together! A comparative study between learning analytics and educational data mining. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 544–553.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29.

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

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A statistically consistent and more discriminating measure than accuracy. *Ijcai*, *3*, 519–524.

López-Zambrano, J., Torralbo, J. A. L., & Romero, C. (2021). Early Prediction of Student Learning Performance Through Data Mining: A Systematic Review. *Psicothema*, *33*(3), 456–465.

Van Goidsenhoven, S., Bogdanova, D., Deeva, G., Broucke, S. vanden, De Weerdt, J., & Snoeck, M. (2020). Predicting student success in a blended learning environment. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 17–25.

Examining pedagogical data literacy: Results of a survey among school teachers at upper secondary level in Switzerland

Konstantinos Michos and Dominik Petko
Institute of Education, University of Zurich
{konstantinos.michos, dominik.petko}@uzh.ch

ABSTRACT: One of the main concerns of learning analytics studies that involve teachers is the concept of data literacy. Pedagogical data literacy includes collecting and analyzing student data and deriving educational interventions, the success of which is again assessed using data. In this paper, we examine this construct to understand pedagogical data literacy levels within a representative survey of *N* = 1059 teachers in upper secondary schools in Switzerland. Preliminary results reveal that more than half of the secondary school teachers indicate having access to digital student data but only one-third is making use of this data to inform their teaching. Only one-fourth of the teachers indicate that they consider themselves proficient in using digital student data to improve their teaching. In-depth analyses will explore different conditions and characteristics that are associated with different teacher profiles of pedagogical data literacy.

Keywords: pedagogical data literacy, school teachers, digital student data, learning analytics

1 PEDAGOGICAL DATA LITERACY AND LEARNING ANALYTICS FOR SCHOOL TEACHERS

There are increasing demands for data-driven decision making (DDDM) in education to improve teaching and learning. During the last decades, several research studies and policy recommendations have proposed data use by teachers and school leaders with the aim to monitor student progress, to adapt teaching to student needs, facilitate data-driven decision making in education and generate school improvement plans (Poortman & Schildkamp, 2016).

While learning analytics has been widely used in Higher Education, the use in K12 contexts has emerged only recently due to the increasing use of digital learning environments in schools. In 2017, learning analytics was ranked among the top technology trends in K-12 education (Freeman et al., 2017). Since then, research has moved from the introduction of learning analytics platforms to teachers' decision making processes and uses (Kovanovic et al., 2021). According to van Leeuwen et al. (2021) this includes general teacher characteristics like technological skills, age and gender and complex characteristics like pedagogical knowledge, professional routines and data literacy. Further research is needed on how these aspects interact and how they influence the pedagogical use and impact of learning analytics.

In educational research, the concept of teachers' data literacy has evolved from a focus on the skills required for teachers to engage, collect, analyze and interpret educational data to the ability for taking pedagogical actions based on data (e.g., changing instructional strategies or providing feedback to students) (Henderson & Corry, 2020). The most common and relevant definition in the literature regarding this issue is the concept "data literacy for teaching" (Mandinach & Gummer, 2016) also referred as "pedagogical data literacy" (Mandinach, 2012). According to Mandinach and Gummer (2016), "data literacy for teaching is the ability to transform information into actionable instructional knowledge and practices by collecting, analyzing, and interpreting all types of data (assessment, school climate, behavioral, snapshot, etc.) to help determing instructional steps. It combines an understanding of data with standards, disciplinary, knowledge and practices, curricular knowledge, pedagogical content knowledge, and an understanding of how children learn" (p. 367). There are several arguments that school teachers' data literacy is lacking and that an improvement in this regard would be beneficial. In particular, it is expected that their engagement with student data can assist teachers in moving from intuitive and undocumented student assessment processes to a systematic way of monitoring student progress (Reeves & Honig, 2015). Nevertheless, in the context of school teachers, few studies examine pedagogical data literacy with regard to digital student data at a larger scale. Therefore, the main goal of this paper is to present a snapshot of descriptive survey results on the state of pedagogical data literacy levels of school teachers in upper secondary schools.

2 METHODOLOGY AND PRELIMINARY RESULTS

A large-scale survey study was designed and is currently being conducted in Switzerland. While the national study is still underway, we are able to present representative data from one of the largest cantons in Switzerland (data collection period: August – October 2021). For the purposes of this study, we adopted an instrument by Wayman et al (2016) to construct a short scale to measure pedagogical data literacy of school teachers. The scale included three items regarding a) available technologies for the analysis of digital student data, b) pedagogical actions based on analysis of digital student data, and c) pedagogical competency regarding the use of digital student data. In addition, the overall survey included several questions regarding demographic information, the profile of the school teachers and the use of digital learning platforms.

N = 1059 school teachers participated in this large-scale study. In particular, teachers' responses show that more than half of the teachers (see Figure 1) agree on having access to adequate technologies for the examination of digital student data (*M*=3.52, *SD*=1.21). At the same time, only one third of the teachers agrees on using this digital student data to plan and adjust their teaching (*M*=2.80, *SD*=1.39). Teachers are even more critical when judging their abilities to improve teaching and student learning based on digital student data (*M*=2.69, *SD*=1.21).

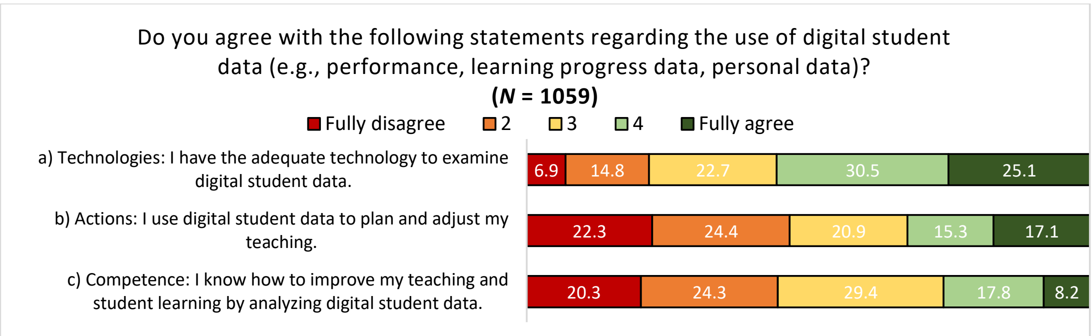


Figure 1: School teachers' pedagogical data literacy levels (weighted percentages)

While each of the three items provides interesting insights on its own, all three items can also be condensed to an overarching scale showing good reliability and a consistent factor structure (Cronbach's alpha = .78; weighted *M*=3.01, *SD*=1.05). The scale will be used to examine additional factors explaining teachers' characteristics and determinants of pedagogical data literacy levels.

3 CONCLUSIONS

Based on a representative sample of Swiss upper secondary school teachers, our results indicate that although more than half of the teachers have access to learning analytics technologies, only a minority is using these technologies to adjust their teaching. In addition, the vast majority of teachers do not feel competent in using learning analytics to improve their educational practice. Based on these findings, research needs to explore the conditions under which higher levels of pedagogical data literacy can develop. We are planning to analyse and present whether different teachers' characteristics (e.g., subject, school type) and the use of different digital platforms (e.g., online communication, LMS) are associated with different levels of pedagogical data literacy. Finally, future work will also explore teachers' dispositions such as the ethical use of digital student data for teaching (Mandinach & Gummer, 2016) within teacher training programs.

REFERENCES

Freeman, A., Adams Becker, S., Cummins, M., Davis, A., & Hall Giesinger, C. (2017). NMC/CoSN Horizon Report: 2017 K–12 edition. The New Media Consortium. Retrieved from https://library.educause.edu/~/media/files/library/2017/11/2017hrk12EN.pdf

Henderson, J., & Corry, M. (2020). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*, *13*(1), pp-pp?. https://doi.org/10.1108/jrit-11-2019-0074

Kovanovic, V., Mazziotti, C., & Lodge, J. (2021). Learning Analytics for Primary and Secondary Schools. *Journal of Learning Analytics*, *8*(2), 1-5. https://doi.org/10.18608/jla.2021.7543

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71-85. https://doi.org/10.1080/00461520.2012.667064

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366-376. https://doi.org/10.1016/j.tate.2016.07.011

Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education*, *60*, 425-433. https://doi.org/10.1016/j.tate.2016.06.010

Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, *50*, 90-101. https://doi.org/10.1016/j.tate.2015.05.007

van Leeuwen, A., Knoop-van Campen, C. A., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12. *Journal of Learning Analytics*, *8*(2), 6-21. https://doi.org/10.18608/jla.2021.7325

Wayman, J. C., Wilkerson, S. B., Cho, V., Mandinach, E. B., & Supovitz, J. A. (2016). Guide to Using the Teacher Data Use Survey. REL 2017-166. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/appalachia/pdf/REL_2017166.pdf

Design of Open-Source Video Viewing Behavior Analysis System

Shizuka Shirai[1*], Masumi Hori[2], Masako Furukawa[3],
Mehrasa Alizadeh[1], Noriko Takemura[1], Haruo Takemura[1], Hajime Nagahara[1]
[1] Osaka University, Japan; [2] NPO CCC-TIES, Japan; [3] National Institute of Informatics, Japan
[*]shirai@ime.cmc.osaka-u.ac.jp

ABSTRACT: This interactive demo presents ViLOG (Video viewing LOG analytics system), an open-source video viewing behavior analysis system developed as the first step of an ongoing research project on engagement analytics for video-based learning. This system was developed as a Learning Tools Interoperability (LTI)-compliant tool so that it would be easy to implement in any environment and to extend features to validate the new framework. This system consists of two main features regarding video viewing log: dashboard page and log collecting module. The dashboard page is focused on confirming the overview of students' progress in real time. In addition, the module saves video viewing log data such as play, pause, fast-forward, rewind, and playback speed changes every 10 seconds, in accordance with the log format defined by National Institute of Informatics. However, video viewing log data is not enough to accurately understand students' learning progress. Further consideration is needed to support students more accurately, such as adopting a multimodal learning analytics approach. We have already started considering an estimation model using multimodal data. We will expand the features to collect multimodal data and explore integrating an engagement estimation model into this system in future work.

Keywords: data visualization, video analytics, dashboard, e-learning, learning analytics

Demo video link: https://youtu.be/W9O-E8EyDnQ

ACKNOWLEDGMENT

This work was supported by MEXT "Innovation Platform for Society 5.0" Program Grant Number JPMXP0518071489.

REFERENCES

Alizadeh, M., Shirai, S., Takemura, N., Terai, S., Nakashima, Y., Nagahara, H., & Takemura, H. (2020). A survey of learners' video viewing behavior in blended learning. Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK20), 768-770.

Kawamura, R., Shirai, S., Takemura, N., Alizadeh, M., Cukurova, M., Takemura, H., & Nagahara, H. (2021). Detecting Drowsy Learners at the Wheel of e-Learning Platforms with Multimodal Learning Analytics. IEEE Access, 9, 115165-115174.

Can Learning Logs Be Useful Evidence in Cheating Analysis in Essay-type Questions?

Tsubasa Minematsu* and Atsushi Shimada
Kyushu University, Japan
***minematsu@ait.kyushu-u.ac.jp**

ABSTRACT: As cheating has become more sophisticated in online examinations, ensuring academic integrity is a challenging task. Previous research has been conducted to detect cheating based on learners' behavior, but not to analyze their learning activities before the examination. We propose a cheating detection score using e-book reading time as the learning activity and essay-type questions as the examination data. Our case study showed that our scoring method could detect students suspected of cheating. In addition, we presented some examples that combining learning activities with the examination data could provide insights that cannot be obtained from the examination data only.

Keywords: learning log analysis, e-learning system, cheating, natural language processing

1 INTRODUCTION

The pandemic of COVID-19 has forced educational institutions worldwide to adopt online teaching and learning via e-learning platforms. The difficulty of ensuring academic integrity has become more apparent than ever. Students can participate in lectures and examinations from various locations in online learning, making it easier to cheat. As a countermeasure, various researchers have proposed to detect cheating, such as response speed analysis during examination time and plagiarism checkers (Rogerson, 2017). The limitation of the previous research is that they did not focus on the learner's learning activities before examinations. Recently, e-book systems can record detailed learners' activities, such as reading time on each page. However, no cheating analysis focuses on such learning logs before examinations to the best of our knowledge. Therefore, our research question of this study is "How can detailed learning logs such as reading time of e-books provide useful insights for cheating analysis?". For investigating our research question, we proposed a cheating detection score based on the reading time of e-books as learning logs and essay similarity. We analyzed the learning logs and the examination data from actual lectures through the detection scores.

2 DATA COLLECTION

In an information science course at our university, we collected e-book clickstream data from April to July 2021. The course was mainly offered for first-year students in a design department. A total of 114 students took an information science course online. The teacher explained the digital textbook provided by an e-book system and streamed the voice-only via Microsoft Teams to encourage the students to read the e-book provided by BookRoll (Ogata et al. 2015). The e-book clickstream data contain the timestamps and the page numbers. In the last lecture, the students took an online examination with 20 multiple choice questions and five essay-type questions on Moodle. The examination started simultaneously, and the time was 40 minutes. The students must not access any teaching materials or discuss with the other students.

3 DETECTION SCORE FOR STUDENTS SUSPECTED OF CHEATING

We propose a detection score based on essay analysis based on natural language processing and the reading time of e-books. Note that we can compute the reading time on each page from the clickstream data. The detection score D represents a suspicious level for students suspected of cheating. We compute D by adding an essay similarity score E and a reading time score R. Therefore, in the student i's detection score, $D(i) = \alpha E(i) + (1 - \alpha)R(i)$. In this study, $\alpha = 0.5$ for evaluating the scores using the same weight. The essay similarity score represents a similarity level between students' essays. To compute E, we apply a pre-trained BERT model (Devlin, Chang, Lee, & Toutanova, 2019) to the written essays for extracting the essay features. The BERT is a deep neural network called a transformer and can extract a feature vector from sentences. If the feature vector of student A's essay is similar to the features of another student's essay, then student A may have shared the answer with the student. This behavior would be considered cheating. We compute E based on the essay feature similarity as follows: $E(i) = \exp\left(-\frac{|f_i - f_{i^*}|^2}{2.5\,\sigma_e}\right)$, where f_i is an essay feature vector from the final hidden vector of [CLS] of BERT, f_{i^*} is the nearest neighborhood of f_i, and σ_e is the standard deviation of distances $|f_i - f_{i^*}|^2$ for all students. The score can be evidence. However, the answers may happen to be similar. To consider cheating from another perspective, we focused on e-book reading time as a learning behavior before the examination. If students are not learning or even accessing enough e-books, they cannot write an essay, and therefore the essays written cannot be similar. In other words, the student is more suspect when a student has not accessed the e-book system, and the essay is similar. We compute R based on the reading time as follows: $R(i) = \exp\left(-\frac{t_i}{2.5\,\sigma_r}\right)$, where t_i is the reading time of student i, and σ_r is the standard deviation of the reading time for all students. We can detect students suspected of cheating using the detection score. In this study, the detection rule is $D(i) \geq \mu_d + \lambda\sigma_d$, where μ_d and σ_d are the mean and the standard deviation of $D(i)$. We set 0.8 to the hyperparameter λ to control the detection sensitivity.

4 CASE STUDY AND DISCUSSION

We analyzed one essay-type question written by the 81 students in the information science course. When grading the essays, the teacher suspected that cheating had occurred. We applied our scoring method to essays written by students and e-book clickstream data from one week before the examination because most of the students reviewed in the term. Reading time was computed for the most relevant page to answer the question. Figure 1 shows detection scores on the essay feature space visualized by t-SNE, a visualization method. In this figure, each marker point corresponds to a student's essay. t-SNE places similar essays close together. The star points mean detected students, and the circle points are non-detected students. Black lines connect students writing the most similar essay. In addition, we show the essay examples written by two students and the result of the essay similarity score only for comparison. The proposed method could detect very similar essays. In Figure 1, the essay of the detected student A was almost identical to the essay of student B. The spelling errors shown in red letters were the same. Note that the spelling error was found by us manually after the detection.

Figure 1: Scoring results in the essay-type question.

As shown in Figure 1, we could detect student A based on the essay similarity score only. However, the proposed method can highlight more suspicious students by combining the reading time. For example, many of the students detected by the proposed method did not review the most relevant page of the e-book. Some of the students had not browsed the relevant pages of the e-book since about 40 days before the exam. Although there was no evidence that these students had learned the contents sufficiently, they wrote similar essays. Our method could provide another insight. We observed that some students, like Student B reading the page, wrote essays similar to the essays of the detected students. Such students might be the original authors of the essays. For investigating the potential of the other learning logs, we focused on the learning logs of student A with the highest detection score. Moodle's auto attendance module recorded this student's attendance, but the student did not access the e-book system for most of the second half of the lecture. This behavior was different from the behavior of the other students. In addition, the examination grade point was lower than the average point. Therefore, we could conclude that student A had not learned the contents enough. This analysis indicates that teachers can use further evidence as reference information by combining the other learning logs, such as Moodle logs. As the limitations of this study, we did not have ground-truth data of cheating students because of the learning log and the examination data obtained from actual lectures. In addition, we did not investigate students' prior content knowledge and other learning activities, such as discussions and notes written by students. Our method might detect students having short reading time and a lot of prior knowledge. However, we showed that our scoring method could provide insights that only the essay analysis could not find. Therefore, as our answer to the research question, combining the learning logs and the examination results could provide more variable insights.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP21K17864.

REFERENCES

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL 2019* (pp. 4171–4186).

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. *Proceedings of the 23rd International Conference on Computer in Education (ICCE 2015)*, pp. 401-406.

Rogerson, A. M. (2017). Detecting contract cheating in essay and report submissions: process, patterns, clues and conversations. *International Journal for Educational Integrity*,13, 1-17.

Analyzing Readability of Scientific Abstracts for ESL Learners

Yo Ehara
Tokyo Gakugei University
ehara@u-gakugei.ac.jp

ABSTRACT: Education in Science, technology, engineering, and mathematics (STEM) is considered important as it often ensures successful careers for learners. However, the advanced content in these fields is available only in a limited number of languages, such as English. Therefore, English-as-a-second-language (ESL) learners of STEM are often additionally required to learn English. Most scientific publications come with abstracts, which are essential in determining how necessary its contents are for the learners of STEM in English. This study investigates this under-investigated but important issue. Ideally, this would be investigated by having ESL learners actually read scientific articles. However, doing so requires prohibitively high costs in terms of both time and money because scientific papers are highly sophisticated and are difficult to evaluate. Instead, in this study, we employed two types of highly accurate automatic readability assessment methods for evaluating text readability of scientific texts for ESL learners. Experiment results show that the percentage of difficult-to-read texts varies with scientific fields, although intermediate ESL learners generally find scientific texts to be readable.

Keywords: Readability, Scientific Texts, English-as-a-Second-Language Learners

1 INTRODUCTION

Scientific publications are predominantly written in English, which is a second language for numerous scientists and students. Hence, the readability of scientific publications for English as a second language (ESL) learners is essential for determining and developing the support they need to learn STEM (Zhou and Bhat, 2021). A language gap between native and ESL learners may result in misinterpretation of scientific papers, which would significantly hinder the development of science. However, few studies have investigated this issue. In this study, we assessed the readability of scientific publications for ESL learners. Specifically, we followed an existing study on readability for ESL learners (Martinc et al, 2021). Readability evaluation of the main body of science papers written in English would be too technical to be appropriate. Therefore, we studied the readability of the title and abstract, which are typically used to determine the importance of the main body. To avoid biasing our analysis to one particular field, we obtained abstracts from the databases of two different fields: 55,410 abstracts in medical and life sciences field from PubMed, and 27,686 abstracts in natural language processing field from the ACL Anthology. A large-scale manual readability assessment is impractical because of financial and time constraints, we constructed two contrastive automatic readability assessors.

2 AUTOMATIC READABILITY ASSESSMENTS

Considering the assesment of N texts: we express their set as $\{\mathrm{T}_i | i \in \{1, \ldots, N\}\}$. Let Y be the set of readability labels that are ordered by their difficulty. For example, in the OneStopEnglish dataset

(Vajjala and Lučić, 2018), we can set Y={0, 1, 2}, where 0 is elementary, 1 is intermediate, and 2 is advanced. The number of levels depends on the evaluation corpus. Using Y, we labeled T_i as $y_i \in$Y. An assessor evaluates each text T_i and outputs its readability score s_i. In a supervised setting, the assessor knows the number of levels in the evaluation corpus from its training examples. Therefore, s_i values range within Y: $s_i \in Y$. However, in an unsupervised setting, it is noteworthy that the assessor knows neither Y nor the number of levels in the evaluation corpus, because no label is provided. Hence, even if only integers are allowed for y_i, s_i values can be real numbers.

Arrays are generally expressed in the form [1, . . . , n]. Given N texts [$T_i | i \in$ {1, . . . , N }], our goal is to make an assessor output an array of readability scores [$s_i | i \in$ {1, . . . , N }] that correlates well with the array of labels [$y_i | i \in$ {1, . . . , N }]. Multiple types of correlation coefficients exist between the arrays of scores and labels. Ap peer the conventions, we used rank coefficients such as Spearman's ρ, defined as the Pearson's ρ between rankings, when s_i is real-valued. Our first type of assessors, which is based on the bidirectional encoder representations from transformers (BERT) (Devlin et al, 2019), was trained using the OneStopEnglish dataset. Each text was annotated by a language teacher. BERT is the current standard for building highly accurate text classifiers and considers textual contexts for its assessment. However, it does not use the information directly taken from ESL learners.

To solve this problem, we built a vocabulary-based assessor, which was trained solely on a dataset of vocabulary test results of ESL learners without using any text readability labels. It perfroms assesment by calculating the bag-of-words probability that the average ability test-taker knows all the words in a given text and regards its negative logarithm as the readability of the text. To obtain the bag-of-words probability that a learner knows a word from a vocabulary test result, we followed the method proposed by Ehara (2018), which is based on the item response theory (IRT) and logistic regression. To train and evaluate our classifiers, we used the publicly-available vocabulary test result dataset provided by Ehara (2018), in which 100 second-language learners took a test and their responses were collected. Their data were published and made publicly available. This dataset uses questions from the vocabulary size test, which is a widely used vocabulary test in applied linguistics (Beglar and Nation, 2007). It consists of 100 questions each of which asks about a word in a multiple-choice format. To better estimate the word difficulty, we considered the frequency values provided by the British National Corpus and the Corpus of Contemporary American English as word features.

3 EXPERIMENTS AND CONCLUSIONS

We used the OneStopEnglish dataset as the source of readability for ESL learners owing to its recency, public availability, and reliability. It has three levels: elementary, intermediate, and advanced, all of which have 189 texts each. This amounts to 567 texts in total. We randomly split these texts into training, validation,and a test sets consisting of 339, 114, and 114 texts, respectively. The training and validation sets were used only to train the supervised methods for comparison. Whereas, the unsupervised methods used only the test set.

We applied the standard BERT-based sequence classification approach involving pre-training and fine-tuning because this approach is known to provide excellent results. Specifically, we used BERT-large-cased-whole-word-masking in the Huggingface models (https://huggingface.co/models) for the pretrained model, which was then fine-tuned using the 339 training texts. This fine-tuned model was named spvBERT, where “spv” implies “supervised”. The Adam optimizer with a setting of 10 epochs and a 0.00001 training rate was used for fine-tuning. For the implementation of conventional readability formulae, we used the readability PyPI package (https://pypi.org/project/readability/). The performance results are provided in Table 1, where TCN RSRS-simple is the previous unsupervised state-of-the-art (Martinc et al, 2021). We can see that vocabulary-based and spvBERT methods achieved the highest scores, implying that the our assessors are highly accurate considering the current state-of-the-art. Despite these differences, our experimental results, whicih are provided in Table 2, showed that the assessments of these two assessor types were similar. First, for both databases, the former assessor judged that the majority of the abstracts were readable to intermediate English learners, as defined by Martinc et al. (2021).

Second, both assessors judged that the abstracts retrieved from the ACL Anthology were easier than those retrieved from PubMed. For the former assessor, this is obvious from the aforementioned classification results into the three levels. While for the latter, the average readability scores for the ACL Anthology and PubMed were 18.45 and 31.25, respectively, where a larger score indicates lesser readablility. Both these results, which indicated that the ACL Anthology abstracts were easier than the PubMed abstracts, were statistically significant (the Mann-Whitney tests, $p < 0\ .01$). This is presumably because medical terminology, which is primarily of Greek origin and frequently used in PubMed, was particularly difficult for ESL learners. The qualitative results of the vocabulary-based assessor confirmed this tendency. For example, the particularly difficult word for ESL learners in PubMed were found to be: hemihydrate and engraftment. In contrast, similarly difficult words in the ACL Anthology were lexicosemantic and colingual. In conclusion, we showed that, 10%-30% of the scientific texts are not readable to intermediate ESL learners, implying that they need assistance to read them. Notably for ESL STEM learning, these percentages varied with academic fields.

Table 1: Performance of Proposed Methods

Method	Spearman's ρ	Pearson's ρ
Flesch-Kincaid	0.324	0.359
TCN RSRS-simple	-	0.615(*)
Vocabulary-based	**0.730**	**0.715**
spvBERT	0.866	0.864

Table 2: Readability of Scientific Publications

-	Elem.	Int.	Adv.
ACL Anthology	0.037	0.860	0.103
PubMed	0.006	0.639	0.305

REFERENCES

David Beglar and Paul Nation. A Vocabulary Size Test. The Language Teacher, 31(7): 9–13, 2007.

Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In NAACL, pp. 4171–4186, 2019.

Yo Ehara. Building an English Vocabulary Knowledge Dataset of Japanese English-as-a-Second-Language Learners Using Crowdsourcing. In LREC, May 2018.

Matej Martinc, Senja Pollak and Marko Robnik-Šikonja. Supervised and Unsupervised Neural Approaches to Text Readability. Computational Linguistics, 47(1): 141–179, April 2021.

Sowmya Vajjala and Ivana Lučić. OneStopEnglish Corpus: A New Corpus for Automatic Readability Assessment and Text Simplification. In BEA, pp. 297–304, 2018.

Jianing Zhou and Suma Bhat. Modeling Consistency Using Engagement Patterns in Online Courses. In LAK21, 2021, pp. 226–236, 2021.

Understanding the Relationship between Learning Motivation and Academic Performance in Mobile Language Learning

Zahra Azizah
Tohoku University, Japan
zahra.azizah.s6@dc.tohoku.ac.jp

Takashi Mitsuishi
Tohoku University, Japan
takashi.mitsuishi@tohoku.ac.jp

Yuichi Ohkawa
Tohoku University, Japan
kuri@tohoku.ac.jp

Xiumin Zhao
Tohoku University, Japan
xiumin.zhao.e2@tohoku.ac.jp

Tomoya Ohyama
Tohoku University, Japan
tomoya.ohyama.e6@tohoku.ac.jp

Yoshinori Hayakawa
Tohoku University, Japan
yoshinori.hayakawa.d4@tohoku.ac.jp

ABSTRACT: In this poster, we present the study results that examined the relationship between students' learning motivation when using mobile learning applications as part of a blended learning environment and their academic performances in class. Learning motivation is a behavior that indicates students' desire to achieve learning goals. The ability to manage learning goals is a sign of Self-Regulated Learning. Our result revealed some types of behavior identified as learning motivation: regularity of studies (RS), early learning, and repetition. These results extend our understanding of the relationship between learning motivation and students' academic performance in second language learning.

Keywords: self-regulated learning, learning behaviors, blended learning, learning analytics

1 INTRODUCTION

Motivation has been widely believed to be one of the critical parameters influencing success in the second language (L2) learning (Gardner, 2019). This study aims to identify students' behavior traits as learning motivation and correlate with academic performances. We use the time engagement concept from (Nguyen et al., 2018), which uses learning time spent per week as a parameter linked to learning design and academic performances. We also adopt parameter "anti-procrastination" from (Li et al., 2018) for measuring self-regulated learning (SRL). We use time spent per week to examine how students' regularity of studies (RS) compared to the teacher's recommendation learning time. As a parameter similar to anti-procrastination, early learning is a score of how early completed learning material compares with total learning activities. Last, we add an observable parameter (repetition) to examine time spent on repetition mode on weekly learning.

2 METHODS

The data comes from Spring-Fall semester 2019 of mobile language learning applications (Ohkawa et al., 2019), that support Chinese language classes, as elective foreign language courses in a university in Japan, for first-year students, from seven classes with different backgrounds. Four Chinese language skills are trained in the apps: pronunciation, speaking, listening, and grammar. The apps focus on helping students review the materials also practice the language. The teacher will check the apps whether the students learned assigned material based on the deadline or not. Our study only considers students who completed Spring – Fall semester and used the apps. Overall, from 239 students, remained 221 students fulfilled these criteria. First, we collect the data from log systems that record students' activities and timestamps. Then, to differentiate one learning session from another, we assume that if there is no activity within 20 minutes, they stop the session. After that, we calculate the total time spent per week. To identify learning motivation traits and their relationship to academic performance, we use three significant parameters, that are:

1) Regularity of studies (RS): total time spent per week compared to teacher's recommendation learning time, 15 minutes per week.
2) Early learning: degree how early students completed learning material compared to total activities.
3) Repetition: total time spent per week based on repeat material that has been learned on the assigned week.

On the RS parameter, we conducted clustering using mini-batch k-means and decided four as the number of clusters. Every cluster has learning time spent across weeks showing different patterns. Compared to the teacher's recommendation learning time (15 minutes), if time spent more than the teacher's recommendation, it indicates regularity of studies. We conclude clusters showed regular and less regular studies patterns.

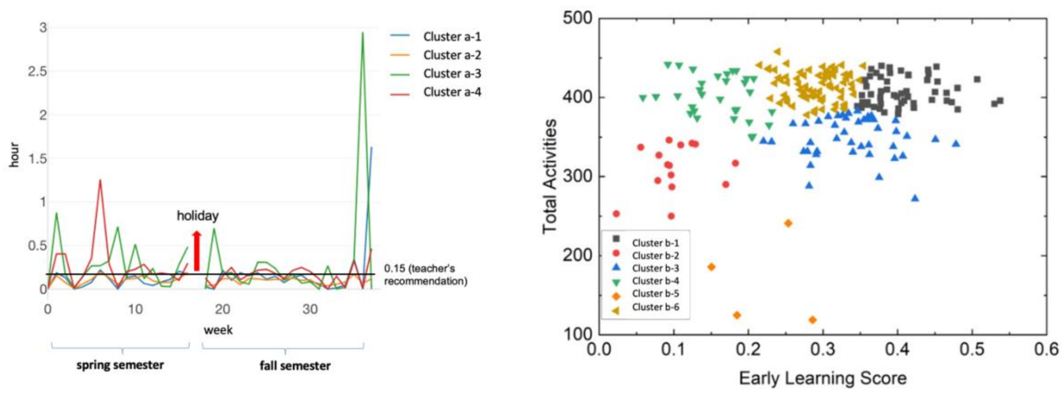


Figure 1: a) study patterns weekly (left), b) cluster of early learning score compared to activities (right)

The second parameter is "early learning." It is calculated by comparing the total available days and the lead days when each learning activity was completed, then sum up. Every student will have a unique early learning score clustered into six clusters, obtained with the Elbow method, then using k-means with the Hartigan-Wong algorithm. Total activities calculated based on unique quiz-taken, watching the video, and listening to audio. The third parameter is repetition; it is computed time spent per week based on repetition mode. We conducted linear regression to compare average grades of Spring and Fall exams with a group of performances students:

1) High-performances: students with average score of more than 80,
2) Medium-performances: students with average score between 60 to 80,
3) Low-performance: students with average score less than 60; and using medium-performances group as the baseline,

overtime with the time spent.

3 RESULTS AND DISCUSSION

From **Figure 1 (a)**, there are two regular studies patterns (cluster a-3 and cluster a-4) and two less regular studies patterns (cluster a-1 and cluster a-2). This result compares learning time spent per week with the teacher's recommendation. In **Figure 1 (a)**, regular studies patterns have many above teacher's recommendations over weeks, while less regular studies patterns only have a few. As a result, 19% of high-performance students use regular studies patterns, while only 3% of medium-performances and 4% of low-performances students use these clusters.

The second parameter is early learning. In **Figure 1 (b)**, we labeled some clusters as:

1) Cluster b-1 and b-6: **highly motivated** (130 students); characterized by a high early learning score and many activities.
2) Cluster b-2 and b-5: **unmotivated** (22 students); characterized by low early learning score and relatively low activities.
3) Cluster b-3: **inconsistent** (42 students); relatively early learner but moderate tasks completer.
4) Cluster b-4: **procrastinator** (27 students); task completer but triggered by the deadline.

Third parameter is repetition. Using linear regression we found that high performance student has positive effect (B = 0.0072, SE = 0.0153, $p < 0.05$) to repetition. While low performance student has no effect (B = -0.0021, SE = 0.6533, n.s.) to repetition.

We consider these parameters because we only use students' log activities as a dataset. We neglect other possible learning motivations, such as the need to master the Chinese language to find a job, maintain the Chinese language skills they already have, etc. It is difficult to measure this kind of motivation as it is intangible. We consider elaborating our study with surveys or interviews for future work to find more learning motivations in second language learning.

REFERENCES

Gardner, R. C. (2019). The socio-educational model of second language acquisition. In *The Palgrave handbook of motivation for language learning* (pp. 21-37). Palgrave Macmillan, Cham.

Li, H., Flanagan, B., Konomi, S. I., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research and Practice in Technology Enhanced Learning*, 13(1), 1-12.

Nguyen, Q., Huptych, M., & Rienties, B. (2018, March). Linking students' timing of engagement to learning design and academic performance. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 141-150).

Ohkawa, Y., Kodama, M., Konno, Y., Zhao, X., & Mitsuishi, T. (2019, October). Development and Evaluation of Smartphone Learning Material for Blended Language Learning. In *2019 4th International Conference on Information Technology (InCIT)* (pp. 108-113). IEEE.

Learning Analytics in Collaborative Online Lab Environments: A Systematic Scoping Review

Hanna Birkeland **Mohammad Khalil** **Barbara Wasson**

Centre for the Science of Learning & Technology (SLATE), University of Bergen

hbi008@uib.no mohammad.khalil@uib.no barbara.wasson@uib.no

ABSTRACT: With the increase in online education arises the challenge of replicating physical laboratories and teamwork online. This poster presents a brief overview of a scoping review of the current research within the area of learning analytics and collaboration in online laboratory environments, including results on the used learning analytics method. A gap of knowledge in the current research where there is a lack of research addressing this topic has been identified. This calls for the need of further research within this area.

Keywords: Collaboration, Learning Analytics, Online Laboratories, Online Learning

1 INTRODUCTION

The importance of technology in education is increasingly growing. The momentum has been evident after the outbreak of the COVID-19 pandemic when typical in-person classroom teaching was suspended worldwide and the educational systems were forced to move into virtual environments. In biosciences, where laboratory sections are at the core of undergraduate education, finding a feasible way of conducting laboratory work in a virtual space is challenging, yet crucial. The challenge is represented by replicating hands-on exercises and teamwork online to meet the pedagogical standards of university discourses and the desired outcome of such exercises. The progress within technology and communication networks has made it possible to develop virtual and remote laboratories that allow students to conduct experiments online and find a way around the limitations of physical laboratories (Alkhaldi et al., 2016). Collaboration and teamwork are at the heart of the common practice in physical laboratories (Teng et al., 2016). With the increase in teaching and learning online, the challenge of facilitating cooperative learning emerges in these online environments. The work of which learning analytics acts to better understand learning performances during collaboration. Learning Analytics may offer students and teachers insight into the interactions within a group. Such information benefits teachers to facilitate their teaching to each group, and students to self-reflect during collaboration. As a part of a master thesis work on this area, and to investigate current research regarding the use of learning analytics in online laboratories and collaboration, a systematic scoping review has been conducted, providing an overview of the volume and nature of existing literature on the given topic. Methods within learning analytics are identified and gaps of knowledge are discussed. Online laboratories will be used as the common term for virtual and remote labs.

2 STATE OF THE ART

The established query string for our search is *(learning analytics) AND (virtual lab* OR online lab* OR digital lab* OR remote lab*)*, which resulted in 419 articles using ELSEVIER Web of Science digital database. We followed PRISMA guidelines by scanning abstracts of the results and then full-text scanning for those fulfilling the criteria. The inclusion criteria were expanded to online learning environments, as few studies address all the factors of the given topic: collaboration, learning analytics and online laboratories. Only 11 articles were selected for the review. Because of the limited space of the poster, we uploaded the included studies via this link https://bit.ly/3E0qx7k

The results from the review revealed few studies addressing all the factors of the given topic: collaboration, learning analytics and virtual/remote/online laboratories. Romero et al. (2015) provides an example of the use of learning analytics in the Weblab-Deusto remote laboratory platform. A software layer over the platform registers data regarding the students' interactions with the experiment, recording clicks and traces within the system and storing it in a database. The analyzed data are presented to both teacher and student through a software showing differences and similarities between the exercises performed by the student compared to the teacher's execution. It is stated that such data can be processed for both individual students and groups of students, however this study focuses primarily on the performance of the individual student. Qvist et al. (2015) present a similar example of learning analytics use in the LabLife3D virtual laboratory environment where they store data of student mouse clicks and time spent on tasks. The analyzed data are presented through timelines of data trails from the executed experiments, enabling teachers to identify occurring errors and students to reflect on their learning process. These laboratory experiments also do not yet provide collaboration amongst students as they focus primarily on the individual learner. The collaboration of students has been further investigated by Orduña et al. (2014) in the same Weblab-Deusto remote laboratory platform as in Romero et al. (2015). Social network analysis was applied to analyze data on uploaded files to the system, as the platform does not store interaction data between students. By checking files shared among students, one can identify who is sharing with who, and who is receiving. This information allows teachers to establish those in need of help, as students who often receive files might be struggling.

There are few remote laboratory environments that allow collaboration within the system, however, Teng et al. (2016) developed one that does. The NetLab remote lab allows students to use the system at the same time as other students and provide them with a built-in chat window for communication. All actions made by students online are broadcasted in their own window. The planned future work of their study is to employ learning analytics methods to analyze those data.

Several research papers regarding discussion forums have been identified. In a more recent study, Doleck et al. (2021) measure the performance of social learning networks in discussion forums through an algorithm which offers to optimize these networks by connecting users with similar tendencies. Pillutla et al. (2020) demonstrate a different example of learning analytics in discussion forums through text classification. A central part of learning analytics is the visualization of learner data. An example of this is presented in Tarmazdi et al. (2015), where the authors have developed a learning analytics dashboard for teamwork in an online computer science course. By combining Natural Language Processing (NLP), information retrieval techniques, and sentiment analysis, the dashboard provides monitoring and analysis of the roles within each group. This allowed the teacher to identify struggling students and teams by looking at how the teams engage in the work.

3 DISCUSSION

There is an apparent knowledge gap within the research of learning analytics in online laboratories and collaboration in labs. Many of the identified studies have not sufficiently developed learning analytics approaches, collaboration, or both. The lack of research in this area could result from the lack of collaborative online laboratory environments. Five years ago Teng et al. (2016) stated that few such environments allow collaboration within the system. This lack seems to persist, given that the results of this scoping review support this claim. Looking at the current work of learning analytics and collaboration in other online learning environments is therefore essential as they provide examples of how it might be implemented in an online lab environment. The most common topics identified in the studies include social network analysis and natural language processing, suggesting these methods are valuable in the analysis of collaboration and are potentially applied in understanding collaboration in online labs. The knowledge gap in this scoping review identifies collaboration in online labs as an area for further investigation.

ACKNOWLEDGEMENTS

This study is co-funded by the Erasmus+ programme of the European Union. Project ENVISION_2027 (2020-1-FI01-KA226-HE-092653).

REFERENCES

Alkhaldi, T., Pranata, I. & Athauda, R.I. (2016). A review of contemporary virtual and remote laboratory implementations: observations and findings. *J. Comput. Educ.* 3, 329–351.

Teng, M., Considine, H., Nedic, Z., & Nafalski, A. (2016). Current and Future Developments in Remote Laboratory NetLab. *International Journal of Online and Biomedical Engineering (iJOE*), 12(08), pp. 4–12. https://doi.org/10.3991/ijoe.v12i08.6034

Romero, S., Guenaga, M., García-Zubia, J., & Orduña, P. (2015). Automatic Assessment of Progress Using Remote Laboratories. *International Journal of Online and Biomedical Engineering* (iJOE), 11(2), pp. 49–54. DOI: https://doi.org/10.3991/ijoe.v11i2.4379

Orduña, P., Almeida, A., Ros, S., López-de-Ipiña, D., & García-Zubía, J. (2014). Leveraging Non-explicit Social Communities for Learning Analytics in Mobile Remote Laboratories. *J. Univers. Comput. Sci*., 20, 2043-2053.

Qvist, P., Kangasniemi, T., Palomäki, S., Seppänen, J., Joensuu, P., Natri, O., Närhi, M., Palomäki, E., Tiitu, H., & Nordström, K. (2015). Design of Virtual Learning Environments: Learning Analytics and Identification of Affordances and Barriers*. International Journal of Engineering Pedagogy* (iJEP), 5(4), pp. 64–75. https://doi.org/10.3991/ijep.v5i4.4962

Kent, C., & Cukurova, M. (2020). Investigating Collaboration as a Process with Theory-driven Learning Analytics. *Journal of Learning Analytics*, 7(1), 59–71.

Doleck, T., Lemay, D. J., & Brinton, C. G. (2021). Evaluating the efficiency of social learning networks: Perspectives for harnessing learning analytics to improve discussions. *Computers and Education, 164*, [104124]. https://doi.org/10.1016/j.compedu.2021.104124

Pillutla, V.S., Tawfik, A.A. & Giabbanelli, P.J (2020). Detecting the Depth and Progression of Learning in Massive Open Online Courses by Mining Discussion Data. *Tech Know Learn 25*, 881–898.

Tarmazdi, H., Vivian, R., Szabo, C., Falkner, K. & Falkner, N. (2015). Using Learning Analytics to Visualise Computer Science Teamwork. In *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education* (ITiCSE '15), 165–170.

New Perspective on Input Feature Analysis for Early Feedback by Student Performance Prediction Considering the Future Effect

Ryusuke Murata*, Fumiya Okubo, Tsubasa Minematsu, Yuta Taniguchi, and Atsushi Shimada
Kyushu University, Japan
*murata@limu.ait.kyushu-u.ac.jp

ABSTRACT: For the student performance prediction task, the analysis of input features is as important as improving accuracy. Such an analysis helps us to pinpoint the features that contribute to the prediction result, which can be used to provide feedback to students and teachers. However, there is a gap between the input features of the early prediction model and the real course setting: the final grade of the course is the result of the learning activities throughout the course, although the early prediction model estimates the final grade from the learning activities during the early stages of the course. This study thus attempted to fill in the gap by using an early prediction model called the RNN-FitNets. In the experiment, it was analyzed whether the RNN-FitNets could consider the input features not only for early stages, but also for the entire course. The experimental results showed that the RNN-FitNets treated the input features in the early stages as if they were the features of the entire course; therefore, the RNN-FitNets can fill the gap in the early feedback, and the stakeholders can understand the input features to be more focused on at the time from the perspective of the entire course.

1 INTRODUCTION

With the emergence of learning analytics, many studies on student performance prediction using machine learning models have been conducted. In the student performance prediction task, it is important to analyze the contribution of input features (learning activities inputted to the prediction model), as well as to improve prediction accuracy. Understanding what input features are important for prediction results helps us to recognize the behavior of black-boxed prediction models, and this can be used to provide feedback to students and teachers. The students and teachers are more likely to accept the prediction result because they can understand why such a prediction result was made. However, there is a concern regarding the feedback on ongoing courses. When providing feedback in the early stages of a course, the prediction model using early input features does not know the future effect of the input features; however, prediction target such as final grade is the result of learning activities during the entire course. The gap may cause an early prediction model to provide wrong feedback, focusing on different input features from the perspective of the entire course. This study thus investigated whether the view of input features throughout a course can be obtained from the input features in the early stages of the course by using an early prediction model termed RNN-FitNets (Murata et al., 2021). RNN-FitNets aims to learn whole time series information in the early time steps. For more detail about the RNN-FitNets, please refer our previous work (Murata et al., 2021). The contribution of this paper is providing a new perspective on input feature analysis for feedback, considering the entire course.

2 VERIFICATION OF THE ABILITY TO CONSIDER THE ENTIRE COURSE

To investigate whether the RNN-FitNets can consider the effect of input features for the entire course from early stages, we compared the RNN-FitNets with a recurrent neural network (RNN) model, which has been recently used and is the foundation of the RNN-FitNets. If the RNN-FitNets can consider the entire course, the contribution of input features during the early stages should become similar to the RNN model for the final lecture week. The procedure of comparison is as follows: (1) train the RNN model for an early stage and for the final week, (2) train the RNN-FitNets for an early stage, (3) apply the prediction models to the test dataset (4) calculate the contribution of input features by SHAP values (Lundberg & Lee, 2017), and (5) compare the results. We selected the fourth week as the early stage because there were two course registration/modification periods by the fourth week in Kyushu University where log data were collected, that is, most students who remained after the fourth week took the course, and the first feedback at the time should be useful.

We used the data collected from the Digital Signal Processing courses during the first semester of 2020 and 2021, held for third-year undergraduate students as elective courses in the Department of Electrical Engineering and Computer Science. We used the 2020 data to train the prediction models and the 2021 data to test the models. The course involved 13 weeks of lectures; however, the first week was excluded from the dataset because it was a syllabus day. For both 2020 and 2021, a total of 89 students took the course. We defined the students who took grade points "A" or "B" as "no-risk" students (2020: 78.7%, 2021: 73.1%), and the students who took grade points "C,” "D,” or "F" as "at-risk" (2020: 21.3%, 2021: 26.9%). Hence, the prediction task became the binary classification for detecting at-risk students. In the courses, a learning management system Moodle and a digital textbook system BookRoll collected learning logs. To input the log data into the prediction models, we aggregated the logs in each lecture week and converted them into active learner points (ALPs) (Okubo et al., 2017), which was made to analyze courses using Moodle and BookRoll. ALPs consist of nine learning activities; however, due to the difference of course design from (Okubo et al., 2017), we used seven leaning activities: three types of activities in Moodle (attendance, report submission, and number of course accesses) and four types of activities in BookRoll (reading time, number of highlights, number of notes, and total number of actions). ALPs evaluate the activities in a range from 0 to 5. Please refer (Okubo et al., 2017) for the detail about ALPs.

Figure 1 shows the contribution of input features for an at-risk student. There were many at-risk students, but the contribution patterns were similar. For this reason, we selected this at-risk student as a representative pattern. In Figure 1, the colors of the stacked bar, *y*-axis, and *x*-axis represent the ALP values inputted into the prediction models, the input features’ name, and SHAP value for input features, respectively. The positive/negative SHAP value means a degree of contribution to predict at-risk/no-risk. Seeing Figure 1, although RNN model in the fourth week (left graph) did not so focus on the late of report submission (yellow bar), the feature exhibited a risk to become “not submitted” (blue bar) in the future and RNN model in the final week (center graph) took it as a high probability of at-risk. In contrast, the RNN-FitNets in the fourth week (right graph) increased the SHAP value for the late of report submission and could consider the future risk. This result shows that the feedback by the RNN-FitNets can encourage attention to the feature to be more focused on from the perspective of the entire course. Moreover, the f-measure value for detecting at-risk students in the fourth week was increased by the RNN-FitNets (0.800) from the value of the RNN model (0.667).

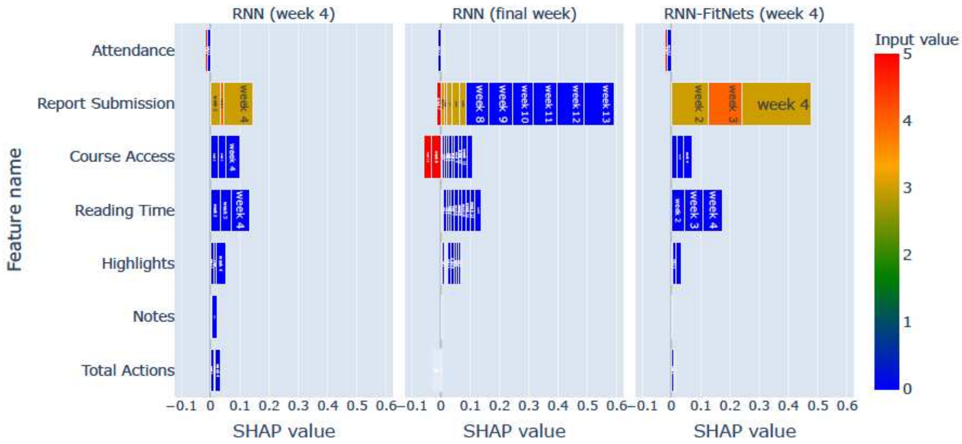


Figure 1: Contribution of input features for a prediction result

3 CONCLUSION AND FUTURE WORK

This study proposed a new perspective on the input feature analysis in the early stages of a course to consider the entire course using the RNN-FitNets. The experimental results showed that the RNN-FitNets could emphasize important learning activities that should be paid more attention during the early stages, which the conventional RNN model cannot consider. Consequentially, we can expect that the RNN-FitNets will show a great potential to provide stakeholders with new insights. For example, in the case of the course in our experiment, the teacher can understand future risk of the late of report submission more seriously and can take early measures to encourage the submission, such as sending a reminder and making time to work on the report assignment during the rest of lecture time following the explanation of lecture content from the teacher. Further work will be required; we need to investigate other case studies, discuss the feedback method, and conduct user evaluations of the effect of the feedback such as a questionnaire survey.

ACKNOWLEDGEMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768-4777).

Murata, R., Shimada, A., & Minematsu, T. (2021). Early detection of at-risk students based on knowledge distillation rnn models. *Proceedings of the 14th international conference on educational data mining* (pp. 699–703). 3

Okubo, F., Yamashita, T., Shimada, A., & Konomi, S. (2017). Students' performance prediction using data of multiple courses by recurrent neural network. *Proceedings of the 25th International Conference on Computers in Education*, (pp. 439-444).

Coding Trajectory Map: Student Programming Situations Made Visually Locatable

Yuta Taniguchi, Tsubasa Minematsu, Fumiya Okubo and Atsushi Shimada
Kyushu University, Japan
taniguchi.yuta.941@m.kyushu-u.ac.jp

ABSTRACT: Understanding student activities in programming exercise is vital for teachers to support students. Because teachers cannot always pay attention to every student's source code, they have difficulties in identifying students who are having trouble with errors and understanding how students write source code. The representative approach of using dashboard systems, however, has limitation on the depth of understanding of students' situations due to the lack of intuitive presentation of the processes in which source code contents evolve. This study aims to provide beneficial information to teachers about the temporal changes of student coding situations. Based on a dataset of source code snapshots taken frequently, we propose a feedback tool for teachers, called Coding Trajectory Map. We conducted an experiment for human evaluation of the tool, whose result shows that teachers can better understand the learning situations of students with the tool.

Keywords: Learning process, Learning situation, Programming, Visualization, Teacher support

1 INTRODUCTION

In novice programming education, many students struggle to write and debug programs (Robins et al., 2003). The difficulties in programming often lower learner motivation which is an important factor of success. Thus, supporting students in difficult situations is crucial. However, identifying such students is hard because help-seeking is known as a difficult metacognitive skill and, in addition, teachers cannot always pay attention to every student's learning situation. Learning analytics dashboards arises to overcome this challenge also in this field (Diana et al., 2017; Fu et al., 2017), which present to teachers an overview of their students' situations through statistics, performance indicators, and visualization, and so on. However, the depth of understanding via prior dashboards is limited because the temporal changes of source code content are less likely to be considered. For example, students' trial and error and coding strategies could be observed only from how source code changes over time, and it is hard to realize from statistical numbers or a single snapshot of source code. The research question in this study is "How can we help teachers obtain insights into student situations when we have a class-wide dataset of student source code snapshots?" Based on the research question, we aim to make student learning situations visually and intuitively locatable making use of a mass of snapshots that were taken frequently. Here, we mean by "locatable" that a learning situation is linked to a specific position on a visual map, where the distance between two positions intuitively reflects the difference in corresponding situations at source code level. In this poster paper, we present a method to create such a visual map, called Coding Trajectory Map, and report the result of a preliminary evaluation experiment on the usefulness of maps for teachers.

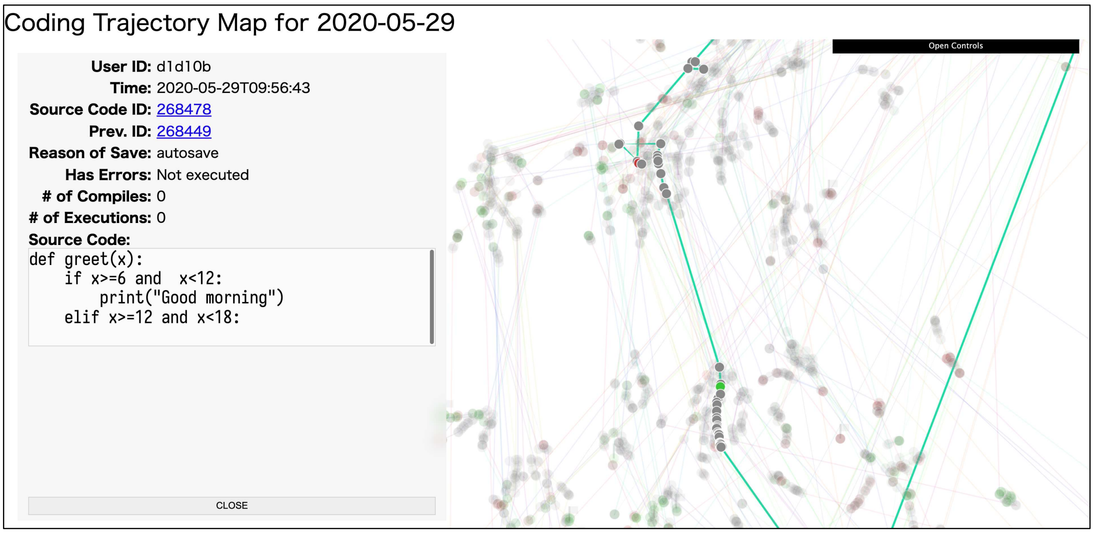


Figure 1: Coding Trajectory Map showing the distribution of source code on the right-hand side and the information of a selected source code snapshot on left-hand side.

2 METHODS

Dataset: We collect a set of source code snapshots in an introductory Python programming course offered in 2020 for first-year students at our university. We used our own online programming environment, called WEVL, to collect snapshots. With WEVL, a snapshot is taken three seconds after a student stops typing or when a student runs a program. In our dataset, about 84% of snapshots were taken by stopping typing.

Visualization: The proposed Coding Trajectory Map shows on a 2D plane a set of source code snapshots. The positions of snapshots are determined with t-SNE algorithm together with edit distance. First, edit distance is measured for every pair of snapshots in a dataset and a distance matrix is formed. An edit distance between a pair of source code is the minimum total number of adding, deleting, or substituting character-wise operations required to transform one into another. Intuitively, this is roughly the same as how much keyboard typing is necessary for the transformation. Second, the t-SNE algorithm is applied to the distance matrix, and we obtain 2-dimensional coordinates for each snapshot. The algorithm computes those coordinates trying to maintain the original distances, i.e., edit distance, as much as possible in the output space. Therefore, it is expected on the map that similar source code snapshots are placed nearby and the distance tell us approximately how much typing effort is required to reach a certain source code starting from one. Furthermore, because our snapshots were taken frequently, it is also expected that a series of snapshots from a student would be like a string of beads when visualized. Figure 1 shows an example of Coding Trajectory Map. On the right, each snapshot is shown as points, snapshots from the same student are connected by a line. On the left, the information of a snapshot that a user selected on the map is provided.

Evaluation: We asked 7 university teachers to evaluate the Coding Trajectory Map tool in the context of reflecting past classes in the same course as described above. Those teachers had experiences of teaching programming before. We provided instructions on how to use the tool in both text and video formats. Teachers were asked to learn the usage of the tools and then to use the tool themselves. We

Table 1: The result of questionnaire for evaluating the proposed tool. (N=7)

ID	Question	Mean	SD
Q1	The tool is useful for understanding the process of each student's exercise in detail by checking the content and status of the source code at any given time, including the presence or absence of runtime errors.	4.43	0.53
Q2	The tool is useful for understanding what kind of code was struggles for students.	4.29	0.76
Q3	The tool is useful for understanding each major and minor programming activity in the class.	4.14	0.69
Q4	The tool can help you identify students to whom you should pay attention.	3.86	1.07
Q5	The tool helps the teacher to reflect on the lesson.	4.00	0.82

also asked them to fill out a questionnaire with 5-point Likert scale to measure the degree of agreement on each question, where 1 stands for strongly disagree and 5 stands for strongly agree.

3 RESULTS AND DISCUSSION

Table 1 shows the result of the evaluation questionnaire. The tool was given the high score of 4.43 regarding the question Q1, which suggests the tool successfully provided the details of the individual coding activities. In terms of source code, we can say that the teachers thought the tool is helpful for finding one that is problematic to students (Q2). Concerning coding activity patterns, it is suggested that common patterns and abnormal activities could be easily identified by teachers (Q3). As for students, we can expect the tool may help us to notice students whom we should pay attention to (Q4) although the score is not very high. Overall, it is shown that Coding Trajectory Map has functions useful to find important code, students, and activities in programming exercise, and teachers agreed that the tool is useful to reflect on classes (Q5). We conclude that we successfully presented the tool which could be beneficial for gaining insights into problematic source code or major and minor coding activities from frequent source code snapshots. While this study focused on teacher support, this kind of feedback would be also useful for students because they will be able to relatively position their own situations on the map. As a future study, we are planning to evaluate the proposed tool in the context of student support.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP21K17863.

REFERENCES

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer science education, 13*(2), 137–172.

Diana, N., Eagle, M., Stamper, J., Grover, S., Bienkowski, M., & Basu, S. (2017). An instructor dashboard for real-time analytics in interactive programming assignments. In *Proceedings of the seventh international learning analytics & knowledge* (p. 272–279).

Fu, X., Shimada, A., Ogata, H., Taniguchi, Y., & Suehiro, D. (2017). Real-time learning analytics for C programming language courses. In *Proceedings of the seventh international conference on learning analytics & knowledge* (pp. 280–288).

Analyzing Correlation between Word Difficulty Levels Assigned by Language Teachers and those by Language Learners

Yo Ehara
Tokyo Gakugei Unviersity
ehara@u-gakugei.ac.jp

ABSTRACT: In second language acquisition, word difficulty is important for determining the order in which learners learn words. However, in many cases, openly available word difficulty scales are created by language teachers based on their own linguistic intuition and not necessarily obtained by directly analyzing second language learners. In this study, we will investigate the extent to which the two correlate. Specifically, we focused on a word difficulty scale for English-as-a-Second-Language (ESL) learners whose native language is Japanese created by applied linguists, namely CEFR-J. We compared this with word difficulty levels calculated from vocabulary test result data on ESL learners whose native language is mostly Japanese using item response theory. As a result of the experiment, we confirmed that they show a statistically significant moderate correlation in Spearman's ρ. This result suggests that language teachers' intuitions roughly fit actual word difficulty for language learners measured on testing. Interestingly, the scatter plot clearly showed that the words that teachers find difficult are not necessarily difficult to learners while those teachers find easy are also easy to learners.

Keywords: Teachers and Learners, Word Difficulty, Item Response Theory

1 INTRODUCTION

Second language learning differs from other types of learning in that it requires learning an enormous number of words. Since it takes a huge amount of time, many previous studies have suggested that these words should be learned at an initial stage (Nation, 2006). Although word difficulty is roughly correlated with word frequency on balanced corpora such as the British National Corpus (BNC), it is counter-intuitive to language teachers that words should be learned in the early stages of learning. Therefore, word difficulty scales have been developed by adjusting the corpus for word frequency, or by having language teachers directly replace the word frequency rankings (Nation, 2006).

In situations where there are many word difficulty scales, which one is appropriate to use? In many cases, we are more interested in what learners can actually do than in the word difficulty scale itself. One widely used measure for assessing what language learners can actually do with a foreign language is the Common European Framework of Reference for Languages (CEFR) [1].

The CEFR itself is merely a scale that defines what learners should be able to do at a given level. In other words, it does not specify what words one should know at what level. This is because the

[1] https://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages

vocabulary that a language learner acquires is greatly influenced by his or her first language. For example, among English as second language (ESL) learners whose first language is French, the words they need to learn in the early stages of learning English may be different from those whose first language is another language because French has many words with the same roots as English. For English learners whose first language is Japanese, the CEFR-J Vocabulary Profile (https://github.com/openlanguageprofiles/olp-en-cefrj) is the result of a survey of what words they should know at each level of the CEFR.

With this concept, word difficulty scales that learners should master have been created based on the linguistic intuition and analysis of language teachers. However, to what extent do these linguistic intuitions and analyses actually correspond to the actual vocabulary knowledge of language learners? Do the word difficulty levels of ESL learners with some experience in learning English match the intuitions of language teachers? These questions are important for ESL learners in the language acquisition process; however, few studies have examined this point. The objective of this study was to investigate this aspect.

2 DATA AND METHODS

First, as mentioned above, ESL learners' vocabulary knowledge is greatly influenced by their native language. The CEFR-J, which is the focus of this study, targets ESL learners whose first language is Japanese. Therefore, for a fair comparison, it is necessary to obtain the results of vocabulary tests for English learners whose first language is Japanese.

To achieve this goal, the dataset by Ehara (2018) was used. This is a publicly available dataset in with the responses of 100 learners to a 100-word test. In this test, each question is on a single word. In each question, a sentence containing the target word is presented within a sentence and test-takers are asked to choose the option that has the same meaning as the presented sentence.

Next, we describe the CEFR-J Vocabulary Profile. The CEFR profile lists the words at the A1, A2, B1, and B2 levels. The CEFR-J first collects the published texts for each level and identifies the words to be learned at each level from the words used in the texts. The total number of words in the dataset is 7,800. When multiple levels are available, the easier level is selected: for exmaple, "time" is used as a noun and verb. The usage as a verb (B2) is more difficult than the usage as a noun (A1).

We used the item response theory (IRT) to obtain word difficulty measures that fit the vocabulary result dataset. Briefly, IRT (Baker, 2004) is a statistical method to automatically set the difficulty level of questions and the ability level of learners from test result data. There are two internal parameters to estimate: the learner's ability and question difficulty. The estimation of whether the learner answered a word correctly (+) or incorrectly (-) is determined by subtracting the question difficulty parameter from the sign of the learner's ability parameter.

To produce consistent results with a variety of data, IRT estimation is typically conducted using publicly available, established libraries. In this case, pyirt (https://github.com/17zuoye/pyirt), the IRT library in Python, was used because it is known to provide stable parameter estimation even when there is a certain amount of missing data. The 2PL IRT model was used for fitting. Further details on IRT models are provided in Baker (2004).

3 EXPERIMENTS AND CONCLUSIONS

Figure 1 displays the scatter plot of CEFR-J against the difficulty parameter by applying IRT to the vocabulary test result dataset (Ehara, 2018) in which ESL learners whose first language is Japanese responded to 100 word questions. Among the 100 words, 90 words were also present in CEFR-J. Along the horizontal axis of Figure 1, 0.0 denotes A1, 1.0 denotes A2, 2.0 denotes B1, and 3.0 denotes B2. We observed that the difficulty based on IRT varies as the difficulty of the CEFR-J increases. More significantly, Figure 1 clearly shows that words that language teachers find difficult (CEFR-J, horizontal) are not necessarily difficult for learners (IRT, vertical); however, words that language teachers find easy are also easy for learners. As CEFR-J levels are discrete, rank correlation was used instead of typical Pearson correlation coefficients. A Spearman's correlation coefficient ρ of 0.52 ($p < 0.05$) was obtained indicating that there was a statistically significant moderate correlation (https://geographyfieldwork.com/SpearmansRankCalculator.html).

We confirmed that the difficulty parameter estimated from the Ehara (2018) dataset shows a moderate correlation with the CEFR-J scale based on Spearman's ρ. This result suggests that the difficulty of words based on language teachers' linguistic intuition is roughly correlated with the difficulty of words for actual language learners.

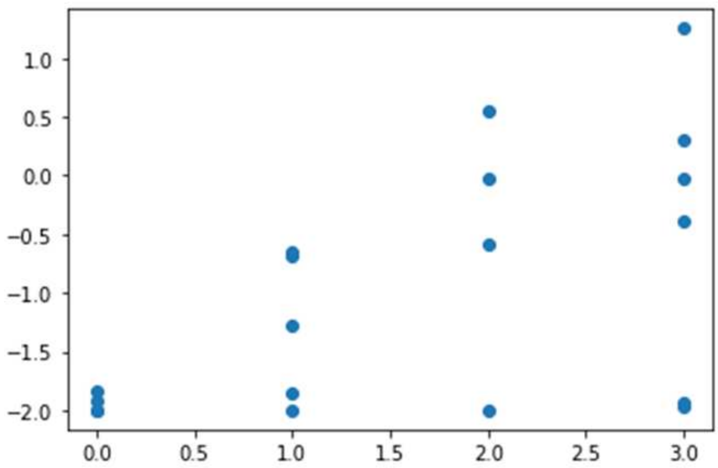


Figure 1: A scatter plot between the CEFR-J scale (horizontal) and the difficulty parameter estimated from the dataset by Ehara, 2018 using the item response theory (Baker, 2004).

REFERENCES

Baker, F.B.andKim, S.-H. (2004). ItemResponseTheory: Parameter Estimation Techniques. Marcel Dekker, New York.

Yo Ehara. Building an English Vocabulary Knowledge Dataset of Japanese English-as-a-Second-Language Learners Using Crowdsourcing. In Proc. of LREC, May 2018.

Nation, I. (2006). How large a vocabulary is needed for reading and listening?. Canadian modern language review, 63(1), 59-82.

Teamable Analytics: A Team Formation and Analytics Tool

Bowen Hui, Opey Adeyemi, Mathew De Vin,
Brianna Marshinew, Keyvan Khademi, Jeff Bulmer, Callum Takasaka
University of British Columbia, Okanagan campus
bowen.hui@ubc.ca,
{opeyadeyemi,mathewdevin03,briannamarshinew,keyvankhademi,bulmer.jeffk,ctakasaka}
@gmail.com

ABSTRACT: Forming effective teams for large classes is a challenge for educators due to the complexity of project needs, the diversity of individual characteristics, and the criteria different educators have for forming teams. In previous work, we proposed an artificial intelligence algorithm that achieves competitive benchmarking results while scaling the performance to handle large class sizes and a large number of constraints (Bulmer et al., 2020). In this demo, we present a web application that uses this algorithm to support the team formation process called Teamable Analytics. Teamable Analytics is compatible with any learning management system (LMS) that uses the LTI protocol. Our tool provides a dashboard for educators to elicit student characteristics and customize how those responses are combined to form teams. We integrated Teamable Analytics with the Canvas LMS and completed four pilot studies with classes that have between 40 to 200 students (Bulmer, 2021). Based on the feedback, we added visual analytics, team regeneration with peer evaluation feedback, and default characteristics about student diversity commonly used in the literature for forming teams. Currently, we are piloting Teamable Analytics in six interdisciplinary classes across multiple university campuses as part of an evaluation for enterprise adoption (Hui et al., 2021).

Keywords: Team formation, team analytics, large class sizes, multiple constraints, LTI integration

1 DEMO VIDEO

The video is available at https://youtu.be/qAevPcDVhkY .

REFERENCES

Bulmer, J., Fritter, M., Gao, Y., & Hui, B. (2020). FASTT: Team Formation Using Fair Division. The 33rd Canadian Conference on Artificial Intelligence (CAI). *Lecture Notes in Computer Science: Advances in Artificial Intelligence*, 12109, 92-104.

Bulmer, J. (2021). *Team Formation Using Fair Allocation.* Master's thesis. Department of Computer Science, Mathematics, Physics, and Statistics, The University of British Columbia, Canada.

Hui, B., Bulmer, J., Khademi, K., Takasaka, C., de Vin, M., Adeyemi, O., Marshinew, B., and Verma, K. (2021). *A Canvas-Integrated Team Formation and Analytics Tool.* Presented at Celebrate Learning Week, The University of British Columbia, Canada.

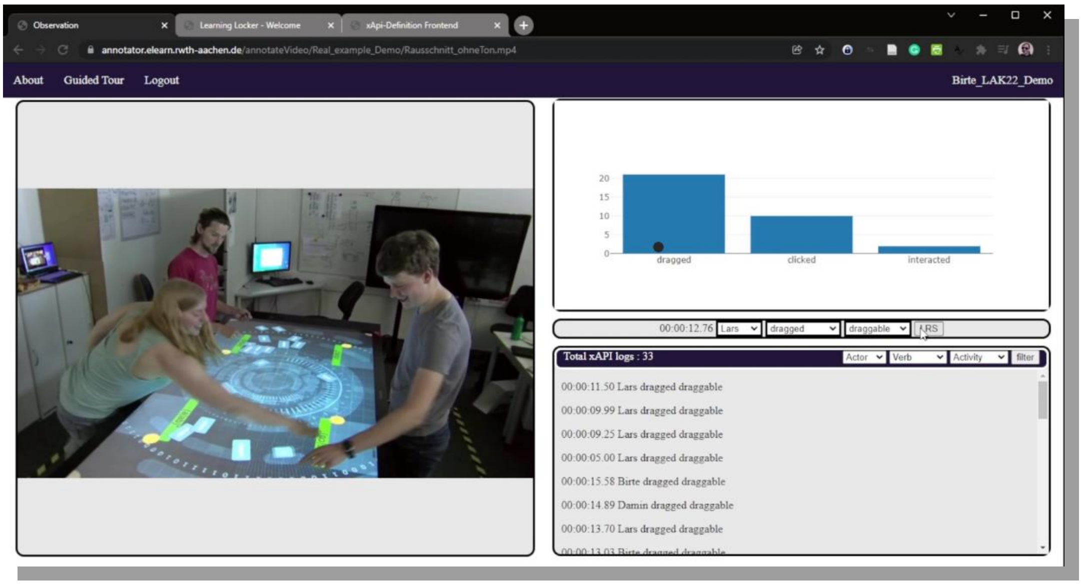
Combining Learning Analytics in remote and virtual lab-based learning with real lab experiences with xAPI video annotation

Authors: Birte Heinemann, Jens Doveren, Muhammad Usman, Ulrik Schroeder
RWTH Aachen University
Heinemann| Doveren|Schroeder@cs.rwth-aachen.de

ABSTRACT: Laboratories are interesting learning spaces where students acquire various competencies. Investigating this competence acquisition with learning analytics is a challenge. Virtualized and remote labs can provide us with data (more) easily, but can they foster the same competencies? We want to investigate this very question and have therefore developed a tool in which we could annotate recordings of real lab experiments with xAPI statements. This allows us to compare it with data collected in the two digital versions of the same laboratory, one remote and one in VR. During the development, we made sure to involve the future users and to meet their needs. This demo shows the first version of the annotating tool that we tested with different stakeholders to check if it is understandable and easy to use. The first results show that teachers quickly wish to facilitate annotation through technical possibilities. The result also shows that the tool helps bring together educators, developers and learning analytics experts and enables active participation of people without a technical background.

Keywords: Learning Analytics, Video Annotation, xAPI, experienceAPI

Video: https://youtu.be/M3ce9ktWX30



The authors gratefully acknowledge the funding by the Federal Ministry of Education and Research (BMBF) for the joint project Open Digital Lab 4 you (grant no. 16DHB2114)

Demo of UpGrade: An A/B testing platform to support learning analytics at scale

April Murphy
Carnegie Learning
amurphy@carnegielearning.com

Steve Ritter
Carnegie Learning
sritter@carnegielearning.com

Zack Lee
Carnegie Learning
zlee@carnegielearning.com

ABSTRACT: In this demo[1], we provide a walkthrough of a working instance of UpGrade, an open-source platform for running large-scale field experiments and A/B tests in educational software. UpGrade integrates with a client EdTech application and operates as a web-based service, providing an interface for researchers or product developers to set up and manage parameters such as condition assignment, start and end logic, and simple metrics monitoring for randomized controlled trials within the integrated application. We walk through experiment creation for a simple application called QuizApp, log in as example students, and show how UpGrade randomly assigns students one of two possible experimental conditions. We then review UpGrade's dashboard and user interface, which displays condition parameters as well as enrollment and metrics data, and review the steps required to use UpGrade with an EdTech app. We conclude the demo by discussing how UpGrade has been used to conduct randomized controlled trials in educational software serving tens of thousands of students, and close with an open invitation for partners and collaborators to help make UpGrade a valuable resource for the learning engineering community.

Keywords: A/B Testing, Educational Technology, Digital Experimentation

[1] Demo URL: https://drive.google.com/file/d/18kmGW9GQQPs09Thydtq16W_uWwsK1g01/view?usp=sharing

IntelliBoard Integrated Predictive Learning Analytics

Author: Elizabeth Dalton
IntelliBoard, Inc.
elizabeth@intelliboard.net

ABSTRACT: IntelliBoard Inc. has provided descriptive learning analytics based on Learning Management System data since 2005. Now IntelliBoard adds integrations with SIS, web conferencing systems, and other educational technology and offers integrated predictive learning analytics. In this interactive demo, we show how to build a predictive learning analytics model and train it using Neural Networks and Logistic Regression. A wide variety of validation statistics are provided. Predictions can be incorporated into reports, charts, and automatic notifications. Use of predictions and notifications can be tracked and correlated to improved learning outcomes. Default models estimate student risk based on common participation indicators, but both labels and features may be customized to any data included in the IntelliBoard data set, including data imported from local systems. IntelliBoard scales from thousands to tens of thousands of students, teachers, and courses.

Keywords: machine learning; learning management system; integration; implementation; scale

1 INTRODUCTION

Link to Demo: https://intelliboard.net/research or https://vimeo.com/671892014/f8f450e33c

In this demonstration, we'll show how to create, train, and generate predictions from a model based on the goal that learners will achieve a passing grade in a course within 90 days of the enrollment start date. This is one example of the many models that can be built with this system.

To connect to our system as a researcher, visit https://intelliboard.net/research. We offer special discounts for limited-access research accounts, as well as affordable pricing for systems that integrate with the tools your institution already uses to improve educational outcomes.

Assessing the new educational model Tec 21 based on dropout and student performance

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

ABSTRACT: The introduction of a new educational model at Tecnológico de Monterrey has posed the challenge of determining whether the new systems are better or at least as good as the previous one. This new model is based in competencies and has a flexible design where students choose a major after the three first semesters. This research proposes an analysis using student data from the new and previous educational models of Tecnológico de Monterrey to create a comparison system between different models. The project aims to use Educational Data Mining, Learning Analytics and Process Mining techniques to identify features that could predict student performance and dropout on earlier stages and determine whether a model performs better than the other on those same points. A prediction model with better precision and a longer prediction window would indicate that the corresponding educational model has better identified and applied the core necessities for student learning. This would also allow for earlier interventions for students that are underperforming.

Keywords: Educational Data Mining, Learning Analytics, Process Mining

1 BACKGROUND

Currently, Tecnologico de Monterrey is implementing the first generation of their new teaching model, called “Tec21”, which is certified by the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC), the Mexican Federation of Private Institutions of Higher Education (FIMPES), and the Accreditation Board for Engineering and Technology, Inc. (ABET), among various others. This model is challenge-based and combines aspects of competence, problem, and case-based methodologies. The main objectives of this teaching model are for the student to acquire the area fundamentals related to their chosen path, as well as developing the competencies and acquiring the skills they will need in their future. In this model, challenges are the central learning unit, with everything else focusing on delivering the needed tools for solving them.

This change has not come without its challenges: aside from the extremely long time that is needed before results can even be seen, there is also the problem of determining if the change was beneficial or not to the students. A large enough change usually makes direct comparison impossible.

Grouping the courses into disciplines might solve this issue. Any large enough University is certified by external organizations to grant degrees on specific disciplines. This accreditation ensures that the programs, even if modified due to a change in model or an update in information or techniques will still cover the same overall themes, and the final objective remains the same. This can be seen on the graduate profiles of the different majors offered by universities. Those same certifications ensure the necessary disciplines are covered before granting a degree.

One advantage that we have is that, by nature, universities record a very large amount if student data each semester, leaving only grouping to be done if we want to work with the disciplines. While most of these attributes are regularly used in data mining, the time aspect of them is often discarded. Since these datapoints have a defined order, we could use an approach based on processes like process mining to obtain additional information. Process mining "refers to the act of discovering, verifying, and improving processes using real world data" (Van der Aalst, 2016). According to the author, process mining assumes that it is possible to extract meaningful data from event logs that are present in almost all databases. While similar in some regards to data mining (the use of large databases, the automatic nature of discoveries, etc.), process mining differs from data mining in some important aspects: it requires a specific type of data that is not always present: a case ID, a timestamp, and the name of the activity. Educational databases have an abundance of these data points, as any respectable organization will have detailed information about which students (case ID) took which classes (events), and on what semester (timestamp).

Finally, it is important to understand the basis of the new model. A competence-based model has as its objective the "mastery of knowledge, skills and abilities that demonstrate learning" (Simonds et al., 2017). While competence-based models are relatively new to higher education, they have been in use for close to 30 years in programs like secondary education (Sullivan & Downey, 2015) and in cases where skills and knowledge should not be without each other, as is in medical care (HARDEN, 1999).

The educational model proposed by Tecnológico de Monterrey (TEC21) is a challenge-based model with 3 distinct phases in its curriculum: discovery, focus, and specialization. New students have the option of enrolling in an undergraduate degree directly or spending 3 o 4 semesters on a more general path (engineering, social science, bio-sciences, etc.). The students that do not choose an undergrad directly have to take a series of elective courses for them to familiarize themselves with their different options and make an informed decision about their future. These electives are perfect data points for process mining that could help better identify the "ideal path" through these electives, which in turn could generate a model and help both students and staff with a series of decisions.

2 GOALS AND RESEARCH QUESTIONS

We believe that machine learning based models for the prediction and classification of student academic success can have applications that go deeper than just making predictions and providing group data for institutional decisions. We believe they can provide a framework for a valuable comparison between different educational models. By developing algorithms for the educational models to be compared using similarly enough features, we can use the precision, recall, and prediction window of the models to determine if one is better than the other in terms of student academic success and/or dropout. A prediction model with better precision and a longer prediction window would indicate that the corresponding educational model has better identified and applied the core necessities for student learning, and as such, could be called "better" in this sense. This would also allow for earlier interventions for students that are underperforming.

Objectives:

The goal of this project is the development of machine learning models for the overall academic success of students (including dropout as a possibility) for the current and previous models to be compared. For the comparison to be valid, we will be using the disciplines of the different majors as attributes in the models. Disciplines are the core themes of a course that do not change, even if the overall teaching strategy does. As part of our research, we will be looking into which one of the various classification or prediction approaches (Neural Networks, Naïve Bayes, Logistic Regression, SVM, to name a few) are best suited to our needs. We are also aiming to develop our models under an explainable AI principle.

Research questions:

Do student's that go through a challenge-based model perform better, worse, or the same as their classic-model counterparts in terms of academic performance (grades) and dropout? What features improved or worsened those results?

Do the specific electives that student take prior to choosing their specialization affect their performance in terms of academic performance (grades) and dropout? Should some of these electives be pre-requisites for other courses? (This only applies to TEC21)

3 PRESENT STATE AND CURRENT SOLUTIONS

To better understand the present state of data mining in education, a systematic review on the topics of Educational Data Mining (EDM), Learning Analytics (LA), and Process mining (PM) was performed. This review looked at a total of 104 unique papers from the SCOPUS database using a set of queries combining the following keywords: "Learning Analytics", "Educational Data Mining", "Process Mining", and "Student performance", "Student Success", and "Dropout". All papers found that regarded education (i.e. teaching) were included, while duplicate files and articles not regarding education were excluded.

We extracted relevant information regarding several important features of the papers like the overall objective, input variables, the techniques used, deployment of the models used, etc. The results of the review showed that the main objectives presently consist of grade prediction (either in ranges by a classifier or points by a predictor) as the most popular one; student behaviours as a runner-up, which we believe is due to the recent surge in both online course popularity and log-data availability; and dropouts in a third place.

In terms of student grades, most studies attempt to predict if a student will fail, is at risk of failing, or pass; a few more even go as far as to try and predict if a particular student will do exceptionally well on a course (Riestra-González et al., 2021; Bravo-Agapito et al., 2021; Zeineddine et al., 2021; Tomasevic et al., 2020; Helal et al., 2018). The most common techniques used were Decision trees, Naïve Bayes, Logistic regression, Artificial Neural Networks, among others. Such techniques have produced results along the 70% to 85% accuracy for identifying failing students (Riestra-González et al., 2021; Zeineddine et al., 2021). Student behaviour articles instead focus on classifying sets of students into either different levels of productive groups as seen in paper by Cerezo and Sunar, to name a few (Cerezo et al., 2016; Sunar et al., 2020), or identify different types of self-regulated learners (Li et al., 2020; Sun & Xie, 2020; Williams et al., 2018). Finally, regarding the third most common objective, student dropout, researchers were attempting to identify what causes students to leave their studies. However, there were fewer studies regarding classic education (i.e. university based courses) than online based ones. Articles by Gregori, Maldonado, and Olaya deal with MOOC dropout, as for them this is a much more key problem (Gregori et al., 2018; Maldonado et al., 2021; Olaya et al., 2020; Zhuhadar et al., 2019).

4 SUGGESTED SOLUTION

In order to tackle the problems of model change and the long waiting period before results can be truly analyzed for universities, we are proposing the development of framework for a comparison system between educational models that takes into account the models statistics as decision variables (precision, recall, f1, prediction window). The inputs will include the large amount of log trace data available from universities (historical records of the courses taken, grades in chronological order, trace information from digitalized courses) along with more common inputs like socio-economic factors. The data inputs to be used will be pre-processed to ensure they allow a meaningful comparison, both by developing the same features for the models so that the models have the same starting point, and to verify the student populations are similar enough.

Along with the overall comparison of systems using data models, the specific features that show an increase or decrease on their effect on education can be identified and used to further improve the current educational model. Ideal "paths" through courses could be identified here and used for curriculum development, while discipline effect changes could then be attributed to the change in model, allowing for a measurement of the impact of the methodology.

5 RESEARCH METHODOLOGY

In order to test our hypotheses, we will collect academic data like grade averages, core and optative course grades, participation on extra-curricular activities, socio-demographic information (age, sex, previous school, parent scholarity, etc.), and historical information of student's trajectory through the university. We will be collecting data from current students of the TEC21 model, as well as from students who studied under the previous methodology. Tecnológico de Monterrey will be providing this data from their servers via their Data Hub and Living Lab departments, specifically the IFE Data Hub. A series of machine learning models including but not limited to classification, process mining, clustering, and regression will be developed for use in the project. Statistical techniques will be applied to ensure the comparisons and findings are valid.

Some ethical considerations in this project have to do with the data we would require. As the data contains information that could be considered sensitive, we need to be careful while handling it in order to keep both it and the student's that it belongs to secure. In order to keep the anonymity of the student's, the data will be given to us without identifying characteristics, and only a randomized identification tag for data managing purposes.

6 WORK SO FAR

At this point in time, I have finished a first draft of a systematic review regarding the present state of data mining in Education (i.e. Educational Data Mining (EDM) and Learning Analytics (LA)) totaling 104 unique articles from the SCOPUS database, with a focus on the objective variable, input types, focus (EDM, LA, or process mining), data mining technique used, and school of thought (Paradigm under which each article was written). This review spanned 11 years, from 2010 to the present year (2021). The data from the university mentioned before has already been applied for, and we are expecting to have full access to it at the latest in December 2021. The project presented here has already passed satisfactorily the proposal defense and comments from both it and the LAK doctoral consortium are being integrated for further improvement.

REFERENCES

Arnold, K., & Pistilli, M. (2012). *Course signals at Purdue*. 267–270. https://doi.org/10.1145/2330601.2330666

Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, *115*. https://doi.org/10.1016/j.chb.2020.106595

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers and Education*, *96*, 42–54. https://doi.org/10.1016/j.compedu.2016.02.006

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Gregori, E. B., Zhang, J., Galván-Fernández, C., & Fernández-Navarro, F. de A. (2018). Learner support in MOOCs: Identifying variables linked to completion. *Computers & Education*, *122*, 153–168. https://doi.org/10.1016/J.COMPEDU.2018.03.014

HARDEN, R. M. (1999). AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education. *Medical Teacher*, *21*(1), 7–14. https://doi.org/10.1080/01421599979969

Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134–146. https://doi.org/10.1016/j.knosys.2018.07.042

Li, S., Du, H., Xing, W., Zheng, J., Chen, G., & Xie, C. (2020). Examining temporal dynamics of self-regulated learning behaviors in STEM learning: A network approach. *Computers & Education*, *158*, 103987. https://doi.org/10.1016/J.COMPEDU.2020.103987

Maldonado, S., Miranda, J., Olaya, D., Vásquez, J., & Verbeke, W. (2021). Redefining profit metrics for boosting student retention in higher education. *Decision Support Systems*, *143*, 113493. https://doi.org/10.1016/J.DSS.2021.113493

Olaya, D., Vásquez, J., Maldonado, S., Miranda, J., & Verbeke, W. (2020). Uplift Modeling for preventing student dropout in higher education. *Decision Support Systems*, *134*, 113320.

Outlook, C. (2020). *Learn more, earn more: Education leads to higher wages, lower unemployment* (Vol. 2021, Issue January 26,). https://www.bls.gov/careeroutlook/2020/data-on-display/education-pays.htm

Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers and Education*, *163*. https://doi.org/10.1016/j.compedu.2020.104108

Rivera, F. Y. M. (2013). *La relación entre ingreso y educación* (Vol. 2021, Issue January 26,). https://www.eleconomista.com.mx/opinion/La-relacion-entre-ingreso-y-educacion-20130501-

0004.html

Simonds, J., Behrens, E., & Holzbauer, J. (2017). Competency-Based Education in a Traditional Higher Education Setting: A Case Study of an Introduction to Psychology Course. In *International Journal of Teaching* (Vol. 29, Issue 2, p. 412).

Sullivan, S. C., & Downey, J. A. (2015). Shifting Educational Paradigms: From Traditional to Competency-Based Education for Diverse Learners. *American Secondary Education*, *43*(3), 4–19. https://www.jstor.org/stable/43694215

Sun, Z., & Xie, K. (2020). How do students prepare in the pre-class setting of a flipped undergraduate math course? A latent profile analysis of learning behavior and the impact of achievement goals. *The Internet and Higher Education*, *46*, 100731. https://doi.org/10.1016/J.IHEDUC.2020.100731

Sunar, A. S., Abbasi, R. A., Davis, H. C., White, S., & Aljohani, N. R. (2020). Modelling MOOC learners' social behaviours. *Computers in Human Behavior*, *107*, 105835. https://doi.org/10.1016/j.chb.2018.12.013

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

Van der Aalst, W. (2016). Process mining: Data science in action. In *Process Mining: Data Science in Action*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

Williams, K. M., Stafford, R. E., Corliss, S. B., & Reilly, E. D. (2018). Examining student characteristics, goals, and engagement in Massive Open Online Courses. *Computers & Education*, *126*, 433–442. https://doi.org/10.1016/J.COMPEDU.2018.08.014

Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, *89*. https://doi.org/10.1016/j.compeleceng.2020.106903

Zhuhadar, L., Daday, J., Marklin, S., Kessler, B., & Helbig, T. (2019). Using survival analysis to discovering pathways to success in mathematics. *Computers in Human Behavior*, *92*, 487–495. https://doi.org/10.1016/J.CHB.2017.12.016

Supplementary page for Doctorate Consortium:

I would really appreciate the consortium's aid on the following issues:

- Multi-department research: What are some recommendations for working with student populations from different areas (i.e. Engineering vs Social Sciences, for example)
- I'm aiming to use machine learning model statistics (precision, recall, f1, prediction window), to compare the **educational** models, what previous steps would you recommend for the comparison to be a completely valid one?
- Any feedback regarding the use of mining techniques and best practices would be greatly appreciated.

As for the researchers I would appreciate feedback from, here is a small list:

- Professor Shirley Alexander
- Professor Samuel Greiff
- Professor Shane Dawson
- Sebastien Lalle
- Özge Nilay Yalçın
- Cristina Conati

Evaluating Young Children's Creation Process for Creative Coding Projects with Learning Analytics

Apittha Unahalekhaka
Tufts University
apittha.u@tufts.edu

ABSTRACT: Millions of animated stories and games have been created on coding applications, such as the free ScratchJr app, by children ages 5-7 as open-ended projects. How can we assess children's learning process from these projects? Learning analytics (LA) has the potential to evaluate children's learning process when they engage in hands-on and unscripted coding activities. The first part of the proposed study focuses on children's patterns trajectories when creating open-ended coding projects with ScratchJr (n = 120 students) in the Coding as Another Language curriculum. The second part focuses on teacher lesson modifications, whether their actions moderate the relationship between students' creation patterns and learning outcomes (n = 28 teachers, 500 students). The result of the proposed study may lead to a better understanding of how LA can assess children's creative coding in classrooms with the integration of teacher instructions.

Keywords: Early Childhood Education, Programming, Project-based Learning, Learning Analytics, Assessment, Constructionism

1 Background

Computing skills are important in the 21st century and have been shown to help develop young children's ways of thinking, including their computational thinking, problem-solving, and critical thinking (Angeli & Valanides, 2020). The creative creation of opened-end coding projects can further promote to acquire computational skills in young children (Bers, 2020; Clements, 1998). Furthermore, the constructionist approach believes that powerful learning can occur when children use technology to create meaningful products that they can share with the people around them (Bers, 2020; Papert, 1980).

To better foster student learning and motivation, not limited to constructionist learning, it is valuable for educators to prioritize understanding students' learning process (Ryan & Deci, 2020, p. 6). While some types of computational assessment, such as task-based, game-based, project-based, and unplugged (Relkin et al., 2020), help to understand children's end learning outcomes, none of these assessments can clearly capture their learning process (Berland et al., 2014). Learning analytics (LA) can fill in this gap to capture children's real-time interactions with the programming platforms (Admiraal et al., 2020). Studies have found that student's learning analytics such as coding block usage patterns can reveal their coding process, mastery levels, and styles (Emerson et al., 2020; Grover et al., 2017).

Usage of LA is growing fast in computer science education discipline. However, current studies on its effectiveness used differing metrics to capture student learning, resulting in mixed outcomes on how LA can effectively capture students' computational competencies. There are even more gaps in the understanding of how LA can assess learning in early grade levels due to insufficient research conducted in this age range. In terms of variables, multiple studies did not find *frequency* of usage as a significant predictor of learning; rather, specific student *learning strategies* were suggested as more useful learning indicators (Gašević et al., 2015). Aligning with Turkle and Papert's (1990, p. 129) ideas, knowledge can be seen as styles, and each style is "equally valid on its own terms". Additionally, external factors such as the instructional condition were shown to moderate student academic performances when interacting with the learning system (Gašević et al., 2016). Consequently, there needs to be more investigation on what usage variables from LA can evaluate young children's constructive learning process, and whether these processes relate to teacher instructions and student learning outcomes.

This study is part of the Coding As Another Language (CAL) project directed by Prof. Marina Bers at the DevTech research group at Tufts University. CAL's overarching goal is to implement ScratchJr into K-2 classrooms across multiple school districts in the US, using a literacy-integrated programming curriculum. ScratchJr is a free downloadable tablet app designed for children ages 5-7 to create animated stories and game projects (Bers & Resnick, 2015) (Refer to Appendix A). The highlight of the work is the application of LA to study young children's learning process when *creating coding projects* on ScratchJr. The learning process in this study will be captured through trajectories of children's behavior when creating projects. This study also examines how teacher instruction modification behaviors might influence children's learning trajectories. The following sections of this proposal will elaborate on two parts of the study, in which a framework to summarize the entire research is shown in Figure 1.

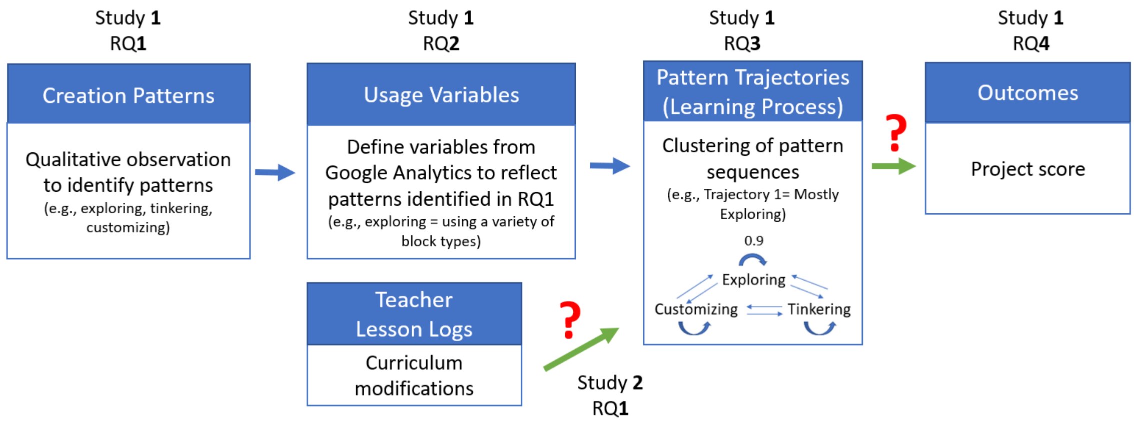


Figure 1: A diagram on different components according to the research questions

2 Goals and Research Questions

2.1 Part One: Children Creation Patterns (*n* = 120 students)

The first part of this study uses learning analytics to investigate how children's learning processes (trajectories of creation pattern) are related to their learning outcomes (ScratchJr project scores). Each child will create one final project over the last three lessons of the curriculum, which is when I will identify their *creation patterns* and assess their projects. This study defines *creation patterns* as the coding and project customizing behaviors identified by observation and learning analytics. Some creation patterns may include exploring, tinkering, and customizing, which are inspired by the EXTIRE framework (Berland et al., 2013).

In the first part of the proposed study, I will utilize qualitative and quantitative research methodologies. Specifically, I will qualitatively observe creation patterns from a subset of children's screen recordings (*n*= 40) and compare these findings to the quantitative ScratchJr analytics data.

- **RQ1:** What are the creation patterns shown on the screen recordings when children create projects?

After identifying creation patterns in RQ1, I will define *usage variables*, features calculated from the raw analytics usage data (e.g., the ratio of block complexity and the average gap time between pulling coding blocks) to capture the observable creation patterns. For example, tinkering patterns might be identifiable when children frequently drag identical coding blocks out from the coding palette at a certain speed. Furthermore, reliability testing is needed to check that the observation-defined and analytics-defined creation patterns are accurately aligned.

- **RQ2:** Can the quantitative analytics usage variables identify the creation patterns from RQ1 reliably?

After the process in RQ2, I will use analytics data to identify creation patterns of 120 participants across the last three lessons. These patterns will be turned into trajectories and compared to the final project scores. *Pattern trajectory* is a crucial factor to reach the goal of this study. It reflects the learning process or how children tend to stick to or change their creation behaviors over different stages of their final coding projects.

- **RQ3:** What are creation pattern sequences and trajectories?
- **RQ4:** How do the project scores vary across pattern trajectories across the three final lessons?

2.2 Part Two: Teacher Lesson Modifications (*n* classroom = 28, *n* student = 500)

The second part of this study aims to understand whether teaching actions affect student learning processes, which may lead to their learning outcomes. Specifically, I am investigating whether teacher modifications of curriculum activities moderate the relationship between the students' learning processes (pattern trajectories) and their learning outcomes (ScratchJr project scores). Using the steps from part one (RQ2 and RQ3) of this study, I will use learning analytics to identify the creation pattern trajectories of all 500 students across 24 lessons.

- **RQ1:** Can teacher lesson modifications moderate the relationship between students' creation pattern trajectory and the average project score at three different time points in the curriculum (beginning, middle, and end)?

3 Novelty of Suggested Solutions

While many traditional assessment approaches focus on assessing students' final learning outcomes, the proposed study will use LA to explore children's learning process (trajectory of creation patterns). Specifically, I will analyze how young children engage in open-ended coding project creation across multiple weeks. This approach in using LA to assess learning process in unscripted activities has been studied with high school and college students but not at the early childhood level (Berland et al., 2014; Blikstein, 2011). With a different coding platform and student developmental level, it is highly possible that young children will have a different coding pattern compared to adults. To check the accuracy of the creation patterns extracted from LA, this study will compare these patterns to qualitative observation of the sessions (screen recording of the moving blocks).

Furthermore, my previous study with ScratchJr's Google Analytics data showed that young children at home seemed to have a more exploratory coding style. This was demonstrated by the greater variety in block complexity levels exhibited by children when creating projects at home compared to children doing similar activities at school (Unahalekhaka & Bers, 2021). However, this earlier study only looked at the block usage frequency and did not connect it to the learning outcomes. For example, it did not address questions related to weather exploratory coding style was related to quality of children's coding projects or their coding competency. Therefore, in the proposed study, I also plan to investigate how children's coding block usage patterns relate to their learning outcomes through ScratchJr project scores.

For the results of the proposed study to be applicable for classrooms, the second part examines how teacher instructions are related to the children's learning process and outcomes. Particularly, the study focuses on whether, when, and why teachers modify their lesson activities from the provided CAL curriculum. It has long been proven that instruction adaptations according to students' needs can positively impact their learning (Parsons et al., 2018). The proposed study explores whether this assumption can be applied to the process of open-ended early childhood coding.

4 Methods

4.1 Participants & Procedure

Participants for this study consist of teachers and students from 28 K-2 classrooms from six schools in Rhode Island, USA, participating in the CAL project. There are two parts to this study. The anticipated number of participants for part one is 120 students from 12 classrooms and 500 students from 28 classrooms for part two. This study plans to collect ScratchJr's analytics from all students across their 24 lessons using the Google Analytics platform. The study will last from November 2021 to April 2022. Participants will participate in a 24 CAL lessons intervention, happening twice a week. Teachers will teach according to the CAL curriculum and ask students to create open-ended ScratchJr projects at three time points across the intervention: beginning (Time 1), middle (Time 2), and end (Time 3). Researchers will collect and assess these projects to measure students' learning outcomes. The summary of the data collection is shown in Appendix B.

While children will create multiple ScratchJr projects, part one of the study will only analyze the one final project created by each child across three consecutive lessons at Time 3 (the end). I will also collect screen recordings of when students do their final projects on the first lesson from three at Time 3 (n = 120 students). A subset of these screen recordings will be used to qualitatively identify creation patterns, then be compared to the patterns found from the analytics data.

Part two of the study focuses on teacher lesson modification across the curriculum's beginning, middle, and ending. Unlike the first part of this study, ScratchJr projects collected at Time 1 to Time 3 will all be analyzed. Teachers will be asked to fill in a lesson log or a survey on how the lesson went after each session. The primary lesson log questions are consistent across lessons; however, the detail within some questions will slightly differ depending on the activities from that lesson. Specifically, the questions will investigate whether teachers modify each activity and whether the modification happens during the lesson preparation, during teaching, or both. The next question explores the reasons behind the modifications, whether they were due to time constraints, student "learning or understanding," "motivation or interest," "behavior (e.g., ability to focus)," or other factors (Parsons et al., 2018; Vandewaetere et al., 2011).

4.2 Analysis Plan

Part one of this study explores children's creation patterns (e.g., tinkering, exploring, customizing, etc.) when creating ScratchJr projects. For RQ1, I will qualitatively observe children's screen recordings (n = 40) at every two minutes to determine a creation pattern at each timeframe. For RQ2, I will test whether usage variables can accurately capture creation patterns at every two-minute timeframe. In other words, usage variables summarize what children did at every two minutes, such as a child used 15 coding blocks with a ratio of 2:5 of advanced to beginner blocks. For RQ3, I will derive creation patterns of 120 participants from analytics data then use Markov Models clustering to find pattern trajectories in each of the three lessons at Time 3 (Saqr & López-Pernas, 2021). For example, in the first lesson, a cluster of students may customize their characters throughout the session with minimal coding. In RQ4, I will conduct longitudinal regression to investigate whether a particular trajectory at three different lessons at Time 3 has a higher average final ScratchJr project score.

Part two of this study tries to understand when and why teachers modify their lessons and how they relate to the children's creation trajectories and learning outcomes. There are 28 classrooms, 24 lessons, and five to six activities per lesson. For each lesson, I will turn teacher modification actions into three continuous and three categorical *teacher variables* as follows:

- Classroom ID
- The ratio of activities unmodified
- The ratio of activities modified during preparation
- The ratio of activities modified during the lesson
- Activities modified due to children's understanding (Yes/No)
- Activities modified due to children's interests (Yes/No)
- Activities modified due to children's behavior (Yes/No)

For RQ1 in part two, I will group student and teacher data from each lesson into three phases (Phase 1: Lesson 1-10, Phase 2: Lesson 11-18, and Phase 3: Lesson 19-24) as students will create ScratchJr projects at three-time points. Using longitudinal logistic regression, the first model will predict the class average ScratchJr project score across three different time points. The predictors of this model will include classroom ID, phase number, teacher variables, and the ratio of student being in each creation trajectory.

4.3 Ethical Considerations

This study has the Institutional Review Board (IRB) of Tufts University approval (protocol #1810044). The data will be safely secured in a password-protected server that will only be accessible by the researchers that are on IRB. This study will obtain educator and parental consent for testing and recording their students or children before the data collection phase. The ScratchJr analytics data is de-identified at an individual level, and only the classroom level identification is known. Lastly, the general trends of the data will be reported, but not individual-level data that can be identified as a particular child. Moreover, general children's identities are further de-identified using randomized IDs instead of full names.

5 Current status of the work

Data collection will last from November 2021 to April 2022. I plan to clean and analyze the data from March 2022 to June 2022. Furthermore, I already conducted an online pilot study with six children. Each child spent two to three 45-minute sessions with me to create ScratchJr projects. This pilot study aims to compare three sources of data: 1) children's reaction from videotaping, 2) usage patterns from screen recordings, and 3) google analytics. The findings from this pilot study will be helpful to determine possible creation processes that can be derived from the analytics.

REFERENCE

Admiraal, W., Vermeulen, J., & Bulterman-Bos, J. (2020). Teaching with learning analytics:how to connect computer-based assessment data with classroom instruction? *Technology, Pedagogy and Education*, *29*(5), 577–591. https://doi.org/10.1080/1475939X.2020.1825992

Angeli, C., & Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Computers in Human Behavior*, *105*, 105954. https://doi.org/10.1016/j.chb.2019.03.018

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. *Technology, Knowledge and Learning*, *19*(1–2), 205–220. https://doi.org/10.1007/s10758-014-9223-7

Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using Learning Analytics to Understand the Learning Pathways of Novice Programmers. *Journal of the Learning Sciences*, *22*(4), 564–599. https://doi.org/10.1080/10508406.2013.836655

Bers, M. U. (2020). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom*. Routledge.

Bers, M. U., & Resnick, M. (2015). *The Official ScratchJr Book: Help Your Kids Learn to Code* (1st ed.). No Starch Press.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, 110. https://doi.org/10.1145/2090116.2090132

Clements, D. (1998). *Young Children and Technology* (p. 23).

Emerson, A., Smith, A., Rodriguez, F. J., Wiebe, E. N., Mott, B. W., Boyer, K. E., & Lester, J. C. (2020). Cluster-Based Analysis of Novice Coding Misconceptions in Block-Based Programming. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 825–831. https://doi.org/10.1145/3328778.3366924

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

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Grover, S., Basu, S., Bienkowski, M., Eagle, M., Diana, N., & Stamper, J. (2017). A Framework for Using Hypothesis-Driven Approaches to Support Data-Driven Learning Analytics in Measuring Computational Thinking in Block-Based Programming Environments. *ACM Transactions on Computing Education*, *17*(3), 1–25. https://doi.org/10.1145/3105910

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M., & Allen, M. (2018). Teachers' Instructional Adaptations: A Research Synthesis. *Review of Educational Research*, *88*(2), 205–242. https://doi.org/10.3102/0034654317743198

Relkin, E., de Ruiter, L., & Bers, M. U. (2020). TechCheck: Development and Validation of an Unplugged Assessment of Computational Thinking in Early Childhood Education. *Journal of Science Education and Technology*, *29*(4), 482–498. https://doi.org/10.1007/s10956-020-09831-x

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

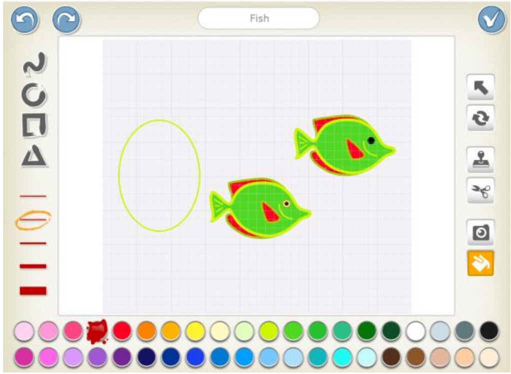
Saqr, M., & López-Pernas, S. (2021). The longitudinal trajectories of online engagement over a full program. *Computers & Education*, *175*, 104325. https://doi.org/10.1016/j.compedu.2021.104325

Turkle, S., & Papert, S. (1990). Epistemological Pluralism: Styles and Voices within the Computer Culture. *Signs*, *16*(1), 128–157.

Unahalekhaka, A., & Bers, M. U. (2021). Taking coding home: Analysis of ScratchJr usage in home and school settings. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-021-10011-w

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

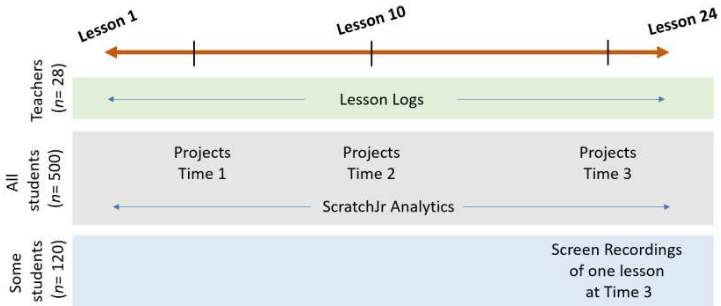
Appendix A



Left: An example of a ScratchJr project, which is showing a parallel coding sequence.

Right: Painting tools on ScratchJr allow children to customize their characters and backgrounds.

Appendix B



Data Collection Plan

Appendix C

	Usage Variables
Coding Blocks	Total blocks
	Ratio of unique blocks
	Ratio of advanced and intermediate blocks over beginner blocks
	Ratio of repeating occurrences (more than 1 identical block next to each other)
Duration	Duration of Coding
	Duration of Painting Tools usage
	Gap time between each block

Potential usage variables from ScratchJr analytics. As I am using Google Analytics to collect usage data from the ScratchJr app, the data that I will have will be quite general, showing event names by timestamps. Some of these events consist of 28 types of coding blocks that will be captured when children drag each block out of the palette and drop on the coding area. Furthermore, the timestamps will also show when children open or close the painting tool, as well as when they add a new page to the project (max. 4). The weakness of this dataset is that there is no event that tells when the child clicks the "play button" (green flag) to execute the syntax, or which page they are working on.

Developing a learning analytics model to explore motivation: A case study in a UK computer science department

Hafsa Omer AL Ansari
University of Huddersfield
Hafsa.alansari@hud.ac.uk

ABSTRACT: This study explores enhancing student learning and performance by investigating student motivation within a learning analytics context. This research is an exploratory case study within the Computer Science Department at the University of Huddersfield, UK. The research aims to study student motivation through Self-Determination Theory and train the identified factors using learning analytics records. More than 10,000 thousand virtual learning environment records were analyzed from year 2019 and 2020, with the study applying a mixed method approach using survey, interview and data records. The survey result has been analyzed using descriptive analysis. The semi-structured interview findings are currently under investigation, and they will be analyzed using thematic analysis. The data records will be analyzed using data mining techniques to assess and map the identified themes with the hidden patterns within the learning analytics data. The final framework will be tested using structure equation modelling on large number of students from CS department.

Keywords: Learning Analytics, Student motivation, Self-determination theory, Computer science

1 INTRODUCTION AND BACKGROUND

Motivation has been considered as a significant factor in a student's academic journey. As discussed by Schweinle, Meyer, and Turner (2006) student motivation and emotions are crucial for their attitudes, behaviors and achievement. In general, many learning theories such as Self-Regulated Learning (SRL) emphasis on motivation for achievement and learning ((Boekaerts, 1999);(Pintrich, 2000);(Zimmerman, 2002)). In fact, nowadays, the availability of big data in the educational domain, along with their analysis of learning analytics, is leading to changes in learning and the learning environment (Siemens & Long, 2011). In learning analytics, information about students is used for assessing, eliciting, and analyzing to optimize a student's learning experience and to help in decision making. Corrin and da Barba (2014), for example, examined how learning analytics can impact student's motivation in a feedback context. Their result shows that students' motivation is affected by their progress awareness and their regulation effort. However, some participants indicated that it had no effect on their motivation. Also, Miltiadou and Savenye (2003) reported that motivation has not received attention in online learning and their study, and they have recommended that in order to increase student success rate and engagement, more studies are required to test motivation theories regarding their effect on student learning. As a result, more research is needed to determine the capabilities of learning analytics in enhancing learning processes, particularly in exploring student learning motivation. Accordingly, the aim of this research is to address these gaps by investigating student motivational factors in learning analytics context.

2 MOTIVATION AND PROBLEM

Student motivation has not yet been investigated for analyses in learning analytics contexts (Schumacher & Ifenthaler, 2018). In learning, motivation plays a significant role, but most current learning analytics research focuses on data privacy, data processing and user systems development. Connecting learning analytics with learning theory and students' motivation is still in its infancy. Lonn, Aguilar, and Teasley (2015) argue that current learning analytics tools do not include investigation of student motivation.

3 RESEARCH QUESTIONS

Research questions guiding this study are:
R1: What are the intrinsic motivational factors that can affect, from a Computer Science student's perspective, their learning within a learning analytics context?
R2: What are the extrinsic motivational factors that can affect, from Computer Science student's perspective, their learning within a learning analytics context?
R3: What are the autonomy-related factors that can affect Computer Science motivation in learning environment within a learning analytics context?
R4: What are the competence-related factors that can affect Computer Science motivation in learning environment within a learning analytics context?
R5: What are the relatedness-related factors that can affect Computer Science motivation in learning environment within a learning analytics context?

4 AIM

This study aims to extract the motivational factors that play a role in shaping students' academic journey within the Computer Science (CS) department. We believe such a study is essential to better understand the different factors that affect CS student motivation. These factors can contribute to providing a better learning experience for students' and teachers as well by informing the future design of learning analytics (LA) tools. Therefore, the new LA tools can include cognitive and noncognitive factors that can be used to enhance student motivation and engagement. Also, institutes would be able to provide learning recommendations based on the students' learning behavior.

5 METHODOLOGY

The study conducts an exploratory case study at University of Huddersfield, UK in the Computer Science Department. It focuses on investigating students' motivation and examines the factors that play a role in them taking their degree. Since the research aims to explore "what questions", an exploratory case study is appropriate (Saunders, Lewis, & Thornhill, 2016). Exploratory case studies are used when there is no predetermined outcome. The exploratory case study investigates what is happening with the analytics in the participating courses and discovers hidden patterns of behavior from the student data transaction. The case study answers the research questions aiming to find out the motivation factors through understanding the relationship between the collected data and the student LA data records. The case study collects primary data that consists of LA transactions records from the University of Huddersfield system during the academic year 2020, as well as surveying and interviewing students. The primary data includes quantitative data collected through (survey and LA transaction records) and qualitative data collected from interviews. Also, the case study collects secondary data, including thousands of LA data records collected from the 2019. The secondary data focuses on the analytics and what patterns convey about students' engagements with different learning tasks.

5.1 MIXED METHOD

A mixed method is used which uses more than one phase of data collection and analysis. In this context, a triangulation of data (Flick, Kardoff, & Steinke, 2004) is used to combine a collection of data from different sources such as interviews, surveys, questionnaires etc. Using mixed methods allows researchers to investigate different perspectives and uncover the relationship between different variables. Mixed methods also enable researchers to obtain a comprehensive view of the study, and view it from different angles (Creswell, 2014). This allows the researcher to validate findings by ensuring that the obtained result is not generated from simply one source and that the final result is therefore supported by other sources of data. Quantitative data includes statistical data about the student activities in the course and student Likert-based survey. The quantitative data provides facts about the current learning behavior and further analysis exposes the relationship between the collected data and the analytics findings. Qualitative data conveys more about reasoning behind the facts. The qualitative data also, helps to provide the interpretation behind the student behavior.

Table 1 Visual model of research method

Phase	Procedure	products
Quantitative data collection (AMS and BNS survey)	Descriptive analysis	Quantitative data
Qualitative data collection (Semi-structured interview)	Thematic analysis	Themes and subthemes
LA data transaction records	EDM techniques	Framework
Quantitative data collection (survey based on the framework from the previous phase)	Structure equation modelling	Framework

5.2 DATA SOURCES

5.2.1 ACADEMIC MOTIVATION SCALE

The academic motivation scale (AMS) evaluates students' motivation based on intrinsic motivation, extrinsic motivation and amotivation. The scale reflects a theoretical concept within Self-Determination Theory (SDT)(Edward L. Deci & Ryan, 2000). There are two versions of the scale used to measure human motivation; the first version assesses seven factors based on 28 items (Vallerand & Blssonnette, 1992). The second version comprises five factors which include intrinsic motivation, identified regulation, introjected regulation, external regulation and amotivation. Both versions have been tested in different contexts; however, the researcher determined that the five factors scale was more compatible with SDT. Also, the AMS has been tested with Italian high school students (Alivernini & Lucidi, 2008). A confirmatory factor analysis was used to confirm the validity of the five factors model of the AMS. The result was consistent with SDT. Recently, the five model has been tested with military cadets (Filosa et al., 2021), with results showing a reasonable fit between the data and SDT.

5.2.2 BASIC NEEDS SATISFACTION SCALE

Basic Needs Satisfaction in General Scale (BNSG -S) was developed based on the Self-Determination Theory (Gagné, 2003). BNSG - S was developed to be used in a general context rather than measuring SDT in a specific domain (Jenkins-Guarnieri, Vaughan, & Wright, 2015). The updated scale was retested and found to have a good overall fit to the data. Jenkins-Guarnieri et al. (2015) were able to provide validity evidence for the modified scale, by calculating bivariate correlations between the final 13 items subscale mean scores. The result showed significant relationships between each subscale. Moreover, Stolk, Jacobs, Girard, and Pudvan (2018) have used the new version of BNS scale by Jenkins-Guarnieri et al. (2015) to measure student basic need satisfaction in an engineering context. The result demonstrates that students in the course show a high level of basic need satisfaction. Also, competence and autonomy results were positively correlated with intrinsic motivation and identified regulation that is supported by SDT (Vallerand & Blssonnette, 1992). Relatedness, on the other hand, was less linked to intrinsic motivation compared to competence and autonomy. This is supported by studies that show that relatedness presents no contribution to the prediction of the outcomes Vlachopoulos and Michailidou (2006), supporting the hypothesis that relatedness might not be a necessary factor in maintaining intrinsic motivation (Edward L. Deci & Ryan, 2000).

5.2.3 STUDENT INFORMATION SYSTEM

The student information system includes data about the student's prior qualification, socio-economic status, ethnic group, module selections and grades.

5.2.4 VIRTUAL LEARNING ENVIRONMENT SYSTEM (BRIGHTSPACE)

Assessments, course information, access to learning materials, interaction with tutors and others through discussion forum and submitting assignments.

5.2.5 LIBRARY DATA

Library data system tracks student visit details, books borrowed and access to electronic journals. For instance, analytics can be used to check if a motivated student is linked with frequent access to library resources.

5.3 ETHICAL CONSIDERATION

There are different ethical issues that need to be addressed in LA that are related to student data privacy. The General Data Protection Regulation (GDPR) is used for data protection and privacy in the European Union (Calder, 2018). The data privacy regulations consist of guides on how student data should be used for research purposes, institutional improvements and when it is required to inform students about using their data (Bennett, 2018). Moreover, students need to be informed that any data used beyond the research purpose has all the personal information removed from the data set to make sure that personal data are treated anonymously (Slade, 2013). The accuracy of the data analysis interpretation is also critical because the result of the analysis has the potential to affect both institutions and students. In this research, all the participants have been informed about the research through a research information sheet and have taken their consent to take part in this research.

6 CURRENT STATE OF THE RESEARCH

The main aim of the study is to explore CS student motivation using SDT and student basic need satisfaction level. An overview of the descriptive statistics, mean, median, standard deviation (SD), skewness and kurtosis are illustrated in table 2. The average of the intrinsic motivation level from our sample is 3.88 and SD= 0.30. Also, the average of the identified regulation is 3.74 and SD=0.51, this result aligns with SDT theory. The average of autonomy is 3.37 and competence is 3.34 which supports the result of (E L Deci & Ryan, 2016). The students who felt competent and autonomous were most likely to have a high level of intrinsic motivation and identified regulation. As intrinsic motivation and identified regulation are highly interrelated with each other. For instance, students who are interested in CS and enjoy the course (Intrinsic motivation) are also, more likely to value the course (identified regulation).

Table 2 Quantative findings

Subscale	Mean	Median	Standard Deviation	Mode	Skewness	Kurtosis
Intrinsic motivation	3.88	4	0.30	4	-2.64	7
Identified Regulation	3.74	4	0.51	4	-2.46	6.24
Introjected Regulation	2.94	3.2	0.83	3.4	-0.63	-0.91
External Regulation	3.05	2.8	0.62	2.8	0.26	-0.99
Competence	3.34	3.4	0.37	3.4	-1.29	2.67
Relatedness	3.57	3.6	0.42	3.6	0.25	-0.79
Autonomy	3.37	3.4	0.68	4	-0.68	-1.25

Findings from the quantitative phase can be explored further with the semi-structured interview to extract CS student motivation factors. The quantitative result would give a broad picture of the student motivation measurements and their basic needs satisfaction level. Using the quantitative method first would give the researcher to expand their understating of student motivation in the qualitative phase and be able to find out the factors behind the quantitative result. The qualitative findings would be presented in themes and subthemes that are related to the research questions. The extracted themes would be trained from student LA data transactions using educational data mining techniques. The final stage would target a large number of students from CS department to test the framework using structure equation modelling (SEM). Using SEM will help to understand the level of connection between the variables, by building a theoretical model that generalizes the first stage findings through describing the statistical relationships among observed variables (Randall E. Schumacker & Lomax, 2005). SEM defines the extent to which the theoretical model is supported by sample data. SEM is particularly well suited to complex situations where psychological social concepts are studied which can otherwise be difficult to measure. Student motivation, which cannot be observed directly, benefits from such an approach (Schunk & Zimmerman, 2012), especially given that SEM explicitly takes measurement errors into account (Chou & Bentler, 1995).

7 Proposed solution

This study contributes to the current understating of student motivation factors. The proposed framework would assist in understating motivation factors for an undergraduate student from computer science department. The study is expected to confirm the result of existing studies that also, which emphasized the importance of motivational factors on student's academic journey ((Lonn et al., 2015), (de Quincey, Briggs, Kyriacou, & Waller, 2019)). However, this study would test the identified factors within the LA data records using data mining techniques.

8 Future work

This research focuses on developing a better understanding of students' motivation from CS department. The motivation factors are investigated based on SDT theory which divide human behaviors intro three main categories, intrinsic motivation, extrinsic motivation and amotivation. Also, the study explores students' autonomy, competence and relatedness. To further this focus in future, we suggest testing the identified behaviors on large sample. Moreover, it would be interesting to examine those behaviors on students from different departments like business and art.

REFERENCES

Alivernini, F. A. B. I. O., & Lucidi, F. A. B. I. O. (2008). The Academic Motivation Scale (AMS): Factorial structure, invariance and validity in the Italian context. . *Testing Psychometrics, Methodology in Applied Psychology, 15*(4), 211-220.

Bennett, C. J. (2018). *Regulating Privacy: Data Protection and Public Policy in Europe and the United States*: Cornell University Press.

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International journal of educational research, 31*(6), 445-457. doi:10.1016/S0883-0355(99)00014-2

Calder, A. (2018). *EU GDPR: a pocket guide* (Second ed.). Ely, Cambridgeshire, United Kingdom: IT Governance Publishing Ltd.

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. . 37–55.

Corrin, L., & da Barba, P. (2014). *Exploring students' interpretation of feedback delivered through learning analytics dashboards.* Paper presented at the ascilite 2014 conference.

Creswell, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches* (Fourth, International student ed.). Los Angeles: SAGE.

de Quincey, E., Briggs, C., Kyriacou, T., & Waller, R. (2019, 2019). *Student Centred Design of a Learning Analytics System*.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological inquiry, 11*(4), 227-268. doi:10.1207/S15327965PLI1104_01

Deci, E. L., & Ryan, R. M. (2016). *Optimizing students' motivation in the era of testing and pressure: A self-determination theory perspective.* . Singapore: Springer.

Filosa, L., Cepale, G., Perinelli, E., Cinque, L., Coscarelli, A., & Alessandri, G. (2021). The Military Academic Motivation Scale (MAMS): A New Scale to Assess Motivation Among Military Cadets From a Self-Determination Theory Perspective. *European journal of psychological assessment : official organ of the European Association of Psychological Assessment, 37*(3), 193-207. doi:10.1027/1015-5759/a000593

Flick, U., Kardoff, E. v., & Steinke, I. (2004). *A companion to qualitative research*. London: SAGE.

Gagné, M. (2003). The Role of Autonomy Support and Autonomy Orientation in Prosocial Behavior Engagement. *Motivation and emotion, 27*(3), 199-223. doi:10.1023/A:1025007614869

Jenkins-Guarnieri, M. A., Vaughan, A. L., & Wright, S. L. (2015). Development of a Self-Determination Measure for College Students: Validity Evidence for the Basic Needs Satisfaction at College Scale. *Measurement and evaluation in counseling and development, 48*(4), 266-284. doi:10.1177/0748175615578737

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in human behavior, 47*, 90-97. doi:10.1016/j.chb.2014.07.013

Miltiadou, M., & Savenye, C. (2003). Applying social cognitive constructs of motivation to enhance student success in online distance education. . *AACE journal, 11*(1), 78-95.

Pintrich, P. R. (2000). Multiple Goals, Multiple Pathways: The Role of Goal Orientation in Learning and Achievement. *Journal of educational psychology, 92*(3), 544-555. doi:10.1037/0022-0663.92.3.544

Randall E. Schumacker, & Lomax, R. G. (2005). A Beginner's Guide to Structural Equation Modeling *Evaluation and program planning, 28*(2), 245-246. doi:10.1016/j.evalprogplan.2005.01.006

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2016). *Research methods for business students* (7th ed.). Harlow, England: Pearson.

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in human behavior, 78*, 397-407. doi:10.1016/j.chb.2017.06.030

Schunk, D. H., & Zimmerman, B. J. (2012). *Motivation and Self-Regulated Learning: Theory, Research, and Applications*: Taylor and Francis.

Schweinle, A., Meyer, D. K., & Turner, J. C. (2006). Striking the Right Balance: Students' Motivation and Affect in Elementary Mathematics. *The Journal of educational research (Washington, D.C.), 99*(5), 271-294. doi:10.3200/JOER.99.5.271-294

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review, 46*(5), 30.

Slade, S., & Prinsloo, P. . (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510-1529.

Stolk, J. D., Jacobs, J., Girard, C., & Pudvan, L. (2018). *Learners' Needs Satisfaction, Classroom Climate, and Situational Motivations: Evaluating Self-Determination Theory in an Engineering Context.* Paper presented at the IEEE Frontiers in Education Conference (FIE).

Vallerand, R. J., & Blssonnette, R. (1992). Intrinsic, Extrinsic, and Amotivational Styles as Predictors of Behavior: A Prospective Study. *Journal of personality, 60*(3), 599-620. doi:10.1111/j.1467-6494.1992.tb00922.x

Vlachopoulos, S. P., & Michailidou, S. (2006). Development and initial validation of a measure of autonomy, competence, and relatedness in exercise: The Basic Psychological Needs in Exercise Scale. *Measurement in physical education and exercise science, 10*(3), 179-201.

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into practice, 41*(2), 64-70. doi:10.1207/s15430421tip4102_2

Metacognitive Judgement and its Influence on Decision-Making with Machine-Generated Explanations

Warren Li
University of Michigan-Ann Arbor
liwarren@umich.edu

ABSTRACT: Technologies in learning environments have potential to improve educational outcomes. By using machine learning models, online behaviors can be used to show students information about their study progress. However, such model outputs can be difficult to understand for users and lead to distrust towards the system. How learners choose to engage with information from these systems is mediated by metacognitive judgements of their own knowledge. This paper describes two studies to better understand this interaction between AI systems and human judgements and facilitate more mindful learning. The first study analyzes how judgements of learning (JOL) are related to usage of machine-generated explanations on a data science task, where JOL is changed through a reflection activity. The second seeks to understand how providing different levels of model transparency will influence the resources allocated towards learning a skill and its relationship with JOL.

Keywords: Metacognition, Metacognitive Judgements, Machine Learning Explanations, Self-Evaluation, AI Trust

1 INTRODUCTION

Many artificial intelligence (AI) systems and machine learning (ML) models have been leveraged in education to support learners. Yet, users of these systems such as students and instructors are not always conscious of how these systems operate. Even to developers, the nature of these models is often described as being a "black box" due to their lack of interpretability (Pieters, 2011). While the outputs of these models can be used in dashboards and are intended to help students, the effectiveness of these machine-generated explanations is also highly dependent on the end-user's interpretation. Opaque systems may lead to skepticism and low trust, especially when there exists a history of machine learning biases exacerbating inequities (Ocumpaugh et al., 2014). Poorly trained models and unrepresentative data sets means that well-intentioned models may not end up helping all learners, which leads to fairness concerns. In an era where information is readily available online and more tools are touted to improve learning, students must selectively evaluate sources, which can be a difficult metacognitive activity.

Metacognition can loosely be described as "thinking about thinking" and metacognitive judgements include Feelings of Knowing (FOK), Judgements of Learning (JOL), and other related constructs (Flavell, 1979). Students who are better able to assess their own knowledge have been shown to attain higher academic performance. Part of the reason for this may be that students can more effectively plan and monitor their own learning if they know what skills need reinforcement and when to initiate practice (Sternberg, 1998). In addition, there are many different sources of information that students can consult, such as analyses of self-assessment from a study tool. How people use this information may be heavily influenced by their metacognitive judgments; if people are not sure how much they know, they may be prone to over- or under-rely on machine-generated explanations (Azevedo, 2005).

Personal trust regarding AI will also affect how someone uses information from these sources, and recent developments have attempted to make such systems more accessible. For example, there are metrics to capture notions of fairness, datasets that might be augmented with samples from representative populations, and explainable ML techniques that provide more transparency. Thus, understanding the relationship between metacognition and machine-supported learning might translate into more effective online education, as learners learn to filter (mis)information from formal or informal sources in greater quantities.

2 RELATED WORK

2.1 Importance of Metacognitive Judgements in Education

Metacognitive regulation consists of monitoring an individual's cognition, as well as evaluating and changing behaviors based on this information. This includes planning out a task, reflecting on task performance, and having metacognitive experiences, examples of which may include finally "getting" a concept or feeling a sense of confusion (Schraw, 1995). Constructs such as JOL fall under this category, which are self-assessments about how well one has learned certain information; "accuracy" compares JOL with actual performance on subsequent tests (Pintrich, 2000).

The importance of developing a set of metacognitive knowledge and skills has been widely affirmed (Sternberg, 1998). When it comes to information gathering and evaluating sources, such as the vast repository of online resources today, metacognitive judgements play a crucial role. Stadler and Bromme (2007) state that synthesizing multiple web documents requires metacognitive judgements and a mental representation of content and sources. Learners must not only render a JOL and recall their existing knowledge in order to map acquired information into their semantic framework, but also evaluate the credibility of the source through indicators such as the author or affiliation. Unfortunately, students ranging from primary to post-secondary education often struggle to demonstrate metacognitive reflection when confronted with a new source, even when the new data results in a contradiction (Mason et al., 2010).

In particular, JOL used in source evaluation often reflects future recall and affects students' study time allocation. It might therefore be considered as a predictive or prospective judgement, though there are also retrospective performance judgements, such as when someone reflects on how well they performed on a task. Mengelkamp and Bannert (2009) also classify judgements into global versus specific judgements. Global judgements are made regarding a wide range of content that is tested with items, such as predicting an overall test score, whereas specific judgements cover a particular item. These varieties of judgements result in different accuracies and leverage different cognitive processes (Fleming et al., 2016), but are both important to study as learning often involves planning ahead as well as reflecting on previous attempts.

2.2 Mediating Interactions Between Students and ML Systems

While use of some AI systems in education have led to increased student performance, including efforts to improve certain aspects of metacognition such as hint-seeking (Roll et al., 2007), the hard-to-interpret nature of many AI tools such as predictive models may be difficult to interpret and lead to hesitancy from students and teachers alike. Explainable artificial intelligence (XAI) is an important nascent field that tackles some of these issues and provides insight into whether a model is making useful predictions. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME) provide the user with a local explanation for a particular prediction: features that were used and their relative importance can be displayed (Collaris and van Wijk, 2020). This enables greater transparency into these systems and can increase user trust and adoption.

However, machine-generated explanations come with their own limitations. People may not use these effectively and such explanations may end up subconsciously altering the user's beliefs in unintended ways (Bauer and Zahn, 2021). Consider a prediction indicating skew based on protected attributes, such as ethnicity or gender, where such considerations are inappropriate; this may inadvertently reinforce stereotypes. Even if an automated system can provide immediate feedback, students who are overly dependent on the machine might not develop the metacognitive skills to recognize when something is going wrong (Azevedo, 2005). Instead, confirmation bias may reaffirm their existing beliefs even when contrasting evidence is shown.

As digital and remote learning continues to grow, interactions with technology increase and so might one's exposure to information. Students need to be able to evaluate sources and distinguish between credible information and falsehoods, especially since misconceptions can be difficult to correct once embedded into long-term memory. Salovich and Rapp (2020) demonstrated that those with higher metacognitive skills had a greater resilience towards exposure to inaccurate information. Yet, this has not been studied in an authentic educational context with modern tools, which is critical given the increasing prominence of online learning platforms.

3 RESEARCH QUESTIONS

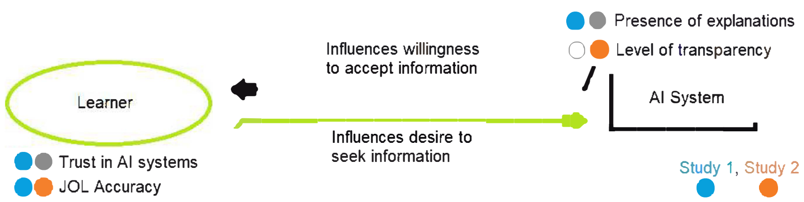


Figure 1: Interactions between learners and AI systems, and variables of interest in Study 1 and 2.

Consequently, I seek to understand the relationship between accuracy in metacognitive judgments, how students engage in decision-making processes given AI explanations, and how that is mediated by their baseline trust in such systems (Fig. 1). Specifically, I aim to answer the following questions:

RQ1: How is the accuracy of learners' JOL related to how they use or disregard machine-generated explanations on a completed task?

To answer this question, two factors will be included in the study design: an (i) AI explanation and a (ii) highlighting factor. I will (i) measure the rate at which learners accept or reject explanations from an AI; (ii) code highlighting activities, explained below, will be used to influence learners' JOL.

Study 1 can reveal the interaction between learners' willingness to use AI explanations and their accuracy of JOL *after* learning has occurred. Given this baseline, I can then compare how JOL accuracy changes when we adjust how the AI portrays information *during* the learning process.

RQ2: Does the level of transparency of a knowledge model and learners' JOL accuracy relate to their decision to continue or stop studying a skill?

Regarding study 2, I focus on the transparency of the knowledge model. By hiding or displaying whether the AI believes the student has mastered a skill, and letting a student decide when to start or stop study for a particular skill, I capture how JOL accuracy is influenced by the AI system design.

4 PROPOSED SOLUTION

4.1 Study 1: Influencing Metacognitive Judgements and Usage of Self- Versus Machine-Generated Explanations

In order to answer RQ1, I will draw participants from a course in an online degree program. This 1 credit-hour course lasts 4 weeks. Each week, students will complete a programming homework assignment, where one problem from each homework will be modified for the purposes of this study. Students will be split into one of 2 (highlighting activity) x 2 (AI explanation) conditions:

In terms of the highlighting activity, the first group will be shown their own code along with a rubric, which details skills being assessed for the relevant problem after each week's corresponding homework deadline. Students will be asked to highlight the lines of code to provide evidence that they understand each skill; students will then repeat the process, finding evidence that casts doubt on skill mastery. This is a metacognitive reflection activity used to change students' awareness of their existing knowledge. Finally, they will be asked to self-evaluate their mastery of the skill on a 5-point Likert scale. Follow-up multiple-choice questions (MCQ) associated with each skill will be used to assess their actual knowledge in a post-test. The difference in their self-evaluation and post-test scores will be the metric for JOL. The second group will not go through the highlighting activity. Instead, students will proceed to the self-evaluation and post-test.

For the AI explanation condition, half of the students from each group above will be given a code highlighting explanation during the self-evaluation phase, which they are told will be generated by an AI; the other half will not. The explanations shown will be the same across all students. Those that are given an explanation will have an opportunity to change their self-evaluations.

During the start of the course, students will take a pre-test; this will be compared with the post-test MCQs to measure knowledge gained. After finishing the last assignment, students will complete the Human-Computer Trust Scale (HCTS), which is a 12-item scale measuring user trust in HCI systems (Gulati et al., 2019). Using a regression model, I will analyze the difference between JOL accuracy for each of the four conditions while controlling for trust levels and prior knowledge to see if those who received AI explanations adjusted their self-evaluations to increase alignment with post-test scores.

4.2 Study 2: Self-Determination in Scrutable Knowledge Models and its Relation to Metacognitive Judgments

To answer RQ2, students will be asked to work through a linear path consisting of several concepts that build on each other. Participants will be drawn from a MOOC on databases. This course includes an SQL tutor, which will be leveraged to collect data on a three-topic progressive sequence. There will be two treatment conditions: one where a knowledge model showing concept mastery is visible, an d one where it is not. There will also be a control condition where the student completes the modules following the prescribed timing of the knowledge model and cannot choose to move ahead or lag behind. Overall efficiency will be measured by both total time spent and performance on a post-test. The difference in efficiency given the student's actual decisions to continue and the computed expected efficiency given the AI's idealized decisions (based on when a mastery probability threshold is first reached) will constitute the metric for JOL.

This is important as many current tools such as intelligent tutoring systems (ITS) do not permit students to modify the models based on their own metacognitive knowledge of themselves, and do not show the computer's estimation of their skills. However, introducing such mechanisms might improve students' JOL accuracy as well as satisfaction, which may affect motivation and general persistence. Furthermore, having students indicate their JOL by deciding when to move on may be more authentic and less tedious than explicitly declaring a confidence level after each question.

5 CURRENT AND FUTURE WORK

5.1 Current Progress

We have presented a paper at LAK that shows disparities in grade predictive performance between subgroups of students when they were given the opportunity to selectively opt-out of data collection (Li et al., 2019). We then explored why students might choose to opt-out and found that low institutional and instructor trust are key considerations when disallowing data usage and might be mitigated by making the interpretation and function of models clearer (Li et al., 2021).

Lastly, I wrote and defended a literature review on metacognition and educational measurement, and found that there are a number of different instruments used to capture this construct that do not converge. Combining various sources of data such as questionnaires or logs from students work ing on computerized tasks, as is the case with the aforementioned studies, may enable us to develop a practical assessment of metacognition along with systems to scaffold and monitor the development of these skills. The studies for my thesis are in the planning stages and the first experiment is expected to be underway next year.

5.2 Future Studies

Even if we are able to adjust students' JOL, or nudge behavior to make students' more effective consumers of information, it may be useful to understand which factors learners with higher or lower JOK attribute their outcomes. According to attribution theory, people may have differing views on their locus of control, or whether or not a circumstance is in their ability to change. For instance, we could intentionally give some problems that are solvable while others that are beyond the student's capabilities, and ask them to explain why they received a correct or incorrect mark.

REFERENCES

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, 40(4), 199-209.

Bauer, K., Hinz, O., & von Zahn, M. (2021). Expl(AI)ned: The impact of explainable artificial intelligence on cognitive processes. *SAFE Working Paper*, 315.

Collaris, D., & van Wijk, J. J. (2020, June). ExplainExplore: Visual exploration of machine learning explanations. *In 2020 IEEE Pacific Visualization Symposium (PacificVis)*. 26-35.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906.

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J. C. (2016). Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, 2016(1), niw018.

Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology*, 38(10), 1004-1015.

Li, W., Brooks, C., & Schaub, F. (2019, March). The impact of student opt-out on educational predictive models. *In Proceedings of the 9th International Conference on Learning Analytics & Knowledge,* 411-420.

Li, W., Sun, K., Schaub, F., & Brooks, C. (2021). Disparities in Students' Propensity to Consent to Learning Analytics. *International Journal of Artificial Intelligence in Education*, 1-45.

Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: evaluating and learning online information. *Metacognition and Learning*, 5(1), 67-90.

Mengelkamp, C., & Bannert, M. (2009). Judgements about knowledge: Searching for factors that influence their validity. *Electronic Journal of Research in Educational Psychology*, 7(1), 163–190.

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.

Pieters, W. (2011). Explanation and trust: what to tell the user in security and AI?. *Ethics and Information Technology*, 13(1), 53-64.

Pintrich, P., Wolters, C., & Baxter, G. (2000). Assessing metacognition and self-regulated learning. *Issues in the Measurement of Metacognition*, 43-98.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, 2(2), 125-140.

Salovich, N. A., & Rapp, D. N. (2020). Misinformed and unaware? Metacognition and the influence of inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 47(4), 608–624.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351-371.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, 2(2), 191-210.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student?. *Instructional Science*, 26(1), 127-140.

From Data to Actions: Unfolding Instructors' Sense making and Reflective Practice with Classroom Analytics

Dana AlZoubi
Iowa State University
dalzoubi@iastate.edu

ABSTRACT: The ultimate goal of using learning and teaching analytics dashboards is to improve teaching and learning processes. Instructors that use an analytics dashboard are presented with data about their students and/or about their teaching practices. Despite growing research in analytics dashboards, little is known about how instructors use the dashboards, make sense of the data they receive, and reflect on it accordingly. Moreover, there is limited evidence on how instructors who use these dashboards take further actions and improve their pedagogical practices. My dissertation work addresses these issues by actively involving the instructors in the design phases of the dashboard and examining their sense making, reflective practice and subsequent pedagogical actions they take based on their classroom analytics. My dissertation includes three phases: (a) systematic literature review that identifies problems & patterns from instructor use of dashboards (current), (b) implementation and examination of instructors' sense-making and reflective practice (current) and (c) human-centered approaches to co-designing instructors' dashboards with stakeholders (current). The findings will contribute to the conceptual basis of instructors' change of their pedagogical practices and practical implications of human-centered principles in designing effective dashboards.

Keywords: instructor dashboards, sensemaking, reflective practice, human-centered design.

1 INTRODUCTION

In the past decade, learning analytics dashboards have been identified as critical emerging technologies that have high expectations on unfolding the black box of teaching and learning complex processes (Johnson et al., 2016; Wise & Jung., 2019). Instructors using analytics dashboards are presented with visual graphs that include information about their students and/or their teaching strategies (Park & Jo, 2015; Verbert et al., 2014). Instructors' ability to transform the information from the data present on dashboards into actionable pedagogical practices is one of the highly touted potential outcomes of dashboard use; yet literature reveals that most instructors struggled in connecting the data with their teaching practices (Wise & Jung, 2019). Moreover, research is still scarce in examining how exactly instructors respond to and make use of analytics from dashboards (van Leeuwen et al., 2017), with the current literature highlighting great differences in how instructors used dashboards, how they made sense of the data, and what actions they took accordingly (if any) (Molenaar et al., 2019). Missing from learning analytics research is an investigation of how instructors interact with data and how they can use this data to change pedagogical practices.

Moreover, although a substantial literature has been dedicated to understanding effective analytics dashboards design (for e.g., Martinez Maldonado et al., 2016; Bodily & Verbert, 2017); yet the impact of using them in improving teaching and learning is still not evident (Viberg et al., 2018). The main challenge presented in the literature is rooted by disconnectedness of the users interacting with the dashboard while focusing more on the technical aspects of dashboard design. There is a recent shift in the field towards human-centered learning analytics that assures that “meanings, interaction opportunities, functions and attributes associated with the system should be defined by the people for whom the system is intended, rather than imposed by designers and researchers” (Buckingham Shum et al., 2019; p2). Human centeredness is defined in different ways and at different levels, yet in all ways it encompasses designing systems with users rather than for users. Thus, there should be consensus on the processes of integrating a human-centered perspective while taking into account the aforementioned identified problems of instructors’ sense making and reflective practice to take subsequent actions accordingly.

2 CURRENT KNOWLEDGE & CONCEPTUAL FRAMEWORK

2.1 Instructors’ Use of Analytics & Perceived Challenges

Instructor dashboards are a specific application of learning analytics dashboards that allows instructors to monitor their students’ progress and evaluate their teaching strategies. The prominence of using analytics seeks to acquire insightful feedback from data that ordinarily would not be captured or recalled, except with the application (Ndukwe & Daniel, 2020). For instance, while some dashboards present instructors with visual graphs that include information about their students (Verbert et al., 2014) , others provide information about their facilitation of teaching strategies such as classroom management (Park & Jo, 2015). Using information about instructors’ and students’ actions can help instructors change their pedagogical practices (Gasevic et al., 2016).

Results from studies with instructors using the dashboards highlighted challenges in the adoption of the dashboard tools and in data literacy, two main themes that are interconnected at many levels. The adoption of such tools can be referred to as the intention to use technology tools and thus was explained through different models with the most common being technology acceptance model (TAM) that suggests the perceived usefulness and perceived ease of use of the tool influences the behavior of the user (Davis, 1989). Data literacy can be defined as the ability can be seen as a combination of being able to read and interpret information and possessing the skill to connect those inferences to relevant pedagogical knowledge to select an adequate follow-up action (Mandinach & Gummer, 2016). As such, data literacy is a complex construct that requires the measurement of several sub-constructs such as graph literacy, etc. (Van Den Bosch, Espin, Chung, & Saab, 2017).

The review on the research of the instructor dashboard revealed that most instructors struggled in connecting the data with their teaching practices (Wise & Jung, 2019). Moreover, there were great differences in how instructors used dashboards, how they interpreted information on the dashboard, and what kind of actions they took (if any) in connection to the data from the dashboard (Molenaar & Knoop-van Campen, 2019). Although the importance of data literacy is acknowledged, Schwendimann and colleagues (2017) indicate in their review of LADs that there is very little research that explores the relation between users’ data literacy and the usefulness of a dashboard.

2.2 Instructor Sense Making and Reflective Practice of Learning Analytics

An important aspect of using dashboards is to go beyond just the use of dashboards by examining what actions instructors take based on the data provided from the analytics. This can be conceptualized through instructors’ sensemaking and reflective practice that will dictate the subsequent actions to be taken. Sensemaking is a social process of searching for meaningful answers that drive the actions that individuals take (Weick, 1995). Sensemaking situated within analytics data requires instructors to understand what is happening, reflect on their practices and decide on what further steps to take to improve pedagogical strategies (Wise & Jung, 2019). Wise and Yung’s (2019) proposed a model of teachers’ process of analytics that embeds teacher activities within a two-part structure of sense-making and pedagogical response. Reflective practice is a critical catalyst for improving pedagogical practices as it allows instructors to identify problems and plan for actions (Walkington et al., 2001). Reflection that occurs through data-informed feedback is significant since it allows making sense of the provided evidence (Avramides et al., 2015). As instructors look at the dashboard analytics, analyze data on it and make meaningful interpretations, the most prominent step is the actions that they take accordingly which can inform their future pedagogical practices. When compared with other traditional feedback from observational protocols, sensemaking can help better understand instructors' actions (Wise & Jung, 2019). Reflective practice can encourage instructors to use the automated feedback to foster teaching strategies.

3 TEACHACTIVE PROJECT

This project is based on an NSF IUSE grant titled "An Integrated Faculty Professional Development Model Using Classroom Sensing and Machine Learning to Promote Active Learning in Engineering Classrooms" (DUE #2021118). TEACHActive model uses machine learning and computational analysis of classroom analytics from an automated observation system, EduSense (Ahuja, 2019), and presents visual analytics via TEACHActive dashboard designed for engineering instructors (AlZoubi et al., 2021). Through the TEACHActive model, engineering instructors in higher education engage in a series of sense making and reflective practices. Raw classroom data is transformed into meaningful metrics, and then these metrics are being displayed as classroom analytics to provide practical feedback for instructors. TEACHActive model includes three main components (a) training on using pedagogical models (for e.g. active learning strategies), (b) automated classroom observation, and (c) feedback in the form of classroom analytics from automated observation followed by reflective prompts. Instructors use the automated feedback on the session display and the progress display to reflect on their pedagogical practices and take actions accordingly.

As such, instructors reflecting on their classroom teaching through the automated feedback can make pedagogical changes in their future sessions. For example, an instructor may devote additional time for group work if the data suggests that a substantial portion of in-class activities were dominated by the teacher.

4 RESEARCH GOALS AND QUESTIONS

The purpose of this research is to examine how instructors' active participation in this model inform change in their pedagogical strategies through sensemaking and reflective practices of analytics data. The research questions that guide my dissertation include the following:

RQ1. How do instructors make sense of classroom analytics displayed on dashboards?
RQ2. How can data from the analytics dashboard be linked/ aligned with instructors' reflection on pedagogical practices?
RQ3. What are effective human-centered design principles used to design dashboards with instructors?
RQ4. How do human-centered principles influence instructors' sense-making and reflective practices?

5 METHODS & CURRENT STATUS OF WORK

Following the designed based research methods and using human-centered design principles, we designed the first iteration and implementation of TEACHActive model and the corresponding feedback dashboard that visualized the automated classroom observation output (AlZoubi et al., 2021, a,b). Results from the pilot data of first implementation revealed that instructors identified some of the metrics to be more meaningful than others. For example, students' hand raises as a function of time (i.e. a scatter plot that lays the number of seconds of hand raises at different moments of the class) was perceived to be more meaningful than numeric metrics that only provided a total frequency (i.e. total seconds of hand raises during class was 22 seconds). Instructors also perceived that the classroom analytics provided with reflective prompts on TEACHActive dashboard provided promises to facilitate feedback on their future teaching. Reflective prompts were particularly interesting to them as they assured instructors that the analytics on the dashboard is not evaluative but rather descriptive. A such, reflective prompts would provide a better context about what was happening during the class. An example would be showing instructors the total number of hand raises and asking questions about why were students raising their hands (whether it is to ask questions, participate in class discussions or another reason). Also, asking instructors whether they would like to set a goal for next session, and whether they find this metric descriptive and indicative of what was happening in their classrooms. Reflection prompts were added on a separate display and in relation to all the metrics provided on the session display.

5.1 Phases of my dissertation work

My dissertation work includes three main phases: a) systematic literature review that identifies problems & patterns from instructor use of dashboards (current), (b) implementation and examination of instructors' sense-making and reflective practice (current) and (c) human-centered approaches to co-designing instructors' dashboards with stakeholders (current). Figure 1 illustrates the three phases.

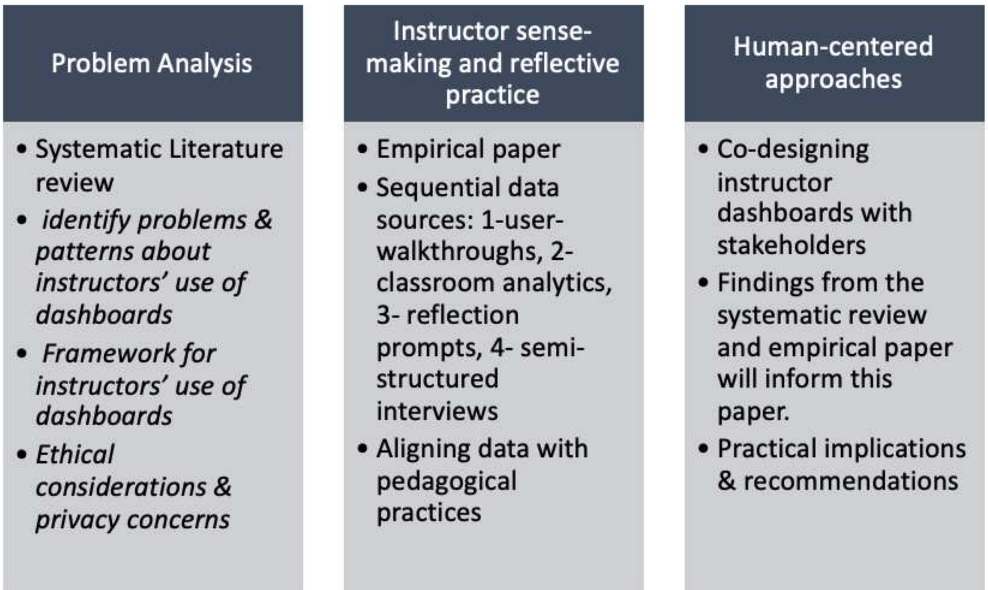


Figure 1. Three phases of dissertation

For the first phase, I am currently preparing a systematic literature review on the state of the art of instructor dashboards and their uses, empirical evidence from the literature on how instructors interacted with the dashboard, how they make sense of the data. Also, I am examining the affordances, the challenges and the practices used while providing instructors with analytics. I am using PRISMA guidelines and I aim to provide a framework for instructor use of dashboards as a contribution from this systematic literature review. This review is currently in the data analysis phase.

The second phase and the third phase of my dissertation will be simultaneously done with the future implementations of TEACHActive model. I aim to examine how instructors make sense and reflect on classroom analytics and what pedagogical actions do they take accordingly. For the second implementation phase of TEACHActive, instructors, first I will conduct user interviews and walk-throughs prior to TEACHActive automated observation implementation. In the user walkthroughs, I will walk instructors through the metrics present on the dashboard, examine which metrics they find meaningful, discuss their interpretations of each of the metrics, discuss expectations from the analytics data, understand their concerns and recommendations. Together with the instructors, we will co-design the reflection prompts displayed in a way that better targets their goals. After that, the automated observation will take place for four-weeks. During the four weeks, instructors will have two cameras in the class (one facing them and the other facing the students) that will track behavioral indicators from classroom through EduSense system (AlZoubi, 2021). EduSense tracks body positions of the instructor and the students in the room, hand raising, and their speech. Our particular features of interest include time spend sitting vs standing, hand raises, movement patterns, frequency of student speech, frequency of instructor speech, moments of silence, student vs. instructor speech (time and distribution over the class period) (Ahaju et al., 2019). After teaching each class session, instructors will be presented individually with the classroom data on the feedback dashboard. Instructors will receive an email notification that the classroom analytics are ready on the dashboard. They will be asked to review the classroom analytics and complete the short reflection prompts on the feedback dashboard. This automated feedback will illustrate the visual representation of behavioral indicators from the classroom in connection with the pedagogical strategies. For instance, changes in class activity (i.e., sit vs stand), student participation through hand raises, body positions, movement patterns, and the frequency and duration of instructor vs. student talk will be displayed on the feedback dashboard at the end of each session. Additionally, the feedback dashboard will present aggregated data of the captured features from instructor behaviors (e.g., sit vs stand, movement patterns, and body positions) and from student behaviors (e.g., hand raises, student vs. faculty speech, and frequency of speech). Comparison stats between sessions that reveals their progress (e.g., "There were 10 more hand raises in the second session compared to the first session," or "Your facilitation behavior increased by 22%") will be provided as well. The feedback dashboard will be used to promote instructors' reflective practice by gaining a better overview of the in-class activities, reflecting on their facilitation strategies, and addressing pedagogical changes in their next sessions. The reflection prompts will be personalized and based on each instructor's data. After the four-week observation period, I will conduct semi-structured interviews with instructors. Data from the interviews provides an in-depth understanding of their sense-making of the analytics data, how they use their pedagogical knowledge to interpret the feedback dashboard data and what actions they took in response to the displayed data. These interviews will also provide input to the future iterations and modification of the dashboard. I plan to measure instructors' sensemaking and reflective practice as well as any pedagogical change and action taken accordingly. I have pilot data for this phase that I am analyzing and modifying the phase accordingly. I need further revision and feedback on data sources and how to measure reflective practice and sensemaking. The third phase of my dissertation includes the human-centered co-design principles and recommendations for practical implications. For this phase, I will use the co-design with human-centered principles that I used with instructors. This part will be dependent on the systematic literature review findings and the empirical piece on how instructors make sense of the data and act accordingly. I already integrated human-centered principles in the first iteration, however, with more data, I will examine what principles were more effective as well as how these principles influenced instructors' sensemaking and reflective practice.

REFERENCES

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

AlZoubi, D., Kelley, J., Baran, E., B. Gilbert, S., Karabulut Ilgu, A., & Jiang, S. (2021, May). TeachActive Feedback Dashboard: Using Automated Classroom Analytics to Visualize Pedagogical Strategies at a Glance. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

AlZoubi, D., Kelley, J., Baran, E., Gilbert, S.B. , Jiang, S., & Karabulut-Ilgu, A. (2021). Designing the TEACHActive Feedback Dashboard: A Human Centered Approach.*11th International Conference on Learning Analytics and Knowledge (LAK21) Online*.

Avramides, K., Hunter, J., Oliver, M., & Luckin, R. (2015). A method for teacher inquiry in cross-curricular projects: Lessons from a case study. *British Journal of Educational Technology*, *46*(2), 249-264.

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405-418.

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

Gašević D, Dawson S, Siemens G (2015) Let's not forget: learning analytics are about learning. TechTrends 59(1):64–71

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. The Internet and Higher Education, 28, 68-84.

Johnson L, Adams S, Cummins M, Estrada V, Freeman A, Hall C (2016) NMC horizon report: 2016 higher education edition. The New Media Consortium, Austin. http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf

Karabulut-Ilgu, A., AlZoubi, D., & Baran, E. (2021, July). Exploring Engineering Faculty's Use of Active-learning Strategies in Their Teaching. In 2021 ASEE Virtual Annual Conference Content Access.

Kelley, J., **AlZoubi, D**., Gilbert, S.B., Baran, E., Karabulut-Ilgu, A., & Jiang, S. (2021). University Implementation of TEACHActive – An Automated Classroom Feedback System and Dashboard. In *Proceedings of the 65th Human Factors & Ergonomics Society (HFES) Annual Meeting.* October 4 - 7, 2021, Baltimore, USA.

Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016, April). Interactive surfaces and learning analytics: Data, orchestration aspects, pedagogical uses and challenges. In *Proceedings of the Sixth international conference on learning analytics & knowledge* (pp. 124-133).

Molenaar, I., Horvers, A., & Baker, R. S. (2019). What can moment-by-moment learning curves tell about students' self-regulated learning?. *Learning and Instruction*, 101206.

Ndukwe, I. G., & Daniel, B. K. (2020). Teaching analytics, value and tools for teacher data literacy: A systematic and tripartite approach. *International Journal of Educational Technology in Higher Education*, *17*(1), 1-31.

Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. Journal of Universal Computer Science, 21(1), 110.

van Leeuwen, A., van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: A case study. Learning: research and practice, 3(1), 42-58.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. Personal and Ubiquitous Computing, 18(6), 1499-1514.

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.

Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Sage.

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.

Conversational Agents for Collaborative Science Learning

Adelson de Araújo
University of Twente
a.dearaujo@utwente.nl

ABSTRACT: In online science labs, where students interact with complex ideas experimentally, the use of chats during the activity can be particularly diverse, sometimes productive and engaging, but other times also unproductive and distracting. In this doctoral project, we will design a conversational agent to support productive talk among students when learning in Go-Lab. Some of the main challenges of this endeavor include addressing when to intervene and what to say. Furthermore, the agent configuration procedure is also a major issue, since it often requires an effort from the teacher for the interventions to work effectively. My project will explore solutions for these challenges by proposing a set of dialogue variables extracted using machine learning and natural language processing to feed a triggering mechanism. The experiments will include an assessment with real classrooms to answer the research questions presented.

Keywords: conversational agents, online science labs, collaborative learning, machine learning

1 INTRODUCTION

Conversational agents (CAs) in online learning platforms have been regarded as means to scale personalized feedback across various contexts in education, showing great potential to intervene in a diversity of tasks and thus deliver a more engaging student experience (Kerly et al., 2007). CAs can introduce dynamic forms of guidance during students' dialogues, particularly in STEM education, where traditional teacher-led instructional approaches may not lead to deep conceptual knowledge, but more student-directed approaches are also needed (de Jong, 2019). In the last decade, CAs have been further modeled to support students learning collaboratively to develop explanations about the topics at hand through intervening in online chats, e.g., (Dyke et al., 2012; Kumar & Rosé, 2011; Tegos et al., 2016). These collaborative settings are the focus of the current project.

When using online labs in their classrooms, teachers face challenges in supporting small groups of students learn using a chat collaboratively. Students can get easily distracte and shift their attention from the task at hand to other topics. Since the teacher is not always there, the conversation can become unfocused and unproductive. To address this issue, a CA could be designed to trigger interventions that generate productive responses on students, and also make the discussion transactive (Dyke et al., 2012). In this project, we will develop CAs with the aim of *supporting productive talk among students*, for example, by asking them to give concrete examples about their arguments, stimulating explicit reasoning, and balancing participation.

Along with the potential to attract students to participate in productive discussions, implementing CAs for collaborative learning is also a significant design challenge. The communication mechanism involves addressing not a single user, as a regular chatbot, but a group of students at an opportune time with an appropriate sentence. To decide on how a CA for collaboration must act, a configuration procedure is required to instantiate triggers and utterances adapted to a specific purpose appropriately. Michos et al. (2020) conceptualized this challenge as an effort that involves three main pedagogical aspects, namely the intervention strategy (how to approach the students), the task design (what are the learning objectives), and the domain model (concepts, topics, and curriculum).

Related work usually investigates variations in the intervention strategy given a fixed task design and domain model, e.g., (Michos et al., 2020; Tegos et al., 2016). They implement intervention strategies with a rule-based approach primarily based on students' use of vocabulary, monitoring target words and synonyms, and other auxiliary variables, such as if the agent already intervened recently and the message counts for each student. These teacher-configured rules have the benefit to be more accessible than those configured by technical experts, but still require intensive teacher effort to configure.

In the first part of my project, machine learning and natural language processing techniques will serve as means to propose a hybrid configuration process that does not rely directly on the teachers' ability to predict students' use of vocabulary. We aim to formulate a triggering mechanism that makes the CA intervene using one of the predefined set of utterances if certain dialogue variables cross threshold ranges. The process workflow on how the CA is triggered is illustrated in Figure 1. As a result, the CAs can operate in online science labs across several concept domains given a fine-tuning configuration made by the teacher, specifying a few topics of interest with which dialogue variables are calculated to feed the triggering mechanism.

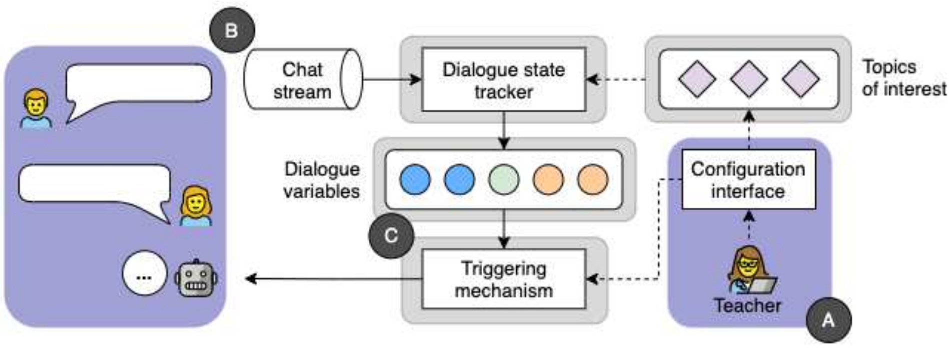


Figure 1: The proposed approach to control conversational agents. In purple, the learning environment and in grey, the components of agent service. (A) The teacher instantiates the agent by configuring topics of interest and optionally triggering parameters. (B) The students talk in the chat, and messages are sent to the agent. (C) Dialogue variables are computed for every message, and the triggering mechanism validate if it is an appropriate moment for intervention.

The second part of this project aims to produce a dashboard module for teachers to follow the progress of and orchestrate chat discussions. Using the same dialogue variables proposed to control the CA, we will design a dashboard with aggregated information about the quality of students' contributions, both at individual and group levels. The purpose is to complement the operation of CAs in chats where they do not support students appropriately. The dashboard module will provide teachers with real-time information about the quality of discussions to identify groups of students to target more specific and timely feedback.

As the main envisioned product, this project aims to contribute with empirical evidence towards more data-informed approaches that teachers can rely on to design and orchestrate collaborative learning activities, particularly when using online science labs. In the following sections, the first part of the project is detailed in more depth, including the research goals, the description of the dialogue variables and triggering mechanism, and the current development status. Thus, a more focused discussion on the second part of this project, which consists of designing a dashboard with the dialogue variables, will be addressed in future work.

2 RESEARCH GOALS

The focus of the first part of this project is to develop CAs with the purpose of supporting productive talk among students, while being able to transfer these CAs to different online science labs. To achieve so, we will propose an innovative configuration process based on machine learning and natural language processing techniques. In this context, the first research question that we will examine addresses the general effectiveness of our approach.

> RQ1: To what extent can a conversational agent for collaboration help students talk more productively in chats when learning with online science labs?

As we discussed above, the rule-based approach often adopted by related work presents difficulties. One of them is that the teacher must predict which words students will use when talking productively. In addition, to transfer or instantiate a CA to another domain model, a completely different set of words are required, and teachers end up putting a reasonable amount of effort to configure the agent and expect it to work appropriately for one single learning activity. The approach proposed in our project is different in the sense that instead of rules explicitly built on students' use of vocabulary, we will design dialogue variables that transform chat messages into indicators that represents the state of the conversation. Then, a triggering mechanism will account for such variables to produce personalized interventions based on them.

One of the main challenges we will face is to produce an effectively transferable configuration process that works reasonably well for different domains. With the combination of dialogue variables and triggering mechanism, we aim to develop a more transferable configuration process in which teachers can spend little effort configuring an agent for several learning activities. Thus, in the second research question of this project, we will examine whether the proposed approach contributes to the prospect of more transferable CAs.

> RQ2: How transferable the proposed configuration process can be to generate effective conversational agents in different online labs?

3 METHOD

3.1 Learning environment

The learning backdrop where this project will focus on is online science labs. The task design follows an inquiry-based learning approach in which students engage with the subject as scientists, solving problems by formulating hypotheses and answering them with experiments. For example, when learning about radiation and its effects on humans, students first read a learning material with key concepts and examples of historical tragedies involving radiation. Then they identify key variables of radiation exposure, make hypotheses about them, interact with the laboratory to collect data, investigate how the data answer the hypotheses, to finally produce evidence about how radiation exposure works. During this process, they can discuss with peers to build arguments together.

More specifically, the CAs will be developed and integrated with the Go-Lab ecosystem (de Jong et al., 2021). In Go-Lab, teachers can create, publish or use "Inquiry Learning Spaces" (ILSs) where students can navigate through the lesson, interact with the labs, complete questionnaires while also interacting with colleagues in the chat. In our experiments, participating teachers will be able to instantiate CAs for a particular ILS while configuring basic preferences about how the CA should operate, for example, defining the relevant topics that the dialogue should be enhanced. The ILSs will be designed to follow a standard set of sequenced inquiry-based learning phases (Pedaste et al., 2015), namely orientation, conceptualization, investigation, conclusion, and discussion.

3.2 Intervention model

Given the purpose of the CAs we highlighted earlier, *to support productive talk among students*, we start by investigating *what to say* to the students. As a trend in the field, the Academically Productive Talk (APT) framework proposed by Michaels et al. (2018) has been effectively used in several recent studies of CAs for collaborative learning, e.g., (Dyke et al., 2012; Michos et al., 2020; Tegos et al., 2016). It defines a set of "talk moves" designed to increase the amount of transactivity in the discussion. We will investigate how to implement APT moves in a suitable way for specific intervention strategies. The talk moves currently considered to be used are listed in Table 1.

Table 1: A selection of APT moves that will be tested in this project. [1]

Talk move name	Utterance
Example	"Can you give an example to illustrate your argument?"
Press for reasoning	"Why do you think that?"
Expand reasoning	"That is interesting. Can you elaborate more on that?"
Add-on	"Would you like to add something to what your partner just said?"
Agree/Disagree	"Do you agree or disagree with what your partners just said? Why?"

3.3 Dialogue variables

We will explore when to intervene by using machine learning and natural language processing techniques. We will extract dialogue variables to feed a triggering mechanism. The dialogue variables currently considered to be used when identifying moments to intervene are listed in Table 2. To extract the codes of collaborative discourse, we will use a classifier that we trained using 20k coded chats to identify two groups of codes using the coding scheme proposed by (Eshuis et al., 2019). The first group of codes is whether the message is "domain", "coordination", or "off-task", and the second

[1] Find more examples of talk moves in this link: https://www.serpinstitute.org/wordgen-weekly/academically-productive-talk

is whether the message is "informative", "argumentative", "asking for information", "active motivating" or none of these. We expect these codes will be relevant to identify appropriate messages for intervention when integrated as part of the triggers available.

Furthermore, for a given learning domain, we will identify topics that represent the main concepts students are supposed to discuss. We will use an automatic approach based on topic modeling to extract topic words from the learning material or Wikipedia. The teacher will be able to manually edit these topic words in the configuration process. The dialogue variable for topic similarity will measure the semantic similarity between a chat message and the main words that compose the topic. This similarity is computed between sentence vectors that we will extract from language models, such as Universal Sentence Encoder or BERT. The topic accumulation is simply the sum of topic similarities throughout the chat, sampled by student. Lastly, we will consider the pace of the student dialogue to see if that indicates a helpful pattern to control agent interventions.

Table 2: Set of dialogue variables measured after each new message from a group of students.

Name	Symbol	Description
Codes of collaborative discourse	$C_{i,j}$	The probability of a chat message being classified as a code *j* from the *i*-th group of codes. This variable is retrieved by a supervised learning model (classifier) that learned from another coded dataset.
Topic similarity	S_t	The semantic similarity (dot product) between the sentence vectors from the chat message and the *t*-th topic's main words.
Topic accumulation	$A_{t,u}$	For each student *u*, the topic similarities for each topic *t* are accumulated to represent how much the student already talked about the topic in the chat.
Pace	P_u	The ratio between the number of messages the student *u* sent and minutes past since the chat started.

3.4 Triggering mechanism

For each talk move shown in Table 1, a different set of dialogue variables and thresholds can be used to trigger it on the appropriate occasion. Our preliminary approach will consist of data analysis on samples of chats that we collected to find these thresholds. By testing different threshold values for each dialogue variable, we will search for ranges of values that generate desired CA behaviors, considering the sensitivity of each variable. With the thresholds found, we will conduct a series of usability tests with the configured CA to assess how well they were triggered and how the students behaved after the interventions. By iteratively improving the CA on different learning activities, we will address the project's research questions.

4 CURRENT STATUS OF WORK, NEXT STEPS & ETHICAL CONCERNS

Currently, we are developing our first prototype. The dialogue variables are already implemented, and we recently submitted a paper where we compute the codes of collaborative discourse training a classifier with a multilingual sentence encoder and contextual information from the chat, achieving around 74% of accuracy and 0.58 of Cohen's kappa. We plan to improve this model in the future, since that is an important part of the dialogue variables. We are investigating the triggering mechanism's configuration, while implementing and testing the web service that will integrate with Go-Lab. After the first prototype is ready, we will design an iterative set of tests to reach real users and collect usability data to improve it in several aspects. After a few iterations of usability testing, the next step will be to design an experiment with real classrooms and collect data to answer our research questions. One of our concerns in our pilot study will be to model the agent's persona, while investigating ethical issues that can arise while the agent is talking with students. That will include a literature search on the main aspects currently discussed about conversational agents' use in education.

REFERENCES

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. https://doi.org/10.1111/jcal.12337

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., Dziabenko, O., Koslowsky, J., Korventausta, M., Law, E., Pedaste, M., Tasiopoulou, E., Vidal, G., & Zacharia, Z. C. (2021). Understanding teacher design practices for digital inquiry–based science learning: the case of Go-Lab. *Educational Technology Research and Development 2021 69:2*, *69*(2), 417–444. https://doi.org/10.1007/S11423-020-09904-Z

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2012). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Eshuis, E. H., ter Vrugte, J., Anjewierden, A., Bollen, L., Sikken, J., & de Jong, T. (2019). Improving the quality of vocational students' collaboration and knowledge acquisition through instruction and joint reflection. *International Journal of Computer-Supported Collaborative Learning*, *14*(1), 53–76. https://doi.org/10.1007/s11412-019-09296-0

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, *20*(2), 177–185. https://doi.org/10.1016/j.knosys.2006.11.014

Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, *4*(1), 21–34. https://doi.org/10.1109/TLT.2010.41

Michos, K., Asensio-Pérez, J. I., Dimitriadis, Y., García-Sastre, S., Villagrá-Sobrino, S., Ortega-Arranz, A., Gómez-Sánchez, E., & Topali, P. (2020). Design of conversational agents for CSCL: Comparing two types of agent intervention strategies in a university classroom. *Addressing Global Challenges and Quality Education. EC-TEL 2020. Lecture Notes in Computer Science*, *12315*, 215–229. https://doi.org/10.1007/978-3-030-57717-9_16

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Tegos, S., Demetriadis, S., Papadopoulos, P. M., & Weinberger, A. (2016). Conversational agents for academically productive talk: a comparison of directed and undirected agent interventions. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 417–440. https://doi.org/10.1007/s11412-016-9246-2

Studying First-Year Student Behavior at an Economics and Business Faculty

Pavani Vemuri
KU Leuven, Belgium
pavani.vemuri@kuleuven.be

[DOCTORAL CONSORTIUM] ABSTRACT: High drop-out rate of first-year bachelors students is a major concern to several higher education institutions, and Learning Analytics (LA) can be a valuable tool to mitigate this problem. This research aims to extract, examine and analyze trace data of first-year courses at an economics and business faculty to provide instructors with insights into student behaviors and learning patterns. These insights can help to predict success and to identify at-risk students. Preliminary visualizations showed that different student behaviors could be associated with different student success levels. Preliminary feature engineering identified key features that are highly relevant across all courses independent of course design and contexts. We are yet to explore potential areas, including: prediction of grades, early prediction, classification, time series analysis to check the consistency of student engagement throughout the semester, featurization of key sequences, longitudinal data across cohorts, and comparative analysis across different study programs.

Keywords: Learning Analytics, Feature Engineering, Regression, Blended Learning

1 MOTIVATION

The Higher Education (HE) sector looks to Learning Analytics (LA) for solutions to issues like student's learning progress, retention, satisfaction, improving teaching quality, and innovations which result in elevated learning experiences, thereby contributing towards institutional performance and ranking. Specifically, drop-out rates and low success rates for first-year students are a concern for many HE institutions. At KU Leuven, it was estimated that around 12% of first-year's drop out after the first-semester exams and up to 35% after the second-semester exams. In response to a university-wide call for initiatives to find feasible solutions to this problem, the "Adaptive Learning Paths for ACtivation and Assessment of Students" (ALPACAS) project was initiated at the Faculty of Economics and Business (FEB) in 2019. The project's central goal is to give more attention to first-year students to ground them in the culture of active learning in their debut year itself. The underlying assumption is that through such grounding in active learning, they will develop self-regulation strategies early on, thereby improving success and retention. As a part of this, LA is to be employed to find insights, informing and allowing instructors to make deliberate decisions about modifying and improving teaching approaches.

While it is essential to know the immediate goals of LA to choose the appropriate methods to be used, it is also vital to identify the factors in focus. For instance, while socioeconomic variables such as "Age, Work, Gender, Stage, Status" may have an impact on student success (Hamoud, Hashim, & Awadh 2018), these variables cannot be addressed by a teacher. Study attitude, on the other hand, maybe affected by instructional design, study counseling and guidance. This dissertation focuses on identifying and analyzing factors that a teacher can address in order to recognize potential problems in an early stage and use appropriate teaching interventions to improve student retention (Palmer, 2013) and success.

2 DATA DESCRIPTION

FEB offers ten programs (taught in English/Dutch languages) at its 4 locations, with approximately 4000+ students enrolling each year. Each program has 4-6 courses taught per semester, summing up to 9-11 courses per year. An instructor or didactic team teaching the course has autonomy over how the course is taught, the pedagogy, use of tools, etc., hence influencing the log data. Overall, there are 110+ courses taught in the first-year bachelor programs. The primary datasets were extracted from the foundation technologies (Gasevic et al., 2019) of an HE institution: collecting log-data from the Toledo-Blackboard Learning Management System (LMS) used at KU Leuven and enrollment and summative scores sourced from the student information systems (SIS).

The datasets across multiple systems were mapped and pseudonymized appropriately in accordance with the university guidelines on data management, privacy, and ethics in using LA. The first extract of the data was transversal in nature and collected over the academic year 2019-2020. When analyzing this data, numerous dimensions are to be considered: a course can be taught by different teachers on different campuses, academic years, and semesters. Data was simplified using a hierarchical structure - students in courses taught by teachers each semester belonging to campus programs. The Toledo extension portal (developed at KU Leuven) of the popular Blackboard LMS allows to collect more than just trace data (navigation between webpages). The interaction of students with the LMS is captured in specific clicks termed Events. Formative assignments and quiz information is captured as Attempts. Efforts are being made to collect video streaming logs, on LMS teacher-student and peer interaction logs and acquire longitudinal datasets to supplement the analysis.

3 RESEARCH QUESTIONS AND RELATED WORK

For this research project, the essential research streams are (1) use of formative tests (van Merrienboer and Kirschener, 2017) and adaptive release learning paths for student motivation and active learning, (2) LA, specifically visualization and prediction of learning outcomes (e.g., Conijn et al., 2016; van Goidsenhoven et al., 2020) to equip instructors with insights to make appropriate teaching interventions. While the former is employed in the implementation of the ALPACAS project, the latter is the primary goal of this analysis.

There are several approaches to deal with trace data. Some predictive LA (PLA) studies treat the outcome variable either as binary – pass/fail (e.g. Malekian et al., 2020; van Goidsenhoven et al., 2020) or grades (Conijin et al., 2016). Compared to online courses and MOOCs with more comprehensive datasets prediction algorithms, PLA studies regarding face-to-face or blended courses are scarce due to high-dimensionality and low sample sizes (van Goidsenhoven et al., 2020). Examining the literature has shown that trace data can indeed give valuable information for predicting success regardless of the context. While the courses in focus vary on a blended scale of entirely face-to-face to completely online, we extend our treatment to all courses in the ALPACAS project.

Some recent studies have examined the probability of a prediction model across different courses (Conijin. et al., 2016; Gasevic et al., 2016) and suggest that generalization of such models is to be handled with care as results vary across courses. Given that the data has a complex hierarchical structure in analyzing the datasets as explained in Section 2, the research questions for this PhD are as follows:

RQ1. Can a predictive model be built with generalized features? Do common key features exist to predict learning outcomes independent of the course contexts, like course design and where the course is on the blended scale?

RQ2. Are there significant factors across the dimensions of the hierarchical structure that affect predicting learning outcomes? Are there differences in learning outcomes of cohorts (1) from different programs attending the same course and (2) from different campuses? Are there differences in the prediction of learning outcomes of cohorts in these contexts?

The data that has been collected, analyzed and visualized should be paired domain knowledge of instructors for interpretation of the results. Ultimately, a LA study should result in actionable items like teaching interventions to help at risk-students, improving courses by updating learning items. In this regard, the research question is:

RQ3. Can student behavior visualization and predictive analyses inform teachers about learning outcomes early in the semester in a useful and actionable way?

4 METHODOLOGY

In order to answer the above research questions, several subsequent studies are planned for the next two years. Most importantly, to find the insights from LA, several steps need to be followed:

1. Data Acquisition: Collecting relevant datasets from LMS server logs is not an easy task. Consulting literature, a list of variables is captured, and after multiple iterations of extracts, verification, validation, and mapping, reasonable datasets were achieved to be analyzed.

2. Data pre-processing: integrating and cleaning data from different sources.

3. Feature selection: A literature study was conducted to list a set of significant features from the literature and map these to the data from the current LMS. Where possible, we will also featurize sequences derived from the mining of trace data. Further, statistical techniques like principal component analysis (PCA) and linear discriminant analysis will be used to identify key features both independent and dependent on course contexts.

4. Model selection and evaluation: different set-ups will be considered: supervised learning for predicting pass/fail, regression to predict actual grades, clustering to identify behavioral patterns, etc. In particular, the aim is to focus on pass/fail prediction (supervised binary classification) and multi-class classification to distinguish further nearly-passed and excellent students within respectively the failed and passed students.

To answer RQ1, firstly features will be extracted from all the courses and key features identified through exploratory factor analysis in order to provide a direction as to how the high dimensionality can be reduced. Next using a two-step approach, courses will be differentiated by their properties (derivable features, delivery mode, assessment approach, content and its delivery choices, success percentage, skill-levels needed etc.), and explore the possibilities of clustering within these groups. Predictive models will be applied to these different groups to find an optimal generalized model for each of them. In a

To answer RQ2, for courses containing different cohorts of students, machine learning based classification (like random forest, logistic regression) or regression methods (like linear regression) will be applied to the complete data for the course and separately to each cohort. The relative importance of different features and accuracy of prediction across cohorts for predicting success will be investigated.

Next, for RQ3, the insights from LA in the form of data visualizations and predictive analyses will be disseminated to the instructors. As stakeholders and domain experts, their feedback on the usefulness of the information will be collected using surveys. Furthermore, we plan to inquire about actions the instructors have taken/will take on the provided information. The responses will be documented for use in future studies.

5 PRELIMINARY RESULTS

We have undertaken two preliminary studies: the first study visualized data of the 2019-2020 cohort to confirm if different student behaviors can be associated with different levels of student success. This study was presented at the CELDA 2021 conference as a short research paper (Vemuri et al., 2021). This initial preliminary analysis of visualization of the events and other insights on student access of content items and how to read the plots was presented to instructors to disseminate knowledge as part of the ALPACAS project. The second study (unpublished) used 2020-2021 cohort data to review what variables and features shown to have predictive value in literature can be extracted from the current LMS logs. In this study we also performed an exploratory analysis to investigate which key features are affected by different course design contexts. PCA was applied to find components that can be constructed from the variables and features and the importance of variables in different courses. The results suggest some key features with high relevance across all courses, such as the count of documents accessed, the time spent on formative assessments, and relative procrastination. Even course-dependent variables such as activity in some particular weeks seem to be common to all courses, even though the exact weeks may vary per course. This exploratory analysis suggests that it is worth investigating generic models that can be applied to all courses.

6 FUTURE AGENDA & PUBLICATION PLANS

This doctoral submission is anticipated to be a collection of published journal articles interspersed by conference papers published at future LAK, EC-TEL, and other conferences. Building on the preliminary studies, which focused on gathering insights from log data, we hope to conduct our analysis as mentioned above and expand and formalize our methodology and replicate it across cohorts (2018-2021). There are many potential areas to be explored which include: prediction of grades (RQ1, RQ2) including early prediction (RQ3), classification (RQ1, RQ2), time series analysis to check the consistency of student engagement throughout the semester (RQ3), featurization of key sequences for prediction (RQ1) and extracting information from the video streaming platform for extra features (RQ1), etc. Ultimately, we hope that future data extractions by (1) instructors at KU Leuven who wish to use LA in their courses have easily accessible and usable scripts to gather insights and (2) researchers for further subsequent studies at KU Leuven or other universities; can use our work to extend it to their own blended learning contexts.

REFERENCES

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*,*10*(1), 17–29.

Gasevic, D., Tsai, Y., Dawson, S., Pardo, A. (2019). How do we start? An approach to learning analytics adoption in higher education. *The International Journal of Information and Learning Technology*, 36(4), pp. 342-353.

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

Hamoud, A. K., Hashim, A. S., Awadh, W. A. (2018). Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. *Intl. Journal of Interactive Multimedia and Artificial Intelligence*, 5(2), pp. 26–31.

Malekian, D., Bailey, J., & Kennedy, G. (2020). Prediction of students' assessment readiness in online learning environments: The sequence matters. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 382–391.

Palmer S. (2013). Modelling engineering student academic performance using academic analytics. *Int J Eng Educ* 29(1), pp. 132–138.

van Goidsenhoven, S., Bogdanova, D., Deeva, G., Broucke, S. vanden, De Weerdt, J., & Snoeck, M. (2020). Predicting student success in a blended learning environment. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 17–25.

van Merriënboer, J.J.G., & Kirschner, P.A. (2017). *Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design* (3rd ed.). Routledge.

Vemuri P., Snoeck, M., Poelmans, S. (2021). Adaptive release learning paths to motivate active learning and engagement in students. *Proceedings of the 18th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2021),* pp. 287-290.

The 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior

Brendan Flanagan[1*], Atsushi Shimada[2], Fumiya Okubo[2], Huiyong Li[1], Rwitajit Majumdar[1], Hiroaki Ogata[1]

[1] Kyoto University, Japan

[2] Kyushu University, Japan

[*] 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 on secondary school reading behavior. As the COVID-19 pandemic has bought about sudden change in learning environments around the world, participants of this year's workshop will be given the unique opportunity to analyze the changes from onsite classes in 2019 and online classes in 2020 in the same education institution. 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.

Keywords: Student Performance Prediction, Data Challenge, Reading Behavior, Onsite Learning, Online Learning

1 WORKSHOP BACKGROUND

Digital learning materials especially digital textbooks are a core part of modern education, and the adoption of digital textbooks in education is increasing. Digital textbooks and e-books are being introduced into education at the government level in a number of countries in Asia (Ogata et al., 2015). This has prompted research into not only the use of such materials within the classroom, but also the collection and analysis of event data collected from the systems that are used for support and distribution (Flanagan et al., 2018; Ogata et al., 2017; Ogata et al., 2015). In addition to its advantages on students' learning, digital text readers are capable of recording interactions regarding students' reading behaviors. As the materials are read by students using the system, the action events are recorded, such as: flipping to the next or previous page, jumping to different pages, memos, comments, bookmarks, and drawing markers to indicate parts of the learning materials that learners think are important or find difficult.

Despite the increase in use, research analyzing students' interaction with digital textbooks is still limited. Recent review study (Peña-Ayala et al., 2014) revealed that almost half of the papers in Learning Analytics (LA) and Educational Data Mining (EDM) fields are using data from Intelligent Tutoring Systems (ITS) or Learning Management Systems (LMS). Previous research into the reading behavior of students has been used in review patterns, visualizing class preparation, behavior change detection, and investigating the self-regulation of learners (Yin et al., 2015; Ogata et al., 2017; Shimada et al., 2018; Yamada et al., 2017). The analysis of reading behavior can be used to inform the revision of learning materials based on previous use, predict at-risk students that may require intervention from a teacher, and identify learning strategies that are less effective and provide scaffolding to inform and encourage more effective strategies. The digital learning material reader can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

The main motivation of this workshop is to foster research into the analysis of students' interaction with digital textbooks, and find new ways in which it can be used to inform and provide meaningful feedback to stakeholders, such as: teachers, students and researchers. This proposal builds upon previous workshops that have focused on student performance prediction based on reading behavior. In previous years at LAK and other international conferences, there have been workshops that have offered open ended data challenges to analyze e-book reading logs and predict the final grade score of learners (Flanagan et al., 2018; Flanagan et al., 2019; Flanagan et al., 2020; Flanagan et al., 2021), with 16, 14, 17, and 12 participants respectively. However, to-date the data challenges have targeted onsite classes in higher education and secondary school settings.

This year we will offer participants a new challenge that focuses on the changes that have been bought about by the COVID-19 pandemic by exploring reading behavior from pre-pandemic onsite classes in 2019 and online classes that were implemented during 2020 (Majumdar et al., 2021). The dataset 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 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.

This year we will provide two datasets: a large labeled training dataset and a smaller test dataset will be distributed. The learner's performance score for the test dataset will be withheld, and participants can upload their scores to the workshop website to check the results of the evaluation periodically. A leaderboard will be provided with the best evaluation score that each participant has achieved to encourage competition between teams. Final data challenge results of prediction models will be confirmed by submission of prediction models for formal evaluation.

2 OBJECTIVES

While we welcome research questions from all participants, and we expect to emphasize the following topic which the organizers feel attention should be paid. Low retention and high failure rates are important problems in education (Villagrá-Arnedo et al., 2017). However, studies have shown that timely interventions for at-risk students can be effective in helping change their behaviors (Arnold et al., 2012; Tanes et al., 2011). Therefore, focusing on the early detection of at-risk students is an essential step to changing student's behavior for greater success.

- This broader task may be approached from the following perspectives:
- Student reading behavior self-regulation profiles spanning the entire course
- Preview, in-class, and review reading patterns
- Student engagement analysis; and behavior change detection
- Visualization methods to inform and provide meaningful feedback to stakeholders

3 OVERVIEW

This workshop was held in a mini-track style with a focus on presentations from participant submitted papers that analyze the data provided by the workshop. The dataset of reading behavior taken from 2019 before COVID-19 and 2020 when university education shifted online was analyzed by three papers (Leelaluk et al.; Lopez et al.; Wangoo et al.) examining the differences in offline/online context and the effect on reading behavior and academic performance. An innovative paper from Guan et al. examined using inexpensive webcam-based eye-tracking in combination with log data to analyze reading behavior. Takami et al. investigated estimating the personality traits of learners to support personalization of recommendations and feedback by examining reading behavior data, and Dai et al. proposed and conducted a preliminary experiment for a knowledge map-based recommendation method for quizzes in BookRoll. The proceedings of the workshop can be found on the following website: https://sites.google.com/view/lak22datachallenge.

ACKNOWLEDGEMENTS

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) 20H01722, (Exploratory) 21K19824, (S) 16H06304 and NEDO JPNP20006 and JPNP18013.

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 267-270). ACM.

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*.

Majumdar, R., Flanagan, B., & Ogata, H. (2021). E-book technology facilitating university education during COVID-19: Japanese experience. Canadian Journal of Learning and Technology 47(4), 1-28.

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.

3rd Workshop on Personalising Feedback at Scale: Systematising the design of personalised feedback

Lisa-Angelique Lim
University of Technology Sydney
lisa-angelique.lim@uts.edu.au

Caitlyn Hayward, Ben Hayward, Holly Derry, Rebecca Matz
University of Michigan
{cholma, haywardb, hderry, rlmatz}@umich.edu

Danny Liu
University of Sydney
Danny.liu@sydney.edu.au

Mette Trier Damgaard
Aarhus University
mdamgaard@econ.au.dk

ABSTRACT: After two successful workshops at LAK, in which presenters explored tools used to provide feedback at scale (LAK18) and focused on students (LAK19), this workshop shifts the attention to educators and practitioners, and how they can design personalised feedback in their own contexts. Stories 'from the trenches' with a focus on educators and practitioners is what will drive the workshop. Continuing to bring together scholars and practitioners to find a common ground for showcasing interesting examples of effective feedback, the workshop will showcase what and how data can be used to improve the process and richness of feedback for both learners and educators. Key outcomes will be a better understanding of personalised feedback processes, which will inform principles for good practice in personalised feedback strategy and ultimately, foster student engagement with this feedback.

Keywords: personalised feedback, nudges, feedback literacy, evaluation

1 INTRODUCTION

In recent years, two successful workshops were delivered at LAK, exploring the affordances and limitations of tools developed to support the scaling of personalised feedback, as well as the impact of this feedback on students. In the third instalment of this series, this workshop shifts the attention to educators and practitioners in the feedback process.

Used effectively, feedback can have significant effects on students' achievement, promoting autonomy and self-regulation (Sadler, 2010). Most recently, with the imperative to shift to online learning due to COVID, now more than ever, it has become increasingly important to be able to support diverse cohorts of students who may only be visible through their digital presence. Advances in learning analytics have led to a proliferation of data-driven solutions for delivering personalised feedback to students at scale. Examples of these include dashboards (e.g., Jivet et al., 2018), instructor-mediated feedback systems such as the Student Relationship Engagement System (Liu et al., 2017) and OnTask (Pardo et al., 2018), as well as dedicated feedback portals such as ECoach (Matz et al., 2021).

With regards to personalised feedback, two further emerging trends may be observed. The first is how student feedback literacy is demonstrated in personalised feedback (e.g. Lim et al., 2021; Tsai et al., 2021). Student feedback literacy can be defined in terms of four processes: appreciating feedback, making judgments, managing affect, and taking action (Carless & Boud, 2018). More evidence is needed to show how educators or researchers are designing personalised feedback in a way that fosters student feedback literacy. The second trend is in the use of nudges in educational settings (e.g. Damgaard & Nielsen, 2018). Within learning analytics, research has been conducted on the use of data-informed nudges as personalised feedback. The intent of such nudges is to promote student engagement with discrete learning activities that lead to successful performance in the course. Data-informed nudges are personalised to students' performance or activity data are similar to process-level feedback, which is also aimed at improving students' learning strategies (Hattie & Timperley, 2007). Data-informed nudges can therefore be considered as a more granular form of process-level feedback. Despite the perceived potential of data-informed nudges, however, the results of data-informed nudge interventions on student learning are very mixed (e.g., Blumenstein et al., 2019; Nikolayeva et al., 2020; O'Connell & Lang, 2018).

One reason for the mixed results around the impact of personalised feedback — indeed, of feedback in general — relates to differences in the way students engage with feedback. Winstone et al. (2017) proposed four groups of factors that can influence students' engagement with feedback: characteristics and behaviour of the receiver, characteristics and behaviour of the sender, characteristics of the message, and characteristics of the context. To date, there has been little research into these characteristics in relation to personalised feedback. Having a body of evidence to illustrate the characteristics of effective feedback processes is important, to inform principles for good practice in automated feedback strategy and ultimately, to promote student engagement with this feedback.

2 SCOPE OF THE WORKSHOP

This workshop brings together scholars and practitioners to explore examples of how educators and system developers can co-design learning analytics feedback processes. The workshop has three primary goals:

1. Reflect on the current adequacy of our multidisciplinary conceptual foundations for practitioners and researchers in LA for effective learning analytics feedback practices in HE;
2. Share approaches for scaling personalised feedback with particular attention to how data are used to inform personalisation, as well as the processes for personalised, dialogic feedback (e.g., timing, content, modality);
3. Promote reflection on both pedagogical and technological approaches to improve feedback practices targeted at the improvement of student learning and their ability to self-regulate learning.

3 ORGANISATION DETAILS

This will be a 3-hour workshop with mixed participation (including selected presenters and interested delegates). Discussions will cover a range of relevant issues including: tools or approaches to personalise feedback/improve message design or feedback processes; implementation processes (e.g. infrastructural, staff skillset & capacity, etc.); challenges and successes (as well as failures); stakeholder engagement, buy-in, and evaluation methodologies & ethics.

3.1 Who is this workshop for?

As a relatively non-technical workshop, this is well suited for a broad audience, including those who wish to understand and apply principles of 'good' data-driven feedback for learning (including Educators/teachers and researchers, Technologists and educational developers, Learning scientists and data scientists/analysts and Academic managers). Given the explicit multidisciplinary nature of the workshop we expect that it will provide an opportunity to discuss and share innovations, impact on learning, and explore future directions in the application of learning analytics for effective feedback that builds students' feedback literacy and ultimately effective self-regulation of learning.

3.2 Proposed workshop activities

After a brief introduction and conceptualisation of the workshop, a series of short presentations will provide a backdrop and provocation to think about current feedback practices. We will discuss both successful and unsuccessful approaches to better understand what works, in which context and for what type of students. Ample opportunity for discussion will be provided to address key themes and issues surfaced during presentations. Similar to the previous two Workshops, a website will be created to provide access to all contributions and presentations as well as a summary from the organisers after the workshop. The workshop will provide an avenue to continue the conversations beyond the session and open opportunities for further collaborations. Participants will also be able to experience personalised feedback for themselves.

4 INTENDED OUTCOMES FOR PARTICIPANTS

We expect a range of presentations that will cover practical, evidence-based approaches to personalising data-driven feedback at scale with a focus on teacher and practitioner perspectives. We expect that participants will obtain a broad perspective of different approaches to using data for personalising feedback or nudges. Participants will also enhance their understanding of the forms of feedback that could improve student learning by discussing cases, issues, and potential solutions to implementing learning analytics-enhanced feedback practices. They will also have an opportunity to connect with researchers and practitioners working to provide personalised feedback or nudges, yielding opportunities for collaboration on approaches and tools across attending institutions. Finally, they will distill principles for effective learning analytics-based feedback and educational nudges. We envisage that a journal special issue may emerge from the workshop.

REFERENCES

Blumenstein, M., Liu, D. Y.-T., Richards, D., Leichtweis, S., & Stephens, J. (2019). Data-informed nudges for student engagement and success. In J. M. Lodge, J. C. Horvath, & L. Corrin (Eds.), *Learning analytics in the classroom: Translating learning analytics for teachers* (pp. 185-207). Routledge.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education,* 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review, 64,* 313-342. https://doi.org/10.1016/j.econedurev.2018.03.008

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *LAK'18: International Conference on Learning Analytics and Knowledge, March 7–9, 2018, Sydney, NSW, Australia.* ACM. https://doi.org/10.1145/3170358.3170421. https://doi.org/10.1145/3170358.3170421

Lim, L.-A., Gasevic, D., Matcha, W., Ahmad Uzir, N. A., & Dawson, S. (2021). Impact of learning analytics feedback on self-regulated learning: Triangulating behavioural logs with students' recall. In *LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 364-374).* ACM. https://doi.org/10.1145/3448139.3448174

Liu, D. Y.-T., Bartimote-Aufflick, K., Pardo, A., & Bridgeman, A. J. (2017). Data-Driven Personalization of Student Learning Support in Higher Education. In *Learning Analytics: Fundaments, Applications, and Trends* (pp. 143–169). Springer, Cham. https://doi.org/10.1007/978-3-319-52977-6_5

Matz, R. L., Schulz, K. W., Hanley, E. N., Derry, H. A., Hayward, B. T., Koester, B. P., Hayward, C., & McKay, T. (2021). Analyzing the efficacy of Ecoach in supporting gateway course success through tailored support. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21), April 12–16, 2021, Irvine, CA, USA.* (pp. 216-225). ACM. https://doi.org/10.1145/3448139.3448160

Nikolayeva, I., Yessad, A., Laforge, B., & Luengo, V. (2020). Does an E-mail Reminder Intervention with Learning Analytics Reduce Procrastination in a Blended University Course? In C. Alario-Hoyos, Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez I., Dennerlein, S.M. (Ed.), *Addressing Global Challenges and Quality Education. EC-TEL 2020. Lecture Notes in Computer Science, vol 12315* (pp. 60-73). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9_5

O'Connell, S. D., & Lang, G. (2018). Can Personalized Nudges Improve Learning in Hybrid Classes? Experimental Evidence From an Introductory Undergraduate Course. *Journal of Research on Technology in Education, 50*(2), 105-119.

Pardo, A., Bartimote-Aufflick, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., . . . Vigentini, L. (2018). OnTask: Delivering Data-Informed Personalized Learning Support Actions. *Journal of Learning Analytics, 5*(3), 235-249. https://doi.org/10.18608/jla.2018.53.15

Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Tsai, Y.-S., Mello, R. F., Jovanović, J., & Gašević, D. (2021). Student appreciation of data-driven feedback: A pilot study on OnTask. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21), April 12–16, 2021, Irvine, CA, USA* (pp. 511-517). ACM. https://doi.org/https://doi.org/10.1145/3448139.3448212

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17-37. https://doi.org/10.1080/00461520.2016.1207538

Towards Trusted Learning Analytics

Srećko Joksimović
Education Futures, University of South Australia
srecko.joksimovic@unisa.edu.au

Djazia Ladjal
Practera
djazia@practera.com

Chen Zhan
STEM, University of South Australia
chen.zhan@unisa.edu.au

Thierry Rakotoarivelo
CSIRO/Data61
thierry.rakotoarivelo@data61.csiro.au

Alison Li
Practera
alison@practera.com

ABSTRACT: Over the last decade, learning analytics has come a long way from measuring shallow proxies of learners' engagement towards capturing fine grained sensor data that allow for measuring complex dimensions such as affect or creativity. The unprecedented amounts of data, both in terms of quantity and quality, brought a plethora of opportunities and raised various ethical and privacy concerns. In that sense, managing and protecting learners' privacy risk reliably and consistently across all datasets in a pragmatic and cost-effective way is of utmost importance for further adoption of learning analytics. In this interactive, half-day workshop session, we will present the “state-of-the-art” work on data privacy risk measurement and reduction. In so doing, we will discuss a variety of methods that provide measurable, policy driven, and provable mitigation mechanisms for maintaining learners' privacy. Participants will also have an opportunity to explore in practice a learning analytics toolbox developed based on the "privacy by design" principles, incorporating some of those novel algorithms. In the final part, participants will be asked to reflect on their perceived usefulness of the proposed solution, as well as to provide an input on their expectations for maintaining learners' privacy.

Keywords: Data privacy, ethics, privacy by design, trusted analytics

1 MOTIVATION AND SIGNIFICANCE

The advancement of learning analytics is directly linked to our ability to collect and process unprecedented amounts of data. The widespread use of technology and connected devices in the learning environment keep increasing the complexity and depth of the data collected (Joksimović et al., 2019). We are currently able to record and store every event and interaction within a learning environment. At the same time, it would be impossible to interpret or analyze these complex datasets without the support of the advanced analytical techniques provided by learning analytics. This includes the ability to store, share and combine the data we have.

Learning analytics, as a field of research and practice, is currently positioned at the intersection of two adverse realities. Recent technological advances allow for the unprecedented data collection possibilities both in terms of quantity and quality (Joksimović et al., 2019). However, ethical and privacy concerns related to the utilization of available data represent a critical issue that needs to be addressed to enable the full proliferation of learning analytics. How pertinent this issue is can be observed through some of the recent examples as well as events that coincide with the emergence of learning analytics. Specifically, the ideas put forward behind the former educational technology company called inBloom Inc are almost perfectly aligned with the goals outlined in learning analytics manifesto. Nevertheless, despite the enormous funding and political support, inBloom failed to gain public trust ending up in a backlash over "inBloom's intended use of student data, surfacing concerns over privacy and protection" (Bulger et al., 2017, p. 4). More recent events with Facebook and Cambridge Analytica[1], numerous data breaches scandals resulting in billions of dollars of damages and fines[2] or failure to use data in ethical way[3], do not contribute to raising trust in data and analytics in general, or learning analytics in particular.

The aim of this workshop is to demonstrate our recent work on developing privacy-preserving learning analytics. This goes beyond just anonymization as we also account for re-identification risk based on the uniqueness of individuals' attributes. We will discuss a variety of methods that provide measurable, policy driven, and provable mitigation mechanisms for maintaining learners' privacy. Participants will also have an opportunity to explore in practice a learning analytics toolbox developed based on the "privacy by design" principles, incorporating some of those novel algorithms.

2 LEARNING ANALYTICS AND DATA PRIVACY

Early after the inception of the field, two important sets of policies emerged in work by Slade and Prinsloo (2013) as well as Pardo and Siemens (2014). Both sets of policies stressed the potential of learning analytics to collect detailed information about how students learn, providing means for changing how learning experiences are conceived and deployed. More importantly, they also highlighted that ethical and privacy issues derived from these new possibilities are not being properly addressed (Pardo & Siemens, 2014; Slade & Prinsloo, 2013). Although, there are certain differences between the two sets of principles, both highlight the importance of student agency, transparency, and accountability. Slade and Prinsloo (2013) further discuss learning analytics as a moral practice, saying that focus should not only be put on what is effective, but on supporting decisions about what is appropriate and morally necessary. The goal should be understanding, not measuring. This is also reflected in Pardo and Siemens notion of accountability and assessment, where assessment has been viewed as the responsibility of the institution to constantly evaluate, review, and refine the data collection, security, transparency, and accountability. To a great extent, these principles have been addressed in the work that followed.

---

[1] https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

[2] https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html

[3] https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

The Open University UK was, arguably, the first institution to develop a policy on ethical use of student data for learning analytics (Slade, 2016). The policy aimed at setting up a framework for ethical use of student data at the Open University, “in order to shape the student support provided” (Slade, 2016, p. 6). Core organizational principles, responsibility, transparency, and trust, as well as students’ involvement as active agents are some of the most prominent dimensions identified within the proposed principles. In addition to the principles discussed in the Open University Policy, JISC released a code of practice for learning analytics where they further introduced the importance of data validity as being essential for developing robust and reliable learning analytics models (Bailey & Sclater, 2015). Finally, Drachsler and Greller (2016) introduced the DELICATE framework, arguably the most referenced work in this area. Although based around a similar set of ethical and privacy concerns, the DELICATE checklist further includes the importance of providing technical procedures to guarantee data privacy and a necessity to consider external providers in dealing with learning related data. The notion of data management and security was further extended in LEA's box data protection framework (Steiner et al., 2016).

As more data became available, it was more important to develop various data protection and privacy legislations on various levels, from educational institutions to government bodies. However, having a legislation (or policy) in place, does not necessarily mean the successful operationalization of the proposed privacy measures. Brown and Klein (2020) analysed 151 university policy statements to identify hidden assumptions; silences; and unintended consequences related to those policies. One of the major issues is that the primary goal of the existing frameworks is driven by the intention to protect institutional liability and reduce potential risk, where data are being treated as a static artifact. More importantly, students are being positioned as “informed, agentic partners in their education” (Brown & Klein, 2020, p. 1159). It is unclear, however, how these policies address individual students’ needs. There is also a striking tension between how data systems are being often conceptualized as static, but also enacted as dynamic technologies.

Although all those frameworks have been essential in establishing a strong agenda for ensuring student privacy there was a need identified to extend towards more practical solutions that would allow for the operationalization of existing policies. To progress more demonstrable processes Hoel and colleagues (2017) provide an extensive list of requirements that should represent essential components of any learning analytics architecture. They also discuss privacy by design, as an essential feature of any learning analytics solution. Two such architectures emerged. Specifically, Gardner and colleagues (2018) introduced MOOC Replication Framework (MORF), a platform-as-a-service architecture for large scale computational research (e.g., predictive modeling and production rule analysis), while protecting student privacy. Whereas Hillman and Ganesh (2019), on the other hand, proposed a system (Kratos) that provides students and schools an “immutable log along with comprehensive access to data that is otherwise scattered across systems and vendors” (Hillman & Ganesh, 2019, p. 5754), placing a particular focus on the aspect of data ownership, data accountability, auditability and transparency.

3 RELEVANCE TO THE LAK THEME

The topic of this workshop is directly linked with the LAK22 conference theme. Acknowledging the importance of privacy, transparency, fairness, and equity as being essential in broadening learning analytics research and practice, LAK conference puts as an imperative on providing means for incorporating privacy by design into the core of learning analytics.

The level of data acquisition and handling raises many ethical issues, security and data privacy concerns (Pardo & Siemens, 2014). It is important that we do not compromise learners' privacy and security to achieve learning analytics goals, as beneficial as they might be to the learners. It is equally important to ensure that we can still derive useful insight and results with learning analytics even when using privacy secured data. This workshop, therefore, aims at contributing to the discourse of developing privacy-driven learning analytics.

4 WORKSHOP ORGANIZATION

Type of event: Interactive workshop session; **Proposed duration:** Half-day.

Type of participation: Mixed participation. The proposed workshop plans to demonstrate various risks related to student privacy and present methods for the assessment and mitigation of potential risks. In the first part, invited experts and researchers who are involved in projects related to the workshop's theme will present their work. In the second part, workshop participants will have an opportunity to work with our privacy-driven learning analytics toolbox, as well as with some of the methods for the privacy risk assessment and mitigation. We will invite participants to bring their own dataset or provide a sample dataset to work with. In the final part of the workshop, participants will have an opportunity to reflect on their experiences with the tools they used in this workshop and share suggestions for improvements.

Proposed activities: Presentations by organizers and invited experts; hands-on activities; and guided discussions around the emerging challenges related to data privacy.

Expected participant numbers: 25-30.

Website: We also plan to integrate all relevant resources and contact information, produced before and during the workshop such as presentation slides and discussion notes, on our project website (https://trustedanalytics.com.au/) to encourage ongoing communication and collaboration after the workshop. We will use our website and other communication channels to coordinate data preparation for the purpose of running the workshop.

Social Media: We will use #TLA22 as the primary hashtag to encourage discussions and communication through Twitter.

REFERENCES

Bailey, P., & Sclater, N. (2015). *Code of practice for learning analytics* (p. 6). Jisc. https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Brown, M., & Klein, C. (2020). Whose Data? Which Rights? Whose Power? A Policy Discourse Analysis of Student Privacy Policy Documents. *The Journal of Higher Education*, *91*(7), 1149–1178. https://doi.org/10.1080/00221546.2020.1770045

Bulger, M., McCormick, P., & Pitcan, M. (2017). *The Legacy of InBloom*. Data & Society. https://datasociety.net/pubs/ecl/InBloom_feb_2017.pdf

Drachsler, H., & Greller, W. (2016). Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. *Proceedings of the Sixth International Conference on Learning Analytics And Knowledge*, 89–98. https://doi.org/10.1145/2883851.2883893

Gardner, J., Brooks, C., Andres, J. M., & Baker, R. (2018). Replicating MOOC Predictive Models at Scale. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. https://doi.org/10.1145/3231644.3231656

Hillman, V., & Ganesh, V. (2019). Kratos: A secure, authenticated and publicly verifiable system for educational data using the blockchain. *2019 IEEE International Conference on Big Data (Big Data)*, 5754–5762. https://doi.org/10.1109/BigData47090.2019.9006190

Hoel, T., Griffiths, D., & Chen, W. (2017). The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems. *Proceedings of the Seventh International Learning Analytics And Knowledge Conference*, 243–252. https://doi.org/10.1145/3027385.3027414

Joksimović, S., Kovanović, V., & Dawson, S. (2019). The Journey of Learning Analytics. *HERDSA Review of Higher Education*, *6*, 27–63.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Slade, S. (2016). *The Open University Ethical use of Student Data for Learning Analytics Policy* (p. 11). https://doi.org/10.13140/RG.2.1.1317.4164

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Steiner, C., Kickmeier, M., & Albert, D. (2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, 66–90. https://doi.org/10.18608/jla.2016.31.5

Writing Analytics for higher-order thinking skills

Antonette Shibani
University of Technology Sydney
antonette.shibani@uts.edu.au

Andrew Gibson
Queensland University of Technology
andrew.gibson@qut.edu.au

Simon Knight
University of Technology Sydney
Simon.Knight@uts.edu.au

Philip H Winne
Simon Fraser University
philip_winne@sfu.ca

Diane Litman
University of Pittsburgh
dlitman@pitt.edu

ABSTRACT: There are a growing number of evaluated writing analytics tools and technologies targeting the improvement of academic writing. As the field grows, there is potential for new writing analytics tools to target formative feedback for higher-order thinking skills. This writing analytics workshop, the sixth in the series at LAK, will explore how writing analytics can potentially support these life skills among learners, and the data, tools, analytics, and pedagogic contexts for such implementations.

Keywords: writing analytics, learning analytics, higher order skills, critical thinking, argumentation, metacognition, reflection, self-regulation, information literacy

WORKSHOP FOCUS

Writing Analytics (WA) has received considerable attention from researchers who aim to advance pedagogical practices using the application of data, tools and analytics on writing. It has evolved as a sub-domain of learning analytics supporting the study of written products and processes in educational contexts through natural language processing (NLP) and other automated text analysis methods (Buckingham Shum et al., 2016). Five previous WA workshops at the Learning Analytics and Knowledge conference have contributed to the growth in the field by defining key areas for attention. These include: An introduction to writing analytics with critical perspectives and community building in LAK (Buckingham Shum et al., 2016), building writing analytics literacy and practitioner capacity (Knight, Allen, Gibson, McNamara, & Buckingham Shum, 2017), a hands-on-training for developing this literacy by understanding technical affordances and aligning them to pedagogical feedback using a socio-technical tool (Shibani, Abel, Gibson, & Knight, 2018), a mapping of the state of the art work in writing analytics along with defining future pathways for the field (Shibani, Liu, Rapp, & Knight, 2019), and a consolidation of writing analytics practice by bringing together writing tool design, writing analytics and writing pedagogy (Rapp, Lang, Shibani, Benetos, & Anson, 2020).

There are an increasing number of evaluations of writing analytics tools and practices that help improve academic writing (Allen, Jacovina, & McNamara, 2015; Knight et al., 2020; Liu, Li, Xu, & Liu, 2017; Woods, Adamson, Miel, & Mayfield, 2017). A distinctive feature of writing analytics is a focus on formative feedback, in contrast to automated essay scoring systems. There is thus potential for these tools to be used to support equipping students with "higher-order thinking" skills such as critical thinking, decision making, and problem solving, in addition to building their knowledge capacity (Miri, David, & Uri, 2007). These skills can facilitate the transition of students' knowledge and skills into responsible action in society (Zoller, 2000). Writing analytics can support the development of complex skills required by learners to thrive in the changing world. While these tools may draw on a range of emerging approaches, there is currently limited research in this area.

Few notable WA applications include the development of learner metacognition using reflective writing analytics (Gibson et al., 2017), a revision assistant for argumentative writing (Zhang, Hwa, Litman, & Hashemi, 2016), knowledge transformation in argumentative writing (Raković, Winne, Marzouk, & Chang, 2021) and self-regulated learning (Winne, 2001; Wollny, Schneider, Rittberger, & Drachsler). Key skills such as critical thinking, argumentation, data and information literacy along with creativity, metacognition, self-regulation, transdisciplinary thinking and lifelong learning are essential to develop among students for their participation in society and sustainable personal, civic, and professional decision making (Buckingham Shum & Crick, 2016). The current workshop will hence focus on how writing analytics can support these higher order thinking skills. Outcomes include an understanding of the current research landscape in writing analytics with respect to higher order thinking skills, and a co-created mapping of how data, tools and analytics can further support the development of these skills in the future.

WORKSHOP ORGANISATION

1 Workshop format and schedule

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:

I. Introductions: Introductions of workshop organizers and participants, and a background to the focus of the workshop.
II. Short Presentations: Authors doing related work present their writing analytics tool or technology, the data collected by the tool, analysis of writing data and how it builds higher order thinking among learners.
III. Round-table discussion: Participants will move around specific topics of interest assigned to tables. These tables will enable more detailed discussion and collaboration based on selected topics from the presentations. Groups will also collaborate on an activity identifying potential writing analytics proxies, lower order natural NLP indices and pedagogical designs that support the development of higher-order constructs.

IV. 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.

V. Concluding remarks and community engagement: Closing remarks on the workshop will be made with future steps. Community engagement among the Special Interest Group on Writing Analytics members will be encouraged.

2 Participation and Dissemination

We will invite specific authors actively researching in this area to present their work at the workshop. Participation will be open to any delegate in addition to presenting authors. A website will be setup for the workshop that will archive the event and disseminate notes to participants

REFERENCES

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Computer-based writing instruction. *Handbook of writing research*, 316-329.

Buckingham Shum, S., & Crick, R. D. (2016). Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics, 3*(2), 6-21.

Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). *Critical perspectives on writing analytics*. Paper presented at the Workshop at the Sixth International Conference on Learning Analytics & Knowledge.

Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). Reflective writing analytics for actionable feedback.

Knight, S., Allen, L., Gibson, A., McNamara, D., & Buckingham Shum, S. (2017). *Writing analytics literacy: bridging from research to practice.* Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference.

Knight, S., Shibani, A., Abel, S., Gibson, A., Ryan, P., Sutton, N., . . . Buckingham Shum, S. (2020). AcaWriter: A Learning Analytics Tool for Formative Feedback on Academic Writing. *Journal of Writing Research*.

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated Essay Feedback Generation and Its Impact on Revision. *IEEE Transactions on Learning Technologies, 10*(4), 502-513.

Miri, B., David, B.-C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in science education, 37*(4), 353-369.

Raković, M., Winne, P. H., Marzouk, Z., & Chang, D. (2021). Automatic identification of knowledge - transforming content in argument essays developed from multiple sources. *Journal of computer assisted learning*.

Rapp, C., Lang, S., Shibani, A., Benetos, K., & Anson, C. (2020). *Bringing together writing tool design, writing analytics and writing pedagogy.* Paper presented at the 10th International Conference on Learning Analytics & Knowledge (LAK20), Frankfurt, Germany, 23-27 March 2020.

Shibani, A., Abel, S., Gibson, A., & Knight, S. (2018). *Turning the TAP on Writing Analytics.* Paper presented at the Learning Analytics and Knowledge, Sydney.

Shibani, A., Liu, M., Rapp, C., & Knight, S. (2019). *Advances in Writing Analytics: Mapping the state of the field*. Paper presented at the Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19), Tempe, Arizona.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. *Self-regulated learning and academic achievement: Theoretical perspectives, 2*, 153-189.

Wollny, S., Schneider, J., Rittberger, M., & Drachsler, H. *Am I planning smart?–Analyzing student goals.* Paper presented at the Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference (LAK'19).

Woods, B., Adamson, D., Miel, S., & Mayfield, E. (2017). *Formative Essay Feedback Using Predictive Scoring Models.* Paper presented at the Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Zhang, F., Hwa, R., Litman, D. J., & Hashemi, H. B. (2016). *ArgRewrite: A Web-based Revision Assistant for Argumentative Writings.* Paper presented at the NAACL-HLT 2016 (Demonstrations), San Diego, California.

Zoller, U. (2000). Teaching tomorrow's college science courses--are we getting it right? *Journal of College Science Teaching, 29*(6), 409.

Writing for Publication: Engaging Your Audience

Yi-Shan Tsai
Monash University
yi-shan.tsai@monash.edu

Melanie Peffer
University of Colorado Boulder
Melanie.Peffer@colorado.edu

Antonette Shibani
University of Technology Sydney
Antonette.Shibani@uts.edu.au

Isabel Hilliger
Pontificia Universidad Católica de Chile
ihillige@ing.puc.cl

Bodong Chen
University of Minnesota
chenbd@umn.edu

Yizhou Fan
University of Edinburgh
yizhou.fan@ed.ac.uk

Rogers Kaliisa
University of Oslo
rogers.kaliisa@iped.uio.no

Nia Dowell
University of California, Irvine
dowelln@uci.edu

Simon Knight
University of Technology Sydney
Simon.Knight@uts.edu.au

ABSTRACT: Over the past two decades, learning analytics has become an established field of research and practice, with a significant increase in the number of academic publications and related results in freely available web search engines. However, professional writing, to either an academic or general audience, can become an overwhelming task, particularly in an interdisciplinary field at the intersection of education, computer science, psychology, and other fields. In this participatory workshop, we invite postgraduate students and early to mid-career researchers to explore the differences among different publication venues in the field, and participate in practical exercises to strengthen their academic writing skills. Through this workshop, we expect to enhance the effectiveness of scientific communications on learning analytics, thereby expanding impact and increase understanding and use of learning analytics.

Keywords: learning analytics, professional writing, science communication

1 WORKSHOP OBJECTIVES

Rigour is the foundation for high quality work that leads to positive impact (knight, Wise & Ochoa, 2019). However, quality research alone may not guarantee publication. Effective communication is crucial in order to engage the audience, particularly for the sustainable impact of scientific research and practice (Iyengar & Massey, 2019). For example, recent calls emphasise the need for meaningful engagement between scientists and the public to improve decision making around public well-being and policy issues (Rose, Markowitz, & Brossard, 2020). Advances in technology have also opened the door to a variety of new options to communicating science (Bubela, et al., 2009), from the ease of developing Blogs (Colson, 2011; Kouper, 2010) to Twitter chats (Young, Tully, Dalrymple, 2017). However, not all published works gain the same visibility. To those new to academia, the process of writing for publication (for academic or general audiences) can appear obscure. To those experienced in academia, writing for publication can still be frustrating at times. Moreover, writing for an interdisciplinary audience raises its own unique challenges. The Society for Learning Analytics Research (SoLAR) has a strategic goal to expand impact and increase understanding and use of learning analytics (https://www.solaresearch.org/about/). In response to this goal, the proposed workshop aims to provide training with regard to writing for publication on learning analytics. Some of the questions that the workshop seeks to address include: How to choose the right venue for publication? How to write for academic and public audiences? How to engage your audience? What will reviewers and editors look for? How to make your work reach the right audience? To answer these questions, the workshop will engage participants in exploring publication opportunities in learning analytics and use SoLAR's flagship publications, including Journal of Learning Analytics (https://www.solaresearch.org/publications/journal/), The International Learning Analytics and Knowledge Conference (https://www.solaresearch.org/publications/conference-proceedings/), and NEXUS (https://www.solaresearch.org/publications/nexus/), to discuss different venues for publication.

2 WORKSHOP ORGANISATION

2.1 Schedule and Activities

The workshop is structured to cover four main topics:

1. Publication venue
2. Audience
3. Working with reviewers and/or editors
4. Digital footprint

Starting with Topic 1, we will explore the differences between writing for academic and general audiences and what to consider when choosing a place to publish work. In Topic 2, we will explore ways to make writing readable, engaging, and applicable to different audiences. We will also cover practical considerations when working in a group. In Topic 3, we will examine quality criteria of publication, specifically in the context of learning analytics, e.g., rigor of research, innovation, methodology, reporting, and contribution. This topic will also include ways to communicate with reviewers and editors, e.g., writing rebuttals and pitching ideas. Finally, Topic 4 serves to help participants better promote and disseminate their work by considering keywords, choice of title, search engine optimization (SEO), and social media.

For each of these topics, we will include information from both the perspectives of writing for an academic audience and writing for the general public. For example, for publication venue, we will discuss what to consider when selecting a journal or conference publication, and introduce the wide variety of options for writing for a larger audience (e.g., *The Conversation,* which serves both academic and general audiences), writing Blogs (e.g., NEXUS), and writing for local newspapers. In the audience section, we will discuss how to write for your audience, and key differences in what scientists versus general readers look for (e.g., 'news you can use' spin on new research). In the third section, we will discuss how to benefit from the peer review process for academic publication, and how to effectively work to pitch an article to an editor of a mainstream publication and build rapport. Finally, in our fourth and final section we will discuss how to build effective digital footprint to promote written works, whether for academic or general publications. This will include discussion of topics such as AltMetrics (Bornmann, 2014), search engine optimisation (SEO), and creating effective titles or headlines.

The workshop will last for half a day and each of the topics will take up 30 to 60 minutes. Sessions will be run by workshop organisers and additional experts will be invited for mini keynotes. We will use Zoom to facilitate the workshop. Participants will be expected to join interactive activities, such as practice writing and discussion in groups.

2.2 Participants

The targeted participants are postgraduate students and early to mid-career researchers, though anyone interested in the topic will be welcome. Information and resources related to the workshop will be made available on the SoLAR website as part of the society's open resources (https://www.solaresearch.org/core/). We expect a maximum of 40 participants.

2.3 Outcomes

The workshop will equip participants with knowledge and best practices to effectively communicate scientific work in writing. It will help build an understanding of the publication process and strategies to captivate the audience's attention. These should result in an improved writing practice and better reach of scientific work in the learning analytics community.

REFERENCES

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. Journal of informetrics, 8(4), 895-903.

Bubela, T., Nisbet, M. C., Borchelt, R., Brunger, F., Critchley, C., Einsiedel, E., ... & Caulfield, T. (2009). Science communication reconsidered. Nature Biotechnology, 27(6), 514-518.

Colson, V. (2011). Science blogs as competing channels for the dissemination of science news. Journalism, 12(7), 889-902.

Kouper, I. (2010). Science blogs and public engagement with science: Practices, challenges, and opportunities. Journal of Science Communication, 9(1), A02.

Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. Proceedings of the National Academy of Sciences, 116(16), 7656-7661.

Knight, S., Wise, A., & Ochoa, X. (2019). Fostering an impactful field of learning analytics. Journal of Learning Analytics.

Rose, K. M., Markowitz, E. M., & Brossard, D. (2020). Scientists' incentives and attitudes toward public communication. Proceedings of the National Academy of Sciences, 117(3), 1274-1276.

Young, R., Tully, M., & Dalrymple, K. E. (2018). Engagement: use of Twitter chats to construct nominal participatory spaces during health crises. Information, Communication & Society, 21(4), 499-515.

CROSSMMLA & SLE Workshop: Learning Analytics for Smart Learning Environments Crossing Physical and Virtual Learning Spaces

Daniel Spikol
University of Copenhagen
ds@di.ku.dk

Miguel L. Bote Lorenzo
University of Valladolid
migbot@tel.uva.es

Multu Cukurova
University College London
M.Cukurova@ucl.ac.uk

Elise Lavoue
University of Lyon
lise.lavoue@univ-lyon3.fr

Michail Giannakos
Norwegian University of Science and Technology
michail.giannakos@idi.ntnu.no

Xavier Ochoa
New York University
xavier.ochoa@nyu.edu

Daniele Di Mitri
Leibniz Institute for Research and Information in Education
dimitri@dipf.de

Davinia Hernandez-Leo
Pompeu Fabra University
davinia.hernandez@upf.edu

ABSTRACT: The workshop proposal for CrossMMLA focused on Smart Learning Environments to collect and analyse multimodal data across the physical and the virtual. The rapidly changing nature of education has begun to embrace hybrid learning, which changes the landscape of how MMLA is used. This year's CROSSMMLA will focus on SLEs supported by MMLA that support diverse learning scenarios. The Workshop proposes an asynchronous format that includes pre-recorded video demonstrations and position papers for discussion, followed by a half-day virtual meeting at LAK'22.

Keywords: Smart Learning Environments, Multimodal Learning Analytics, Hybrid Learning, Sensor-based Analytics

1 INTRODUCTION

Smart Learning Environments (SLEs) use adaptive technologies and are designed to include innovative features and capabilities that improve understanding and performance for learners and teachers. SLEs provide learners with adequate support at the right time and place based on individual learning needs, which are determined by analysing their learning behaviours, performance and contexts (Spector, 2014). Given that SLEs afford various features and attributes like adaptation, flexibility, thoughtfulness there are clear connections to research with SLE and Multimodal Learning Analytics (MMLA) that have brought together diverse fields that combine educational, computational, psychological, and related research into how people learn and how this can support complex processes with technology. SOLAR's Special Interest Group on Multimodal Learning Analytics Across Spaces (CrossMMLA SIG) aims to promote research that considers the challenges of making sense of complex educational data that involve multiple interaction modalities, people, and smart learning spaces. Understanding and improving learning traces from the real world and smart environments require new technology, learning, and design approaches. For LAK'22, CROSSMMLA will focus on SLEs that can support a myriad of learning scenarios that connect formal and informal learning.

With this aim, SLEs may collect data about learners' and educators' actions and interactions related to their formal and informal learning activities, with the challenge to create a continuum between these spaces. They also may collect different aspects of education in their contexts (e.g., learners' emotions, physiological states, ambient temperature) from sources such as clicking and scrolling behaviour in the Learning Management Systems, handheld devices, computers, cameras, microphones, wearables, and environmental sensors. These data can then be transformed, fused and analysed using different computational and visualisation techniques to obtain actionable information that can trigger a wide range of interventions to promote better learning in formal and informal contexts and in addition to maintaining learners' motivation and engagement across hybrid spaces (Lavoué, Ju, Hallifax, & Serna, 2021).

1.1 Background

The rapid global shift to virtual learning challenges the research and practice for this field and educational practices overall. SLEs and MMLA need to develop different theories about analysing human behaviours during diverse learning processes across spaces and educational materials to create useful tools that could augment the capabilities of learners and instructors. Furthermore, these tools and practices must be designed and implemented ethically and sustainably to provide value and equity for all learners

LA and MMLA combine the power of affordable sensor technologies and advances in machine learning to analyse, analyse and make inferences from the collected sensor data (Blikstein, 2013; Ochoa, 2017; Worsley, 2012) This technology acts as a virtual observer and analyst of learning activities across multiple contexts between stakeholders, devices, and resources. Recent work explores how real-time and automatic analysis of video and audio can support learning by automating the analysis of these activities through the development of innovative tools (Chan, Ochoa, & Clarke, 2020; Chejara, 2020; Kasparova, Celiktutan, & Cukurova, 2020). Martinez-Maldonado and colleagues (2020) are conducting work to make the stream of multimodal data into meaningful layers that explain critical insights to teachers and students.

LA research is expected to provide data insights into the aforementioned actionable information required by SLEs (Papamitsiou & Economides, 2016). Important research efforts are being carried out with this goal (Tabuenca & et al., 2021). However, researchers have also realised that multimodal data collection in the learning sciences demands new and powerful methodological and analytical techniques and technologies that can/should be theory-driven. Identifying and merging these opportunities has been a focal point of MMLA, CrossMMLA and LA4SLE workshops that have been conducted at research conferences over the past ten years (Prieto, 2017; Scherer, Worsley, & Morency, 2012).

1.2 Aims for the Workshop

The Workshop will serve as a forum to exchange ideas on how we as a community can use our knowledge and experiences from CrossMMLA to design new tools to analyse evidence from multimodal data and SLEs. In addition, we are interested in how we can extract meaning from these increasingly fluid and complex data coming from different kinds of transformative learning situations and how best to provide feedback on the results of these analyses to support those learning processes positively.

The dimensions and contexts of SLEs that utilise MMLA are complex and layered and provide researchers with multiple challenges (Cukurova, Giannakos, & Martinez - Maldonado, 2020). In addition, education is embracing hybrid approaches for learning and teaching, and the CrossMMLA community needs to find ways to research, design, and further develop tools and methods to investigate these new landscapes.

The Workshop aims to discuss the main issues to further research, development, and implementation with SLE and how these overlap with LA. Plus, how SLE research and practice can utilise the latest advances in LA. Contributions from the following topics are welcome (but not limited):

- Ethics-driven LA for SLEs
- Human-centred MMLA and SLEs
- Personalisation for data-driven interventions
- New IoT scenarios for MMLA and SLEs
- Multimodal data approaches for SLEs
- Theoretical foundations for the design and evaluation of SLE considering learners' needs and motivation
- Visualisation techniques for learners and teachers
- Multimodal interfaces to provide feedback to learners and teachers

2 WORKSHOP OBJECTIVES

Apart from being a venue to present the current prototypes and ideas in the field of Multimodal Learning Analytics across Spaces, this Workshop has three main objectives:

Interchange of Technical Knowledge: CrossMMLA, including SLEs, requires an extensive technical infrastructure to be developed (recording equipment, synchronisation techniques, feature extraction algorithms, etc.). While not attractive from the research perspective, implementing and testing this infrastructure is time-consuming. The community's information interchange on the suitable sensor or library accelerates research and practical projects.

Setting Ethical Standards: Discussions on ethics are the ideal moment in which projects are confronted with constructive criticism about the possible impact on their users' cognitive, meta-cognitive, emotional, and social well-being. This Workshop's objective is to serve as the opportunity to, as a community, discuss our ethical standards for research and practice, especially considering the changing regulations like the proposed European Artificial Intelligence Act.

Fostering New Research Initiatives: To continue to connect researchers, many of the current SLE and MMLA projects started as discussions in previous workshops. CrossMMLA and SLE communities have relied on cooperation and collaboration to tackle new and ambitious projects as a small but open community. We expect that this workshop serves as a unique opportunity to think and discuss the future of this sub-field that needs to include SLEs.

As the results of previous CrossMMLA workshops have corroborated it, we expect that this new meeting at LAK'22 will serve as another step towards the use of diverse and relevant traces of learning activities, collected in the different SLEs in which the learners transit, to understand better and improve their learning experience.

REFERENCES

Blikstein, Paulo. (2013, 2013-01-01). *Multimodal learning analytics.* Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13.

Chan, Man Ching Esther, Ochoa, Xavier, & Clarke, David. (2020). Multimodal learning analytics in a laboratory classroom. In *Machine learning paradigms* (pp. 131-156): Springer.

Chejara, Pankaj and Prieto Luis P. and Ruiz-Calleja Adolfo and Rodrguez-Triana Mara Jess and Shankar Shashi Kant and Kasepalu Reet. (2020). *Quantifying collaboration quality in face-to-face classroom settings using mmla*.

Cukurova, Mutlu, Giannakos, Michail, & Martinez ‐ Maldonado, Roberto. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology, 51*(5), 1441-1449. doi:10.1111/bjet.13015

Kasparova, Angelika, Celiktutan, Oya, & Cukurova, Mutlu. (2020). *Inferring student engagement in collaborative problem solving from visual cues.* Paper presented at the Companion Publication of the 2020 International Conference on Multimodal Interaction.

Lavoué, Elise, Ju, Qinjie, Hallifax, Stuart, & Serna, Audrey. (2021). Analyzing the relationships between learners' motivation and observable engaged behaviors in a gamified learning environment. *International Journal of Human-Computer Studies*.

Martinez-Maldonado, Roberto and Echeverria Vanessa and Fernandez Nieto Gloria and Buckingham Shum Simon. (2020). *From data to insights: A layered storytelling approach for multimodal learning analytics*.

Ochoa, Xavier and Lang Awdg Charles and Siemens George. (2017). Multimodal learning analytics. *The handbook of learning analytics, 1*, 129--141.

Papamitsiou, Zacharoula, & Economides, Anastasios A. (2016). Learning analytics for smart learning environments: A meta-analysis of empirical research results from 2009 to 2015. *Learning, design, and technology: An international compendium of theory, research, practice, and policy*, 1-23.

Prieto, Luis Pablo and Martnez-Maldonado Roberto and Spikol Daniel and Hernndez-Leo Davinia and Rodrguez-Triana Mara Jess and Ochoa Xavier. (2017). *Joint proceedings of the sixth multimodal learning analytics (MMLA) workshop and the second cross-LAK workshop*.

Scherer, Stefan, Worsley, Marcelo, & Morency, Louis-Philippe. (2012). *1st international workshop on multimodal learning analytics.* Paper presented at the Proceedings of the 14th ACM international conference on Multimodal interaction.

Spector, Jonathan Michael. (2014). Conceptualizing the emerging field of smart learning environments. *Smart learning environments, 1*(1), 1--10.

Tabuenca, Bernardo and Serrano-Iglesias Sergio and Carruana-Martin Adrian and Villa-Torrano Cristina and Dimitriadis Yannis A. and Asensio-Perez Juan I. and Alario-Hoyos Carlos and Gomez-Sanchez Eduardo and Bote-Lorenzo Miguel L. and Martinez-Mones Alejandra, & et al. (2021). Affordances and Core Functions of Smart Learning Environments: A Systematic Literature Review. *IEEE Transactions on Learning Technologies*.

Worsley, Marcelo. (2012). *Multimodal learning analytics: enabling the future of learning through multimodal data analysis and interfaces*.

Networks and learning analytics: Addressing Educational Challenges

Mohammed Saqr
University of Eastern Finland
mohammed.saqr@uef.fi

Sonsoles López-Pernas
Universidad Politécnica de Madrid
sonsoles.lopez.pernas@upm.es

Ángel Hernández-García
Universidad Politécnica de Madrid
angel.hernandez@upm.es

Miguel Ángel Conde
Universidad de León
mcong@unileon.es

Oleksandra Poquet
University of Paris / University of South Australia
sasha.poquet@learningplanetinstitute.org

ABSTRACT: Network Analysis is an established method in learning analytics research. Network Analysis has been used to analyze learners' interactions, to inform learning design, and to model students' performance. The workshop entitled "Using Network Science in Learning Analytics: Building Bridges towards a Common Agenda", carried out within the LAK2021 conference, resulted in valuable insights and outcomes: guidelines for better reporting, methodological improvements, and discussions of several novel research threads. Traditionally, the focus of the conversation has been on methodological issues of network analysis. This year, we would like to extend the conversation by slightly shifting the focus to what network analysis can do to improve learning and educational opportunities. As such, this new edition of the workshop aims to build on the fruitful achievements of the previous iteration to address new themes, which we refer to as "challenges and opportunities" in relation to practice. This edition of the workshop sought contributions around examples of applications and impact, including those that can help address societal challenges embedded within educational practices and those that foster an open conversation about privacy and ethical implications of network data.

Keywords: Social network analysis, network science, learning analytics, network analysis

Networks are the fundamental building blocks behind several related fields, such as Social Network Analysis, Network Science and Network Analysis. In education, these terms have been loosely and interchangeably employed across several domains of applications. Therefore, we use the term network analysis to refer to any application of network-based techniques in education. Network analysis predates the birth of learning analytics as a field. However, a significant upsurge of research at the intersection of networks and education has been kindled by the growing interest in learning analytics (Kaliisa et al., 2022). Such an upsurge has made network analysis methods gain popularity in learning analytics research. The wealth, ease, and flexibility of network analysis have contributed to a wide range of applications across several domains (Borgatti et al., 2009). Early applications of network analysis included tools for the analysis of learners' interactions and the patterns derived from those interactions, as well as using SNA to inform learning (Lockyer et al., 2013). Combining network analysis with other quantitative methods has augmented our understanding of collaborative learning interactions and discourse (Poquet et al., 2021) and improved the identification of relevant actors and roles in the learning process (Hernández-García et al., 2015), among others. Recently, a growing number of research articles have looked at the usefulness of network analysis as a method for modelling students' performance (Saqr et al., 2019) and modelling novel interaction platforms, such as instant messaging (Conde et al., 2021), to mention a few.

The rapid development of the field of social network analysis and the closely related fields (e.g., network science, complexity science and psychological networks) has progressed our understanding of the world: from the structure of genes to the spread of diseases and the development of efficient algorithms (Newman, 2018). Nonetheless, the pace of research in learning analytics has not harnessed the full potential of emerging and rapidly-developing network methods (Dado & Bodemer, 2017; Poquet et al., 2021). Several recent systematic reviews, discussion papers and scholars have highlighted some of the challenges in this research area: predominance of descriptive methods, paucity of temporal network analysis, underusage of network inference methods, lag of adoption of modern analytics techniques (e.g., psychological network methods) and sub-optimal reporting (Dado & Bodemer, 2017; Kaliisa et al., 2022; Poquet et al., 2021). The challenges faced by network researchers and educationists can be varied, complex and rapidly changing. However, issues that address methodological rigor and impact are arguably more urgent to address. Rigor, better reporting and replicability allow researchers to build on research findings and advance our understanding of theory and practice (Poquet et al., 2021), while impactful research helps researchers, practitioners and society at large better adopt the methods.

This workshop aims to address new themes, building on the fruitful accomplishments of the previous iterations. The themes emerge as "challenges and opportunities" in relation to practice. The workshop was designed to foster the sharing of perspectives, strengthen the community of scholars working in the network community, and generate new ideas regarding the future of the field. To address the current challenges of SNA in the field of learning analytics, participants were invited to contribute their latest research in one of the following themes:

- How can methodological rigor and reporting be balanced with practical relevance?
- How can researchers address the issues of practical applications and impact? In other words, how can we translate research results from academic and scholarly publications into actionable tools or methods that teachers can use in their teaching practice.
- How can network approaches inform socio-emotional and communication skills?
- How can network approaches contribute to the development of equitable educational practices?
- What are the ethical and privacy applications for using network analysis in learning analytics?

- How to implement successful network analytical applications that analyze systems to help advance impact in learning analytics?
- What interdisciplinary bridges between the field and other closely related fields can contribute to addressing educational challenges via network analytics approaches?
- How can educational researchers use modern network methods (e.g., psychological networks, inference network methods and network science) to advance current theoretical models and frameworks?
- How can education researchers contribute to the development of novel network methods, measures or techniques that considers the contextual factors that are specific to learning and teaching?
- How can network researchers contribute to more reproducible research?

The workshop also welcomed contributions discussing the potentials of novel methods or novel applications of existing methods and reflections on their relevance to practical applications, including but not limited to:

- Recent advances in SNA methods and approaches including new tools, measures, and techniques applied to learning.
- Innovative data collection, analysis, presentation, and visualization methods.
- Temporal networks, and temporal aspects of networks in general.
- Psychological networks.
- Generative network models.
- Software demonstration for analysis of learning networks.

Six submissions were accepted for presentation in the workshop:

1. "Participation and Interaction Patterns for MOOC Learners with Different Achievements: A Collective Attention Network" by Ming Gao and Jingjing Zhang.
2. "Analysis Of Discussion Forum Interactions for Different Teaching Modalities Based on Temporal Social Networks" by Nidia Guadalupe López Flores, María Óskarsdóttir and Anna Sigridur Islind
3. "Research on the Representation, Extraction, and Evolution of Networked Knowledge" by Wang Huaibo and Li Jihong.
4. "Knowledge Graph of Educational Concepts for Structured Learning in Indian Regional Languages" by Satish Kumar.
5. "Towards Theoretical and Generalizable Text Network Analysis: A Case of Forma Mentis Networks" by Oleksandra Poquet and Massimo Stella
6. "Instant Or Distant: The Tale of Two Interaction Platforms and their Influence on Collaboration" by Mohammed Saqr and Sonsoles López-Pernas

REFERENCES

Conde, M. Á., Rodríguez-Sedano, F. J., Hernández-García, Á., Gutiérrez-Fernández, A. & Guerrero-Higueras, Á. M. (2021). Your teammate just sent you a new message! The effects of using telegram on individual acquisition of teamwork competence. *International Journal of Interactive Multimedia and Artificial Intelligence*, *6*(6), 225.

Dado, M. & Bodemer, D. (2017). A review of methodological applications of social network analysis in computer-supported collaborative learning. *Educational Research Review*, *22*, 159–180.

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I. & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior*, *47*, 68–80.

Kaliisa, R., Rienties, B., Mørch, A. I. & Kluge, A. (2022). Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies. *Computers and Education Open*, *3*(100073), 100073.

Newman, M. (2018). *Networks* (2nd ed.). Oxford University Press.

Poquet, O., Saqr, M. & Chen, B. (2021). Recommendations for Network Research in Learning Analytics: To Open a Conversation. *Proceedings of the NetSciLA21 Workshop*.

Saqr, M., Nouri, J. & Jormanainen, I. (2019). A Learning Analytics Study of the Effect of Group Size on Social Dynamics and Performance in Online Collaborative Learning. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies* (Vol. 11722, pp. 466–479). Springer International Publishing.

Broadening the NLP Agenda in Learning Analytics for the Age of AI

Pete Smith, Elizabeth Powers, Henry Anderson, Justin T. Dellinger, George Siemens
University of Texas at Arlington
psmith@uta.edu

ABSTRACT: While the field of learning analytics (LA) has made enormous strides in improving and understanding learning processes with data-driven approaches, some questions have remained difficult to answer using quantitative data alone. Incorporating qualitative, textual data can help address many of these questions, while also providing rich context and nuance that may be otherwise difficult to discover. The field of natural language processing (NLP) has itself undergone a renaissance over the past decade, turning more to massive datasets, enormous compute budgets, and rapid tool development. NLP has the potential for an expanded role in LA and emphasis on increasingly debated artificial intelligence (AI) topics, such as inclusivity, interdisciplinarity, impact, and roles of educator-researchers.

This workshop focuses on connecting scholars and applied experience across fields by asking how: 1) underexplored questions in LA can be addressed by developments in NLP and 2) how the LA field can further dialog in and investigate key areas that are at the center of AI research and application. Participants will 1) explore how LA researchers can incorporate NLP's approaches and tools into their current practices and amplify them into the broader areas of AI research and applied practice and 2) develop lasting networks for future scholarly exchange.

Keywords: natural language processing (NLP), unstructured data, linguistics, language datasets, NLP algorithms, educational contexts

1 INTRODUCTION AND BACKGROUND

1.1 Introduction

As natural language processing (NLP) tools, techniques, and resources have exploded onto the research landscape over the past decade, learning analytics (LA) researchers have made good, oftentimes pioneering use of algorithms and toolsets (e.g., Joksimovic et al., 2014; Kovanovic et al., n.d.; McCaffrey et al., 2021; Öncel et al., 2021; Ullman, 2019). In select cases, LA has led to the development of NLP tools (e.g., Shum et al., 2016). And today, as the field of NLP matures, scholars are openly questioning data, algorithmic, and tool practices within the discipline (Bender & Koller, 2020; Bender et al., 2021; Paullada et al., 2020).

What does NLP look like in an LA context as both fields continue to develop? As scholars move forward to research with and through language in educational contexts, Bender (2021) posits that linguistics will become *broadly inclusive* of different languages and cultures; *integrative,* with increased focus on interdisciplinarity and complexity*; computationally-assisted*, using technology to support larger investigations; and *impact-oriented,* rather than benchmark-oriented.

Will LA researchers play a leading role in the future of NLP in the years ahead, especially given LA's increasing focus on inclusivity, interdisciplinarity, human-machine issues, and an orientation in the "real world" of learning and educational practices? The LA field is uniquely poised to play a leading role in the major shifts to come in NLP research and practice as well as in artificial intelligence (AI) more broadly.

2 WORKSHOP OBJECTIVES

2.1 Objectives

The broader goals of this workshop are to: 1) build capacity for LA research using NLP centered on inclusivity, interdisciplinarity, innovative human-computer approaches, and impact-orientation by the sharing of research and application by leveraging efforts that are currently taking place in existing communities that are centered on educational language processing, LA, or AI; and 2) forge connections between these existing communities to promote and share with a worldwide network.

In addition to the organizers, this workshop will partner with the Learning Analytics Learning Network (LALN), both prior to the conference event (solicitation of papers and presenters) and following the proposed workshop, where LALN will house a curated set of resources created by events as well as continue conversations generated at LAK. The workshop organizers will also produce an event summary report.

2.2 Hosting, Sharing, and Communication

LALN will host and openly share all materials via its website and resource hub. Hosting will include standard presentation artifacts (slide sets, recordings, etc.) as well as the event report.

Any information will be shared through newsletters, social media, and individual networks. An evocative hashtag such as #LALeadstheNewNLP (or similar) will be used.

3 ORGANIZATIONAL DETAILS OF THE PROPOSED EVENT

3.1 Type of Event

This event will take the workshop format. Following the lead of the overall LAK event, this workshop will take place online.

3.2 Proposed Duration

This workshop will have a half-day format.

3.3 Workshop Activities

Following the recommended dates and deadlines, a separate call for papers, demonstrations, and speaker/panel participants will be conducted by the workshop organizers to assemble presenters for different sessions during the event. Speakers will coordinate with event organizers to develop take-away themes into question prompts for breakout group discussions and future networking.

3.4 Proposed Schedule

10 minutes	Introduction
45 minutes	Keynote (30 mins) + Q&A (15 mins) *Power Couple: AI and learning analytics to improve the human experience* Dr. Nia Dowell Assistant Professor and Faculty Innovation Fellow School of Education University of California, Irvine
5 minutes	Short break

60 minutes	Short papers (20 mins each) + Q&A (10 mins each) *Predicting perceived achievement in MOOCs using NLP techniques as LA tools* Dr. Eyal Rabin, The Open University of Israel Dr. Vered Silber-Varod, The Open University of Israel *Data Privacy Challenges in LA* Dr. Djazia Ladjal, Practera Dr. Chen Zhan, University of South Australia Dr. Dinusha Vatsalan, Data61 Dr. Thierry Rakotoarivelo, Data61 Dr. Srecko Joksimovic, University of South Australia
5 minutes	Short break
40 minutes	Demos (15 mins each) + Q&A (5 mins each) *LIWC '22* Dr. Jamie Pennebaker *Others TBD*
15 minutes	Session: Where do we go from here? Dr. George Siemens
5 minutes	Wrap-up

3.5 Expected/Maximum Participant Numbers

This workshop is expected to attract up to 50 participants, which includes presenters and attendees.

4 COMMUNICATION AND RESOURCE SHARING PLAN

4.1 Event Recruitment

In addition to the LAK website, the organizers will utilize personal contact lists, networks, and social media to issue a call for participation as well as other networks dedicated to learning analytics or educational data mining. A website will also be created for the event through LALN. Potential participants include: LAK registrants with beginning- or intermediate-level skills and interests in NLP; advanced NLP participants with research and experience to share, both at this event and in the networked events to follow.

4.2 Information Sharing

Information will be shared with participants before, during, and after through the event by email (primary), the event website, LALN resource hub, and social media platforms.

4.3 What tools will be utilized for organizational purposes?

For event planning and organizational purposes, email will be the primary tool. For resource sharing, the website and resource hub will be used.

REFERENCES

Bender, E. M. (2021, January 6). *Linguistics, living its best life*. Max Planck Institute for Psycholinguistics.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623).

Bender, E. M., & Koller, A. (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185-5198).

Joksimovic, S., Gasevic, D., Kovanovic, V., Adesope, O., & Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The Internet and Higher Education*, *22*, 1–10.

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In D. Gasevic, G. Lynch, S. Dawson, H. Drachsler, & C. P. Rosé (Eds.), *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 15–24). New York, NY: ACM. doi:10.1145/2883851.2883950

McCaffrey, D., Holtzman, S., Burstein, J., & Klebanov, B. B. (2021). What can we learn about college retention from student writing? In *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, (pp.64-66).

Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021). Automatic student writing evaluation: Investigating the impact of individual differences on source-based writing. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference* (pp. 620–625).

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2020). Data and its (dis) contents: A survey of dataset development and use in machine learning research. *arXiv preprint arXiv:2012.05345*.

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.

Raji, I. D., Denton, E., Bender, E. M., Hanna, A., & Paullada, A. (2021, August 28). *AI and the everything in the whole wide world benchmark*. Open Review.net. https://openreview.net/forum?id=j6NxpQbREA1

Shum, S. B., Sándor, Á., Goldsmith, R., Wang, X., Bass, R., & McWilliams, M. (2016). Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 213–222).

Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, *29*(2), 217–257.

Culturally Aware Learning Analytics

Ioana Jivet
Goethe University Frankfurt, Germany
DIPF Leibniz Institute for Research and Information in Education
jivet@dipf.de

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

Maren Scheffel
Ruhr University Bochum, Germany
maren.scheffel@rub.de

ABSTRACT: Learning analytics (LA) have been implemented in various countries, often in different ways and on a limited scale. This makes the transfer of LA solutions from one country to another challenging due to varying contextual, technical, and cultural factors. This interactive workshop aims to explore the role of stakeholders' cultural values and preferences in (1) the acceptance of learning analytics services and related privacy concerns, (2) the design of LA tools, and (3) the evaluation of LA interventions. Throughout the workshop, we will discuss and identify possible cultural differences and similarities–the factors that have hitherto not been extensively studied by LA researchers–for the wider adoption at scale. Through this workshop, we will explore whether the stakeholders' cultural values are some of the factors that the LA community should consider in the design and evaluation of LA systems. In the workshop, we will introduce the participants to *culture-sensitive* and *value-sensitive* design methods, and practice some of them.

Keywords: learning analytics, cultural awareness, value-based approach, scalability, human-centered design

1 INTRODUCTION

In the past years, we have seen many examples of learning analytics being implemented in various countries, but we also increasingly realize that learning analytics are applied very differently in those countries. This makes the transfer of LA solutions from one country to another challenging, due to varying contextual, technical, and cultural factors. In this workshop, we aim to identify and discuss possible cultural differences and similarities–the factors that have so far not been extensively studied by LA researchers–for the wider adoption of LA at scale. Furthermore, the idea that a 'one size fits all' paradigm does not lead to effective LA tool designs and implementation has become accepted within the LA community. Yet, there is still a big question about what factors define the 'right size'. Through this workshop, we will explore whether the learners' cultural values are some of these factors. In an increasingly international educational landscape, how and to what extent should the LA community consider such factors in order to make a relevant impact (i.e., improve learning) at scale? What opportunities are offered by LA technologies to consider learners' cultural preferences and values, and how can we design culturally aware LA services which account for these values?

While there were some initial attempts to focus LA on cultural differences (Vatrapu, 2011), this critical topic is underrepresented in current LA work. However, some research has shown that student learning patterns and learning strategy use in higher education differ across cultures (Marambe et al. 2012). Cultural differences play a role also in online learning contexts as they affect students' collaborative learning (Vatrapu & Suthers, 2007), and educational technology acceptance and use (Nistor et al. 2013). Also, the few LA studies that included the cultural dimension show that cultural differences influenced the effectiveness of LA interventions (Davis et al., 2017; Kizilcec & Cohen, 2017). All these examples show the importance of designing culture-sensitive LA services that would increase the acceptance and adoption of LA at a global scale.

The LAK community would benefit from starting a discussion and drafting a set of suggestions on how to create more inclusive tools that put users and their needs at the centre of the design process. Following such principles could lead to more meaningful tools that do not put certain stakeholder groups at an advantage over others. We believe that the proposed workshop is of particular interest to the LAK community for several reasons, including a need to:

- offer sustainable LA solutions that would facilitate and enhance the process of digital transformation of education
- provide inclusive and equitable quality education
- scale up LA efforts across institutions worldwide
- enable and enhance stakeholders' agency in online learning settings.

2 WORKSHOP GOALS AND STRUCTURE

The present workshop aims to:

- explore and raise awareness of possible effects of stakeholders' cultural values and preferences on: i) the acceptance of LA services and related privacy concerns, ii) the design of LA tools, and iii) the evaluation of LA interventions.
- introduce the workshop's participants to culture sensitive and value sensitive design
- introduce and practice selected design methods 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. 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, e.g., 'power distance' (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. That is, 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.

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. If there is interest, we would also be happy to support the formation of a SIG interested in the topic.

REFERENCES

Baker, R., Ogan, A., Madaio, M., & Walker, E. (2019). Culture in computer-based learning systems: challenges and opportunities. https://edarxiv.org/ad39g/

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* (pp. 454-463). https://doi.org/10.1145/3027385.3027411

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

Marambe, K., Vermunt, J., & Boshuizen, H. (2012). A cross-cultural comparison of student learning patterns in higher education. *Higher Education, 64*(3), 299-316. https://doi.org/10.1007/s10734-011-9494-z

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., Göğüş, A., & Lerche, T. (2013). Educational technology acceptance across national and professional cultures: a European study. *Educational Technology Research and Development*, *61*(4), 733-749. https://doi.org/10.1007/s11423-013-9292-7

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

ACKNOWLEDGEMENTS

Work by Ioana Jivet was funded by the Hessian Ministry for Science and the Arts (HMWK) under the project HessenHub – Netzwerk digitale Hochschullehre Hessen, Innovationforum Trusted Learning Analytics.

Third International Workshop on Human Centred Learning Analytics

Yannis Dimitriadis
Universidad de Valladolid
yannis@yllera.tel.uva.es

Roberto Martinez-Maldonado
Monash University
roberto.martinezmaldonado@monash.edu

Juan Pablo Sarmiento
New York University
jps651@nyu.edu

Fabio Campos
New York University
fcc261@nyu.edu

Patricia Santos
Universitat Pompeu Fabra
patricia.santos@upf.edu

Carla Barreiros
Graz University of Technology
cbarreiros@know-center.at

ABSTRACT: The term human-centred learning analytics (HCLA) was recently coined to refer to the subcommunity of LA researchers and practitioners interested in utilising the body of knowledge and practice from design communities, such as participatory design and co-design, into data-intensive educational contexts. Although there is a growing interest in designing LA systems with students and teachers, several questions still remain regarding how the LA community can appropriate design approaches from other communities and identify best practices that can be more suitable for LA developments. This workshop intends to address some of these questions.

Keywords: design, human-centred, co-design, participation.

1 INTRODUCTION

This workshop seeks to build on the momentum from recent years within the LAK community, around the contributions that Human-Centred Design theory and practice should make to Learning Analytics system conception, design, implementation and evaluation. The theme of LAK18 was Towards User-Centred Design, where there were two sessions devoted to this topic (LAK18-UCD, 2018). At this conference, the first LAK Participatory Design workshop was convened, providing an identity to this strand of work (Prieto-Alvarez et al., 2018). This was consolidated in a special issue of the Journal of Learning Analytics on Human-Centred Learning Analytics (Buckingham Shum, et al. 2019) and four PhDs have recently been completed with explicit attention to participatory design for LA, reflecting interest in the emerging generation of researchers (Dollinger, 2019; Echeverria, 2020). Moreover, the best practitioner paper at LAK20 focused on co-designing with learners (Sarmiento et al., 2020). A version of this workshop was successfully held in LAK20 and in ECTEL21.

1.1 Background

(Mis)understandings of real-world users, stakeholders, contexts, and routines can make or break LA tools and systems. However, the extent to which existing human-centred design methods, processes, and tools are suited to address such human and societal factors in the context of LA is a topic that remains under-explored by our community. In response, the term human-centred learning analytics (HCLA) was recently coined (Buckingham Shum et al., 2019) to refer to the subcommunity of LA researchers and practitioners interested in utilising the body of knowledge and practice from design communities, such as participatory design and co-design, into data-intensive educational contexts. Holstein et al. (2017) were the first in adapting various co-design techniques to identify teachers' data needs and build prototypes of awareness tools with them. In fact, teachers have been the most commonly involved stakeholders in LA co-design studies. For example, Dollinger et al. (2019) discussed implications for the use of participatory semi-structured interviews with teachers in long-term LA projects. Wise and Jung (2019) combined LA interface walkthroughs and transcript analysis to make design decisions for a dashboard intended to be used by teachers. Prestigiacomo et al. (2020) explained how generative tools can be used to investigate the broad challenges that teachers are facing to then focus on those that can be addressed by automatically generating evidence for reflection. Holstein et al. (2019) featured a number of co-design techniques, namely card sorting, directed storytelling, semi-structured interviews, prototyping and behavioural mapping, to co-design a classroom analytics innovation with teachers. Whilst some examples of LA design processes have focused on engaging with students, these are just starting to emerge (Chen & Zhu, 2019; de Quincey et al., 2019; Prieto-Alvarez et al., 2018, Prieto et al., 2019; Sarmiento et al., 2020).

1.2 Aim of the Workshop

The studies presented above make it evident that there is a growing interest in designing LA systems with students and teachers. But several questions remain regarding how the LA community can appropriate design approaches from other communities and identify best practices that can be more suitable to LA developments. However, little work has been done in proposing the steps that other researchers or designers can use as a guidance to structure participatory sessions to understand critical aspects of the envisaged use of LA tools and the actual data needs that stakeholders may have. This workshop aims at consolidating the subcommunity of LA researchers and practitioners interested in the human factors related to the effective design of LA innovations. In doing so, we plan to address questions such as: What has been done so far in HCLA, and what have we learned from these experiences? Within the context of our field, how do we define some fuzzy concepts such as "participatory", "co-design" and "human-centeredness"? Finally, as a community, what do we want to know (research agenda) from now on? Thus, the intended outcome of this workshop is twofold:

Outcome 1: A plan for the consolidation of a new SoLAR SIG dedicated to the study and practice of HCLA within the larger LAK community; and

Outcome 2: The publication of a report summarising the workshop experience and, hopefully, a "roadmap manifesto" setting a research agenda for HCLA.

2 ORGANISATIONAL DETAILS

2.1 Workshop Format, Participation, and Pre-workshop Task

The workshop is envisioned to be a half-day, fully online workshop. Between 12 and 24 participants, with a shared interest in human-centred learning analytics, are expected to be part of this workshop. This workshop welcomes everyone with an interest in the field, from beginners to experts. We will not have a call for papers. Instead, participants will be asked to fill a survey which will capture previous experiences in HCLA as well as current understandings of design aspects that will be relevant for the discussions during the workshop. In particular, participants will be asked to share their experience with human centred design or human centred LA; to define human-centred design; to share what design methods they are familiar with; future plans to adopt human-centred design methods in LA projects.

2.2 Workshop Activities

The workshop is planned to take place during the pre-conference activities of the main conference and is planned for a half-day format of up to 4 hours (March 19 or 20, 2022). 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. **Human-Centred Design challenge**. The second part of the workshop is a collaborative *design challenge*. Participants will engage in creating a research design plan by using human-centred methodologies. They will be grouped in teams of 4-5 people, and go to virtual breakout rooms. They will be presented with a design need and asked to work together designing a human-centred design project to handle the need. Groups will be prompted to consider methodologies such as Value-Sensitive Design, Co-design and generative tools in planning their projects.

3. **Sharing and guided critique**. The third part will be a discussion based on the experience co-designing the human-centred plans. A number of discussants from other communities (e.g. human-computer interaction, interaction design, participatory design and information visualisation), and some that critique human-centred design methods, will be invited to the workshop to give their critical

points of view on the ideas posed in the design plans. We expect that this will lead to a discussion of the pros and cons of human-centred design techniques, what needs to be adapted to fit LA purposes and the differences of meaning of human-centred design for different people.

4. **Discussion on next steps**. All participants will be invited to contribute with ideas to set a potential HCLA research agenda and the potential configuration of a dedicated SIG.

2.3 Dissemination Strategy

A workshop website will be made available upon acceptance of this workshop. A call for participation will be generated and published via the website, and through the twitter accounts and mailing lists the workshop organisers have access to. The website will also include an overview of the aims of the workshop, information about the workshop organisers, contact details and reports and other outputs from the workshop.

REFERENCES

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

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics, 6*(2), 10–26.

Echeverria, E. (2020). *Designing Feedback for Collocated Teams using Multimodal Learning Analytics.* Doctoral Dissertation, Connected Intelligence Centre, University of Technology Sydney, Australia.

Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 257-266).

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: co-designing wearable cognitive augmentation for K-12 teachers. In *Proceedings of the 8th international conference on learning Analytics and knowledge* (pp. 79-88).

LAK18-UCD (2018): User-Centred Design Sessions I and II*, Proceedings of the 8th International Conference on Learning Analytics.*

Prieto, L. P., Rodríguez-Triana, M. J., Martínez-Maldonado, R., Dimitriadis, Y., & Gašević, D. (2019). Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology*, *35*(4).

Prieto-Alvarez, C., Martinez-Maldonado, R., & Shum, S. B. (2018). *Mapping Learner-Data Journeys: Evolution of a Visual Co-Design Tool*. Paper presented at the ACM Australian Computer-Human Interaction Conference, OzCHI'18.

Sarmiento, J. P., Campos, F., & Wise, A. (2020). Engaging Students as Co-Designers of Learning Analytics. *Companion Proceedings 10th International Conference on Learning Analytics & Knowledge* (LAK20).

Wise, A. F., & Jung, Y. (2019). Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making. *Journal of Learning Analytics, 6*(2), 53–69.

LAK22 WORKSHOP: Towards a Philosophical Framework for Learning Analytics. Year Two.

Andrew Gibson[1] and Pablo Munguia[2]
1 Queensland Institute of Technology, Brisbane, 4000 Australia
2Flinders University, Bedford Park, South Australia 5001 Australia
andrew.gibson@qut.edu.au

ABSTRACT: This workshop aims at discussing and sharing ideas to help construct a philosophical framework that learning analytics needs as a field. LAK'22 would be the second year the workshop runs and it builds on the success and reflection of its 2021 iteration. This workshop is the first step towards the development of a philosophical approach to help practitioners collaborate, interrogate, and develop this foundation. The workshop is a half-day event. Participants will be invited to submit a brief position paper for review in advance of the workshop. During the event there will be brief presentations of these papers followed by collaborative activities to create robust, but intellectually stimulating and constructive conversations. The workshop will be synthesized via a multi-authored publication summarizing the points discussed to share with the broader field.

Keywords: Philosophical framework, theories, learning analytics as a discipline.

1 BACKGROUND

Learning analytics is maturing as a discipline. Yet as practitioners from diverse backgrounds bring their varying skill sets, they also bring their own disciplinary based philosophies to the field, which at times can introduce confusion or even dissonance. We argue that there is a need to carry a conversation about how learning analytics should be philosophically grounded and argue for a philosophical framework for the field. We believe a workshop that facilitates this discussion is timely given the questions posed for the 12th Annual Learning Analytics and Knowledge conference. This workshop will be running for a second consecutive year as the 2021 workshop at LAK proved to be popular and successful and previous participants have co-authored a manuscript currently in review. Yet, there is a strong need to continue the discussion and maintain momentum in this important and foundational topic for learning analytics.

We can assume that all learning analytics practitioners share the same ultimate passion and goal, to find qualitative and quantitative ways to improve the learning experience for learners. However, the pathways we take are highly diverse, which can result in resistance when trying to bridge discourse between disciplines. We suggest that learning analytics as a transdisciplinary field would benefit from building on a philosophical construct to allow improved collaboration and uptake by academics and institutions.

The lack of a philosophical framework for learning analytics has significant ramifications which may prevent the field from maturing. For example, it causes confusion when trying to explain what is and is not in scope for the field. Further, the absence of a philosophical framework slows the field down when attempting to test, experiment and scale up ideas and methods. We are still debating at length the ethics surrounding the discipline (e.g., Corrin et al., 2019, Ferguson 2019, Kitto and Knight 2019, West et al. 2020), which a philosophical framework would help resolve.

Selwyn's (2020) provocations express the need to dig deep and assess whether the current direction of learning analytics is indeed what we want for the field. More importantly, Selwyn questions what is actually needed in society and what is missing from our background disciplines when moving into this transdisciplinary space. Finally, such a philosophical framework would allow for the creation of momentum, as the field is reaching a critical turning point: it is needs to move beyond a few practitioners working in isolation or practicing in few classrooms to institutional or national plans to adopt and follow ethical use of learner data for pedagogical purposes. This last point is being made frequently (e.g., Ferguson 2012, Selwyn 2020 West et al. 2020), and while a recent survey showed that institutions are willing (Tsai and Gasevic 2017), when attempting to put in place these methods, we often fail (Ferguson 2012, Buerck 2014, Munguia et al. 2020).

2. ORGANIZATIONAL DETAILS

WORKSHOP TYPE: Interactive Workshop session.

WORKSHOP SIZE: we are targeting 12-15 Participants. With 5-10 key submissions.

DURATION: half a day.

TYPE OF PARTICIPATION: mixed participation (interested delegates may submit a short paper, see below).

EXPECTED ACTIVITIES: short presentations by participants that submitted papers, discussion groups, working on a publication authored by interested participants.

3. OBJECTIVES:

We are creating a sharing and collaborative workshop for two groups of people:

1. those that have an a priori contribution to make about a philosophical framework of learning analytics as a field; and

2. those who have a more general interest in the topic and would like to engage with the ideas proposed by others.

The workshop is designed to meet the following objectives:

1. Initiate a conversation around developing a philosophical framework for learning analytics

2. Provide a forum of friendly critique for existing ideas

3. Present the discussion of ideas in a form that can be disseminated to the wider community.

In order to meet objective 3, we intend that an output of the workshop will be a synthetic paper harvesting participants' input. Our intention is to submit this paper for publication in the JLA.

3. POSITION PAPER SUBMISSION

Workshop participants can submit a paper for the workshop in the form of an extended abstract. The length should be a maximum of 3 pages. Authors should focus on answering the question:

What philosophical ideas should be considered as a foundation for the field of learning analytics and how might practitioners in the field engage with them?

Deadline for submission of contributions is 17 December 2021.

WORKSHOP FORMAT

Time	Activity
1 hour	Contributors present key ideas lightning talk style (4-5 mins each). Other participants write PMI (Plus/Minus/Interesting) short reactions on post-it notes which are grouped.
1-1.5 hours	Group break-out around key philosophical ideas presented - 'birds of a feather' style. Groups engage in dialogue around (a) the potential role of the idea in LA, (b) the value to ALL LA stakeholders, (c) how the idea might advance or secure the field moving forward.
1.5-2 hours	Bring ideas together from all groups and begin writing synthesis paper. Break out in smaller groups as necessary.

REFERENCES

Buerck, J. P. (2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. Journal of Learning Analytics, 1(1), 129-139. https://doi.org/10.18608/jla.2014.11.7

Corrin, L., Kennedy, G., French, S., Buckingham Shum S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper.https://melbourne-cshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4, 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R. (2019). Ethical Challenges for Learning Analytics. *Journal of Learning Analytics*, *6*(3), 25–30. https://doi.org/10.18608/jla.2019.63.5

Kitto, K. and Knight, S. 2019. Practical ethics for building learning analytics. Br J Educ Technol, 50: 2855-2870. doi:10.1111/bjet.12868

Munguia, P., Brennan, A., Taylor, S., & Lee, D. (2020). A learning analytics journey: Bridging the gap between technology services and academic need. The Internet and Higher Education. 46: 100744.

Selwyn, Neil. "Re-Imagining 'Learning Analytics' … a Case for Starting Again?" *The Internet and Higher Education* 46 (July 2020): 100745. https://doi.org/10.1016/j.iheduc.2020.100745.

Tsai, Y. S., & Gašević, D. (2017). Learning analytics in higher education—challenges and policies: A review of eight learning analytics policies. Proceedings of the 7th International Learning Analytics & Knowledge Conference (pp. 233–242).

West, D., Luzeckyj, A., Toohey, D., Vanderlelie, J., & Searle, B. 2020. Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics. Australasian Journal of Educational Technology, 36(2), 60-70. https://doi.org/10.14742/ajet.4653

Answering the Right Questions: Aligning decision-making models with curriculum theory

Author: Elizabeth Dalton
IntelliBoard, Inc.
elizabeth@intelliboard.net

ABSTRACT: Data-informed decision-making tools help participants in the learning process predict likely outcomes and how to affect those outcomes. However, implementations of these systems depend upon assumptions, often implicit, about the curriculum theory paradigm(s) prioritized by the learning institution in deciding what outcomes to support. Making these assumptions explicit can help us to construct predictive learning analytics models that generate predictions and guidance in keeping with the needs of all participants and stakeholders. This half-day tutorial workshop begins with an interactive exercise to "unpack" our assumptions about the nature of teaching and learning and explore how those assumptions can guide model design. We will explore the practical considerations of collecting and processing data most appropriate to those curriculum priorities, using data sets provided by the organizer. Participants may use sample data or to import data to a supplied secure instance of a model building tool to construct features from commonly available data elements and to experiment with model construction and validation. At the conclusion of the workshop, each participant will present a model design to the group that aligns with their chosen curriculum paradigm, including ideal features and labels, possible proxies, and risks to avoid. Future directions will be discussed.

Keywords: Predictive learning analytics, data-informed decision-making, outcomes, feature engineering, success criteria, curriculum theory

1 WORKSHOP BACKGROUND

A common purpose of predictive learning analytics systems is the identification of at-risk students (Arnold, 2010; Smith, Lange, & Huston, 2012; Whitmer, 2013; Jayaprakash et al, 2014; Hlosta, Zdrahal, & Zendulka, 2017) for the purpose of providing opportunities to intervene and improve learning outcomes. Several well-known large-scale implementations of learning analytics decision-making tools focus on on-time course completion and pre-determined, objectively defined outcomes (Pistilli, Arnold, & Bethune, 2012; Whitmer, 2013). These systems provide substantial value when the goal is achievement of a set of "core" competencies by as many learners as possible, as efficiently as possible. Identification of at-risk students can be used to support decisions about where and how to focus teacher and learner effort to greatest efficiency. But this is not the only valid model of education.

As with all learning analytics systems, any proposed decision-support tool must be validated against agreed-upon outcomes for the learning system (Baer & Campbell, 2012). Curriculum theory addresses the overarching goals of an educational program, as well *who* determines those goals. Is an educational program to be judged on the basis of its benefit to each learner, or to a community? Is the value of knowledge in its source (a tradition of scholars or the construction of knowledge internally by the learner), or the uses to which knowledge may be put? Are objective evaluations of educational programs or outcomes even possible? Who will determine what the intended outcomes are for a given program, and whether those outcomes have been met? These questions must be answered in order to define what a decision-making system aims to help participants and other stakeholders achieve.

The purpose of this tutorial workshop is to acknowledge the varying intentions, priorities, and practices of education, whether in public schools, institutions of higher education, corporate settings, or other contexts, and to explore how learning analytics and data-based decision-making tools can relate to and support these different needs.

As one lens of analysis, the curriculum theory work of Michael Schiro (Schiro, 2008) can be used to explore and frame definitions of education and educational purposes. Schiro proposes two axes along which curriculum theories align: a duality between objective and subjective reality, and an orthogonal dimension of whether primacy is given to the source (authority) of knowledge, or the use to which knowledge is put.

Source of knowledge

Objective Reality	Academic Scholar	Learner Centered	**Subjective Reality**
	Social Efficiency	Social Reconstruction	

Use of knowledge

Figure 1: Curriculum Theory Axes (Schiro, 2008, p 179)

Viewed along these axes, we can see an alignment between curriculum theory and pedagogy: between the decisions made about what is "of worth" in education and how we believe teaching and learning occur. We also see implications for who will be using a decision-making system.

Table 1: Curriculum Theory, Pedagogy, Related Terms

Curriculum Theory	Pedagogy	Common Terms	Decision Makers
Academic Scholar	Cognitivism	Traditional teaching	Scholars
Social Efficiency	Behaviorism	Pragmatic Workplace training Competency-based learning Self-paced learning	Funders Employers
Learner Centered	Constructivism	Individualist Learner-centered Open pedagogy	Learners (as individuals)
Social Reconstruction	Social Constructivism Connectivism	Idealist Social justice Transformative learning	Learners (as a group) Activist Organizers

These axes also provide the context in which "quality" and "success" in education are defined, and necessarily, how learning analytics must be implemented in order to support those definitions of quality and success.

As Schiro emphasizes, a single educator or educational program usually references two or more of these purposes, and we see elements of all four of these purposes in programs from local community schools to the most prestigious academic universities to corporate internal training. No single predictive learning analytics model will likely meet all purposes of a program, but a decision-making system can include multiple components to support different priorities.

2 OBJECTIVES

The primary goals of this workshop are as follows:

1. Introducing concepts of curriculum theory to learning analytics researchers and practitioners, including the relationships between curriculum theory, pedagogy, decision-makers, outcome measures, and predictive variables

2. Providing participants with opportunities to explore the availability of outcome measures and predictors appropriate to a program's curriculum theory within currently available data, and to discuss ways to gather necessary data that may not be available within current data systems

3. Developing an instrument to evaluate prior and future studies based on alignment between curriculum theory, data, and decision support tools, including roles supported by tools

3 WORKSHOP ORGANIZATION

This half-day tutorial workshop will contain the following activities:

3.1 Curriculum Theory Exercise (45 minutes)

To establish a common working vocabulary and a starting point for discussions, participants will complete a modified version of the Curriculum ideology inventory (based on Schiro, 2007, p39). This form of the inventory uses an analog 2-dimensional plot to allow participants flexibility in response. Plots will be constructed on a shared workspace (physical or virtual, as applicable) and discussed as a group at the end of the exercise. Participants will explore how the results can be used to inform construction of data-driven decision-support systems.

3.2 Data Demonstration (15 minutes)

The workshop facilitator will demonstrate data import and model building using IntelliBoard, a commercial tool providing SAAS ETL and ad-hoc and scheduled reporting, with examples related to each of the 4 discussed curriculum ideologies. (Concepts and techniques demonstrated in this tool will be applicable to other commonly used ETL and analysis tools.) Techniques related to feature construction, decision-support reporting, proactive notification, and suggested actions and action tracking will be reviewed.

3.3 Participant Model Construction (60 minutes)

Participants will each be provided with an individual secure instance of the IntelliBoard service and may explore existing data or import their own data. This access constitutes a free 30 day trial (no purchase obligation) with an option to extend as part of qualified pilot. (The vendor is open to discussing granted extensions for individual or student researchers.) Each participant will have an opportunity to design a decision-support system, including predictive model, notifications, and recommended actions. Participants may use an alternative tool if preferred. The facilitator will be available to answer questions, and participants can also collaborate.

To connect to the IntelliBoard service prior to the workshop (highly recommended), please visit https://intelliboard.net/research and complete the contact form.

3.4 Participants Present Results to Group (30 minutes)

Each participant will have 5 minutes to present their preliminary results to the group, including implementation of outcomes and predictors related to curriculum priorities, limitations of currently available data sources, and thoughts toward gathering additional necessary data.

3.5 Literature Discussion and Survey Instrument Draft (30 minutes)

Participants will discuss whether a survey of existing literature with respect to curriculum alignment may be helpful, and if so, what selection and categorization criteria may be helpful and who would like to participate. If time allows, participants will discuss establishing a channel for ongoing discussions and outreach.

REFERENCES

Arnold, K. E. (2010). Signals: Applying Academic Analytics. *Educause Quarterly*, *33*(1), n1.

Baer, L., & Campbell, J. (2012). From metrics to analytics, reporting to action: Analytics' role in changing the learning environment. *Game Changers: Education and Information Technologies, Educause*, 53-65.

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 6–15.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Pistilli, M., Arnold, K. E., & Bethune, M. (2012). Signals: Using Academic Analytics to Promote Student Success. *EDUCAUSE Review*.

Schiro, M. (2007). *Extension Activities for Curriculum theory: Conflicting visions and enduring concerns*. Sage. http://www.sagepub.com/schiroextensionactivities

Schiro, M. (2008). *Curriculum theory: Conflicting visions and enduring concerns*. Sage.

Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive Modeling to Forecast Student Outcomes and Drive Effective Interventions in Online Community College Courses. *Journal of Asynchronous Learning Networks, 16*(3), 51–61.

Whitmer, J. C. (2013). *Logging on to improve achievement: Evaluating the relationship between use of the learning management system, student characteristics, and academic achievement in a hybrid large enrollment undergraduate course* [University of California, Davis].

LAK Theory 2022: Workshop on Theory and Learning Analytics

Kathryn Bartimote
The University of Sydney
kathryn.bartimote@sydney.edu.au

Sarah K. Howard
University of Wollongong
sahoward@uow.edu.au

Dragan Gašević
Monash University
dragan.gasevic@monash.edu

ABSTRACT: This half day workshop is the 3rd 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 Roger Azevedo to give the keynote presentation on 'Emotion Theories and Learning Analytics'. 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, along with a theoretical framework of interest. In groups, participants will consider how nominated projects can demonstrate the role of theory in design, model validation and interpretation of findings.

Keywords: Conceptual rationale, framework, paradigm, theoretical model

1 BACKGROUND

LAK Theory will run for its third consecutive year in 2022. Enrolment in previous years has been healthy, even amidst the COVID-19 pandemic.

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, as well as the broader social sciences 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.

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	'Emotion theories and learning analytics' 20 minutes presentation, 10 Q&A	Roger Azevedo
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: https://forms.gle/9QLPKvpJ2s38zv8U8. 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 https://sites.google.com/view/lak22-theoryworkshop. 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.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71.

Reimann, P. (2016). Connecting learning analytics with learning research: The role of design-based research. *Learning: Research and Practice*, *2*(2), 130-142.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5-13.

LAK22 Assess:

The 2nd Workshop on Learning Analytics and Assessment

Dragan Gašević and Mladen Raković
Monash University
{dragan.gasevic, mladen.rakovic}@monash.edu

Naif Aljohani
King Abudlaziz University
nraljohani@kau.edu.sa

José A. Ruipérez Valiente
University of Murcia
jruiperez@um.es

Sandra Milligan
University of Melbourne
s.milligan@unimelb.edu.au

Saeed Ul Hassan
Information Technology University
S.Ul-Hassan@mmu.ac.uk

ABSTRACT: The 1st Workshop on Learning Analytics and Assessment was successfully organized as a part of LAK21 conference. At the 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. The workshop organizers jointly submitted a proposal for Special Issue (SI) on Learning Analytics and Assessment to the British Journal of Educational Technology (BJET). Recently, this proposal has been accepted and the organisers have received more than 30 abstracts from potential contributors. To take advantage of this momentum and continue productive discussions on this important and emerging research topic, we propose the 2nd workshop on Learning Analytics and Assessment. The intent of this workshop is to address some of the key open challenges in learning analytics that are related to reliability and validity of data collection and analysis, use of learning analytics in formative and summative assessment, measurement of learning progression, and assurance of assessment trustworthiness. An open call for contributions will be distributed to solicit brief descriptions of current research and practice projects for roundtable-style discussions with workshop participants. Expected outcomes are the formation of a community of practice and a possible follow-up publication.

Keywords: assessment, learning analytics, educational measurement

1 BACKGROUND

The field of learning analytics aims to harness the potential of digital traces of user interaction with technology. Through the analysis of digital traces, learning analytics seeks to advance understanding and support learning process, and improve environments in which learning occurs. Many promising results in learning analytics have promoted vibrant research and development activities and attracted much attention of policy and decision makers. To date, learning analytics demonstrated very promising results in several areas such as prediction and description of learning outcomes and processes (e.g., Baker, Lindrum, Lindrum, & Perkowski, 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, Roda, van Boxtel & Sleegers, 2012), and frameworks for ethics, privacy protection, and adoption (e.g., Tsai et al., 2018).

1.1 Challenge

Regardless of many promising results, the field still needs to address some critical challenges, including those at the intersection between learning analytics and assessment. For example, how can learning analytics be used to monitor learning progress? How can learning analytics inform formative and summative assessment as learning unfolds? In which ways can validity and reliability of data collection and analysis in learning analytics be improved? These challenges are of high significance in contemporary society that more and more requires development and use of complex skill sets (Greiff et al., 2017). Therefore, learning and assessment experience are closely associated. A growing body of research in educational data mining has been done on developing techniques that can support intelligent tutoring systems with the mechanisms for skill development (Corbett & Anderson, 1994; Desmarais & Baker, 2012). Yet, there is limited research that looks at how data collected and methods applied in learning analytics can be used and possibly constitute a formative or summative assessment. Moreover, can such data and methods satisfy requirements for assessments articulated in psychometric properties, methodological models, and different types of validity and reliability.

The role of learning analytics in analysis of assessment trustworthiness is another open research challenge. This has particularly been emphasized during the COVID19 pandemic with the emergency transition to distance and online education that also required different approaches to assessment that go beyond proctored exams. Several studies proposed the use of data analytic methods for detection of potential academic dishonesty and cheating behaviors. Although some interesting insights are ported and a strong potential to detect suspicious behaviors is demonstrated, there are many open challenges related to technical, ethical, privacy, practical, and policy issues of the development, implementation, and use of such data analytic methods.

1.2 Objective

Motivated by the successfully organized 1st Workshop on Learning Analytics and Assessment during LAK 21, we propose the 2nd iteration of this workshop which main objective will be to further promote research and practice that looks at the intersection of learning analytics and assessment. This workshop will examine approaches that build upon established principles in assessment to improve reliability, validity, usefulness of data collection and analysis in learning analytics. The workshop will also look into the ways how learning analytics can contribute to the future developments in assessment for both summative and formative purposes. The workshop will examine practices for the use of learning analytics to assure assessment trustworthiness with the 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
5 minutes	Welcome and plan for today
60 minutes	Roundtable (Part 1). Work in progress roundtables[1]: 10 minutes to introduce project, summarise progress to date, outline challenges to be overcome, and gather input from the group, followed by 10 minutes discussion with colleagues [x3 before break]
30 minutes	*Break and socialization*
60 minutes	Roundtable (Part 2). Continued [x3 after break]
10 minutes	Roundtable report back: group representatives to summarize conversation and potential impact on the work
15 minutes	Next steps plenary discussion, and close: Gauge interest in further activities around theory and learning analytics e.g. LAK 2023 workshop, LASI 2023 workshop/tutorial, mid-year check in, etc

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 01 November 2021, deadline for abstract submissions 17 December 2021, and notification of acceptance 14 January 2022. We anticipate a registration of up to 30 participants. Please use #LAKAssess when referencing this event on social media.

3 OBJECTIVES/INTENDED OUTCOMES

The workshop will provide a space for both capacity building and connection, and it is hoped that the event will spark the formation of a community of practice. The outcomes of the event will be housed on the Google Site. A possible follow-up publication will be organized in the form of a journal special issue.

[1] Roundtable session presenters will be asked to indicate the stage of their work at the time of submission of a 250 word abstract, e.g. data collection/extraction, data analysis, write up.

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.

Processing and Visualizing Clickstream Data Using R

Fernando Rodriguez
School of Education, UC Irvine
fernanr1@uci.edu

Hye Rin Lee
School of Education, UC Irvine
hyerl@uci.edu

ABSTRACT: Student clickstream data is a valuable data source in learning analytics research. These time-stamped records of student click events within a Learning Management System (LMS) provide researchers with rich and finely detailed information about the learning process. One of the main drawbacks of student clickstream data, however, is that it is complex, requiring a nuanced understanding of the structure of the data, as well as advanced data processing techniques. Adding to this issue is that there are few formalized trainings that teach the specific methods and techniques for working with clickstream data. Thus, the goal of this pre-conference tutorial is to directly train researchers on processing, inspecting, and visualizing clickstream data using the R programming language. During this tutorial, attendees will learn about the general structure of clickstream data and methods for working time-stamped variables. In addition to learning fundamental data cleaning and processing techniques, attendees will also learn how to construct and visualize measures of engagement. Intermediate-level experience with the R programming language is required, as well as familiarity with the Tidyverse family of libraries.

Keywords: Clickstream data, learning management system data, R programming language, data visualization

1 WHAT IS STUDENT CLICKSTREAM DATA?

Student clickstream data refers to time-stamped records of student click events within a LMS (but it can also refer to any time-stamped data collected from a digital learning platform) (Baker et al., 2020). These data provide information about the specific url (or location of a click) and the corresponding timestamp (See Figure 1).

deident_id	url	created_at
24	https://canvas.system.edu/courses/chemcourse/modules/items/229291	3/19/18 5:43
24	https://canvas.system.edu/courses/chemcourse/modules/items/229305	3/19/18 5:15
24	https://canvas.system.edu/courses/chemcourse/modules/items/229287	3/19/18 5:13
24	https://canvas.system.edu/courses/chemcourse/modules/items/229303	3/19/18 5:05
24	https://canvas.system.edu/courses/chemcourse/files/2621536?module_iter	3/19/18 4:53
24	https://canvas.system.edu/courses/chemcourse/modules	3/19/18 4:41
24	https://canvas.system.edu/courses/chemcourse/modules/items/230019	3/19/18 4:41
24	https://canvas.system.edu/courses/chemcourse/modules/items/230019	3/19/18 4:41
24	https://canvas.system.edu/courses/chemcourse/modules/items/230020	3/19/18 4:41
24	https://canvas.system.edu/courses/chemcourse/modules/items/229231	3/19/18 4:40
24	https://canvas.system.edu/courses/chemcourse/modules/items/229229	3/19/18 4:25
24	https://canvas.system.edu/courses/chemcourse/modules/items/229230	3/19/18 4:13
24	https://canvas.system.edu/courses/chemcourse/modules/items/229226	3/19/18 3:51
24	https://canvas.system.edu/courses/chemcourse/modules/items/229224	3/19/18 3:40
24	https://canvas.system.edu/courses/chemcourse/files/2621459/download?d	3/19/18 3:21
24	https://canvas.system.edu/courses/chemcourse/files/2621459?module_iter	3/19/18 3:21
24	https://canvas.system.edu/courses/chemcourse/modules/items/252009	3/19/18 3:20
24	https://canvas.system.edu/courses/chemcourse/modules/items/229291	3/19/18 2:11
24	https://canvas.system.edu/courses/chemcourse/modules/items/229191	3/19/18 0:41

Figure 1: A sample of clickstream data generated from one student

Student clickstream data helps researchers answer important questions about how student learning takes place in online and digital environments. These questions can be practical in nature (How often do students visit the course's LMS?) or pertain to effective course design (Do short format lecture videos increase engagement versus longer format lecture videos?). Clickstream data has also been used to provide theoretical insights into students' self-regulated learning and other motivational processes (Cicchinelli et al., 2018; Li et al., 2020).

Because clickstream data generates a high volume of individual data points, researchers have used it to develop robust predictive models of learning (Ding et al., 2019; Zhou & Bhat, 2021), data dashboards (Diana et al., 2017), and other rich visualizations of the learning process (Park et al., 2018; Rodriguez et al., 2021). Clickstream data can also be used to detect whether students are engaging in potential academic misconduct, such as exploiting features in the learning platform (Paquette et al., 2014; Trezise et al., 2019), or performing well on exams without engaging with course materials (Alexandron et al., 2016; Sangalli et al., 2020).

2 BARRIERS TO LEARNING HOW TO WORK WITH CLICKSTREAM DATA

The main drawback of working with clickstream data is that it is complex, requiring researchers to carefully map specific url pages to course content. Processing clickstream data also requires researchers to inspect whether the timestamp is in the correct time zone, or whether specific clicks and timestamps represent real activity or are simply artifacts in the data (e.g., clicks that were autogenerated when a web-browser activated a cookie). Lastly, researchers must also construct measures of engagement (e.g., number of clicks per day), as clickstream data does not contain aggregate level information in its raw form.

While clickstream data is a valuable source of data for learning analytics research, the inherently complex nature of this data, coupled with limited professional development training, creates a significant barrier for those who wish to conduct learning analytics research using clickstream data, but do not have access to any formalized entry points into the field.

The goal of this pre-conference tutorial is to directly train researchers on processing, inspecting, and visualizing clickstream data using the R programming language. During this workshop, attendees will learn about the general structure of clickstream data and methods for working time-stamped variables. In addition to learning fundamental data cleaning and processing techniques, attendees will also learn how to construct and visualize measures of engagement.

3 TUTORIAL STRUCTURE

This is a three-hour tutorial. During Hour 1, attendees will learn about the structure of clickstream data and what the various elements in the dataset represent. Attendees will also learn how to prepare the data for processing, such as formatting timestamps, categorizing urls by resources type, and checking the data for potential issues.

Hour 2 will focus on cleaning the clickstream data and constructing measures of engagement. These include general indicators (e.g., total number of clicks, number of clicks per day, number of unique visit days) and indicators by resource type (e.g., total number of video lecture clicks, number of video lecture clicks per day). Attendees will also inspect these measures by learning how to generate descriptive diagnostics.

Hour 3 will focus on building data visualizations of the clickstream measures of engagement. This will include histograms, line graphs, boxplots, and scatterplots.

4 TUTORIAL PREREQUISITES

Intermediate-level experience with the R programming language is required, as well as familiarity with the Tidyverse family of libraries. (Wickham et al., 2019).

ACKNOWLEDGEMENTS

The data used for this tutorial is based upon work supported by the National Science Foundation (Grant Number 1535300). The training is supported by the Institute of Education Sciences (Grant Number R305B210006).

REFERENCES

Alexandron, G., Lee, S., Chen, Z., & Pritchard, D. E. (2016). Detecting Cheaters in MOOCs Using Item Response Theory and Learning Analytics. In UMAP (Extended Proceedings).

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018, March). Finding traces of self-regulated learning in activity streams. In Proceedings of the 8th international conference on learning analytics and knowledge (pp. 191-200).

Diana, N., Eagle, M., Stamper, J., Grover, S., Bienkowski, M., & Basu, S. (2017, March). An instructor dashboard for real-time analytics in interactive programming assignments. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference (pp. 272-279).

Ding, M., Yang, K., Yeung, D. Y., & Pong, T. C. (2019, March). Effective feature learning with unsupervised learning for improving the predictive models in massive open online courses. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 135-144).

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. The Internet and Higher Education, 45, 100727.

Park, J., Denaro, K., Rodriguez, F., Smyth, P., & Warschauer, M. (2017, March). Detecting changes in student behavior from clickstream data. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference (pp. 21-30).

Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). Understanding student procrastination via mixture models. In Proceedings of the 11th International Conference on Educational Data Mining. International Educational Data Mining Society.

Paquette, L., de Carvahlo, A., Baker, R., & Ocumpaugh, J. (2014, July). Reengineering the feature distillation process: A case study in detection of gaming the system. In Educational data mining 2014.

Rodriguez, F., Lee, H. R., Rutherford, T., Fischer, C., Potma, E., & Warschauer, M. (2021, April). Using clickstream data mining techniques to understand and support first-generation college students in an online chemistry course. In LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 313-322).

Sangalli, V. A., Martinez-Muñoz, G., & Cañabate, E. P. (2020, April). Identifying cheating users in online courses. In 2020 IEEE Global Engineering Education Conference (EDUCON) (pp. 1168-1175). IEEE.

Trezise, K., Ryan, T., de Barba, P., & Kennedy, G. (2019). Detecting academic misconduct using learning analytics. Journal of Learning Analytics, 6(3), 90-104.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Zhou, J., & Bhat, S. (2021, April). Modeling Consistency Using Engagement Patterns in Online Courses. In LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 226-236).

DesignLAK22: Connecting learning design and learning analytics to develop visualisations for feedback and course review

Nancy Law
The University of Hong Kong
nlaw@hku.hk

Sandra Milligan; Sadia Nawaz
University of Melbourne
s.milligan@unimelb.edu.au; nawazs@student.unimelb.edu.au

Linda Corrin
Swinburne University of Technology
lcorrin@swin.edu.au

Aneesha Bakharia
University of Queensland
aneesha.bakharia@gmail.com

ABSTRACT: The 7th Annual DesignLAK Workshop addresses the challenge of visualising the alignment between learning outcomes, activities, behaviour, and achievement to inform feedback on learners' progress and refinement of course design. Many advances have been made in the field of learning analytics (LA) that have focused on ways that learner behaviours and performance can be visualised for educators and learners. However, there is an ongoing tension between the value of representations of learner behaviour and activity versus a focus on performance and progression outcomes. To reap the potential benefits of LA greater consideration needs to be given to how learning design and models of learner progress can inform the design of analytics and visualisations to enable the provision of pedagogically appropriate feedback to learners while also providing information that can be used by educators to refine course design. In this interactive, half-day workshop participants will be given the opportunity to explore these points of intersection between LA measures, models of learner progress, and learning design. Using the Learning Design Studio tool (Law et al., 2017), participants will also have a chance to create visualisations that can provide feedback to learners and educators to enhance learning progress and design.

Keywords: Learning analytics, learning design, visualisation, feedback

1 BACKGROUND

A primary goal in learning analytics (LA) research and development is to advance the understanding of student learning within authentic educational settings and to provide data-driven insight to learners, educators, and administrators to improve learning and learning design (LD). The intersection between LA and LD enables the development of meaningful insights for educators and learners that can inform learner feedback and the refinement of course design. In a review of the literature, Law and Liang (2020) identified two key features of efforts that are successful in making the LA/LD connection: (1) the existence of a tight coupling between the intended learning outcomes, task sequence, analytics used, and feedback generated; and (2) that the LA to be adopted and the feedback to be given to the learners are part of the LD process rather than a separate post-design activity. To scaffold the deployment of design-appropriate LA requires LD frameworks and technology platforms that enable the operationalisation of these frameworks.

Over time there have been several efforts to develop taxonomies and frameworks to bridge the LA/LD gap. For example, the **C**ognitive **OP**eration framework for **A**nalytics (COPA) (Gibson, Kitto & Wills, 2014) bridges LA and LD through integrating LD-related constructs relating to learner cognition into the LA taxonomy. Seufert et al. (2019) further highlighted the importance of connecting LA objectives to pedagogical concerns when proposing a 2-dimensional framework for the categorisation of analytic objectives: the context of learning (individual vs. social), and the analytic purpose (providing prediction vs. supporting reflection). Building on insight from existing literature, Law & Liang (2020) proposed a multilevel LA taxonomy that connects parameters in five aspects of LA decisions (measures, data type, functionality, techniques, and stakeholders) to three different levels of pedagogical decision making (course, curriculum component, and task levels).

In other emerging work, patterns of performance have been proposed to help educators map how the LD can help learners to achieve certain learning goals, and learners to understand how they are progressing in the trajectory of their learning (Milligan et al., 2020). This can allow learners to monitor and manage their own learning, while also providing educators with feedback on whether the LD is effective in enabling student attainment of learning goals. The models of progression that underpin the analysis of these patterns of performance are strongly influenced by the educator's LD decisions relating to tasks, sequence, and assessment measures. The models are premised on a clear developmental focus which provides learners with an opportunity to demonstrate growth as well as to adapt to the context and content of the course.

The DesignLAK22 workshop will explore how these different LD and learner progress models can be combined with LA measures to be translated into meaningful visualisations of feedback for learners and educators. The workshop session will involve a series of activities on the process of developing these visualisations in practice supported through the use of the Learning Design Studio (LDS) platform (Law et al., 2017). The design of the LDS was informed by Goodyear's (1999) four level framework of pedagogical decision making and Alexander, Ishikawa and Silverstein's (1977) model of a design pattern language to encapsulate the hierarchically embedded granularities in design. The platform provides a learning task taxonomy that reflects the pedagogical orientation of each of the selected tasks and a task setting template to record associated social, technological and assessment settings. LDS enables the exploration of learners' behaviour and outcomes by based on LD principles and assumptions. Educators can also use the LA tool to specify LA visualisations.

Over the past six years, the DesignLAK series of workshops have explored the intersection between learning analytics and learning design from a number of different perspectives. DesignLAK workshops have previously focused on key concepts around feedback processes (Milligan et al., 2016), assessment design (Ringtved et al., 2017), and validity of assessment measures (Law et al., 2019). In 2018, the DesignLAK workshop showcased different LA/LD tools from around the world (Corrin et al., 2018), and in 2021 (our first online workshop) we explored and used MIT's DIVE prototyping tool to allow participants to rapidly prototype LA visualisations with reference to LD (Corrin et al., 2021).

DesignLAK22 will provide an opportunity to bring together key elements from these previous workshops, through the application of models of LA/LD and learner progression that have been inspired by the outcomes of the conversations and sharing of practice through these previous events.

2 OBJECTIVES OF THE WORKSHOP

The objectives of the DesignLAK22 workshop are to explore (1) the alignment between intended learning outcomes, pedagogy, task sequence, and assessment design with LA, (2) the ways that models of learner progression can be developed from a LD to inform how feedback can be visualised, and (3) how the resulting LD-informed LA visualisations can be interpreted and utilised for feedback to learners and educators/designers. The workshop is designed for a wide audience including learning analytics researchers and practitioners, as well as learning designers interested in the use of LA to inform their research/practice (i.e., LAK attendees).

3 WORKSHOP DESIGN

DesignLAK22 is proposed as a half-day workshop and designed to be delivered in an online format. The workshop will be made up of a series of interactive, small-group activities, interspersed with whole-group discussions and opportunities for feedback. The DesignLAK22 team will facilitate the workshop in an online synchronous tool that enables chat communication as well as breakout rooms. The online nature of the workshop can allow for registration of up to 40 participants. The workshop will be open to all educators, learning designers, researchers, and learning analytics practitioners who have an interest in how learning analytics and visualisations can be deployed to improve learning design and the provision of pedagogically appropriate feedback to learners.

1.1 Pre-workshop preparation

The DesignLAK website will be updated to outline the design of the 2022 version of the workshop and provide information on how to get involved for prospective participants. This will include: a summary of the workshop design, information about the workshop facilitators, the workshop schedule, and further resources that may be of interest to participants on the connections and alignment between learning design and learning analytics. The DesignLAK team will promote the event through a range of social media platforms (e.g., Twitter, etc.) and mailing lists of professional societies (e.g., SoLAR, ASCILITE, etc.). Prior to the workshop an introduction email will be sent to remind participants of the workshop schedule and collect information about their roles, context, and interests (which will be used to allocate participants to groups within the breakout sessions).

1.2 The workshop

The workshop will start with an icebreaker activity to allow participants to get to know one another (this may be done in activity groups if the numbers are large). The DesignLAK team will then provide a brief overview of the ideas, models, and previous DesignLAK outcomes that will inform the rest of the activities in the workshop. The workshop will then be split into three main parts: **Part 1** begins with an introduction to the conceptual framework underpinning the LDS platform, followed by a brief introduction to an authentic course design for use in the activities. Emphasis will be placed on the multilevel, hierarchically nested nature of pedagogical decision making, and the need for alignment between intended learning outcomes, pedagogical approach, task sequence and assessment.

Participants will then be divided into small groups of 5 - 6 members to evaluate the strengths and weaknesses of the course design, and to identify one learning outcome and the efficacy of the associated learning tasks that they would like to investigate. The whole workshop group will then reconvene in plenary, during which each activity group will report on the LA questions they would like to explore in the third part of the workshop. This will be followed by a 15-minute break. In **Part 2** participants will reconvene in their groups where they will be asked to consider the alignment of the elements they identified in Part 1 to the assessment design, with reference to a learner progress model. The purpose of this activity is to open up ideas around the ways that the analytics could be visualised for educators and learners. **Part 3** of the workshop begins with a brief introduction to the visualisations that can be generated in LDS, and provision of a set of anonymised data from an authentic implementation of the exemplar course. Participants will then return to their breakout groups to develop a visualisation(s). Each group will evaluate the appropriateness of the visualization for LD refinement and feedback to learners based on their group's exploration. The overall outcomes of these activities would be reported in plenary. The workshop will close with participants sharing reflections on the intersection of LA/LD, the models and systems presented, and the resulting visualisations.

4 WORKSHOP OUTCOMES

The workshop offers participants the opportunity to use the LDS platform to assess design alignment across different components and levels in a course design, to consider learner progress models in mapping LD, and to design feedback visualisations for educators and students. Throughout the workshop participants will be encouraged to Tweet about their learnings, using the hashtag #DesignLAK22, to encourage wider engagement with the ideas presented in the workshop. The DesignLAK22 team will make a summary of the outcomes available on the website post the event. An evaluation of what participants learned from the workshop and what they might do with this new knowledge will be conducted as part of the wrap up session at the end of the workshop, then combined with any ideas of interest that emerge so as to inspire the focus of DesignLAK23.

REFERENCES

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, Construction*. OUP: USA.

Corrin, L., Law, N., Ringtved, U., & Milligan, S. (2018) DesignLAK18: Evaluating systems and tools that link learning analytics and learning design (Workshop). 8th International Conference on Learning Analytics and Knowledge, University of Sydney.

Corrin, L., Bakharia, A., Law, N., & Milligan, S. (2021) DesignLAK21: Rapid prototyping of learning analytics visualisations for learning design. 11th International Conference on Learning Analytics and Knowledge, University of California Irvine.

Gibson, A., Kitto, K., & Willis, J. (2014). A cognitive processing framework for learning analytics. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 212–216). New York: ACM. https://doi.org/10.1145/2567574.256761

Goodyear, P. (1999). Pedagogical frameworks and action research in open and distance learning. European Journal of Open, Distance and E-Learning, 2(1). https://old.eurodl.org/?p=archives&year=1999&article=35

Law, N., Corrin, L., Milligan, S., & Ringtved, U. (2019) DesignLAK19: Ensuring validity in assessment design and analytics (Workshop). 9th International Conference on Learning Analytics and Knowledge, Arizona State University.

Law, N., Li, L., Farias Herrera, L., Chan, A., & Pong, T. C. (2017). A Pattern Language Based Learning Design Studio for an Analytics Informed Inter-Professional Design Community. *Interaction Design and Architecture(s)*, 33, 92 - 112.

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://doi.org/10.18608/jla.2020.73.8

Milligan, S., Ringtved, U., & Corrin, L. (2016) Learning design and feedback processes at scale: Stocktaking emergent theory and practice (Workshop). 6th International Conference on Learning Analytics and Knowledge, University of Edinburgh.

Milligan, S. K., Luo, R., Hassim, E., & Johnston, J. (2020). Future-proofing students: What they need to know and how to assess and credential them. Melbourne Graduate School of Education, the University of Melbourne: Melbourne.

Ringtved, U., Milligan, S., & Corrin, L., Littlejohn, A., & Law, N. (2017) DesignLAK17: Quality metrics and indicators for analytics of assessment design at scale (Workshop). 7th International Conference on Learning Analytics and Knowledge, Simon Fraser University, Vancouver.

Seufert, S., Meier, C., Soellner, M., & Rietsche, R. (2019). A pedagogical perspective on big data and learning analytics: A conceptual model for digital learning support. Technology, Knowledge and Learning, 24(4), 599-619. https://doi.org/10.1007/s10758-019-09399-5

Building Open Science and Scholarship in SoLAR

Bart Rienties
Open University UK

Christopher Brooks
University of Michigan

Simon Buckingham Shum
University of Technology Sydney

Rebecca Ferguson
Open University UK

Maren Scheffel
Ruhr-Universität Bochum

ABSTRACT: This SoLAR initiated workshop aims to bring together the SoLAR community to explore how open science and scholarship can be incorporated into our culture and practice using a three-part interactive format. Amongst the questions to be discussed are: What could and should SoLAR do to encourage open science and scholarship?; What prevents researchers in SoLAR from contributing to open science?; Which learning analytics approaches could (not) be made open, and why?; And what could SoLAR do to make open science and scholarship more attractive and relevant? We aim to share the outcomes of the workshop broadly within the SoLAR community and invite the community to respond to both the outcomes of this workshop and the follow-up SoLAR position paper.

Keywords: Learning Analytics, Open Science, Open Scholarship,

1 PURPOSE OF OPEN SCIENCE AND SCHOLARSHIP WORKSHOP

Across the science community and society at large there are increased calls to make science, its data, analyses, and approaches more transparent (Burgelman et al., 2019), open **(**Kuzilek et al., 2017**)**, accurate **(**van der Zee & Reich, 2018**)**, and reproducible **(**Nosek et al., 2015**)**. At the Annual General Members (AGM) meeting at LAK2021 a motion was discussed originally proposed entitled "Let's prioritize open data, analyses, and science in LAK and JLA!" **(**Brooks, 2021**)**. As Brooks argued, "we should help our society improve our game with respect to open science! Educational research is difficult to replicate and thus doesn't get replicated, leading to spurious results going undetected as such. As one of the largest educational technology societies which has a strong history of empirical research, we are in a great place to help push forward a change!"

The 2021 SoLAR AGM supported the motion on 15 April 2021, which states that SoLAR "recognizes the importance of open, accessible, reproducible, repeatable, and replicable data and analyses approaches. SoLAR also recognizes a diversity of epistemological, ethical, and legal challenges and opportunities which such approaches face. SoLAR will develop a statement that will inform and guide research and practice associated with these approaches." The AGM motion reflected both the priorities and concerns raised in connection with open science and scholarship as the motion was formulated.

This workshop will explore these issues in more depth, using a three-part format. For the first part, participants will be asked to propose ten-minute presentations on any aspect of the relationship across open science, open scholarship, and learning analytics. In the second part, participants will be split into groups, where they will discuss the issues raised in the presentations, relating these issues to their own experiences where appropriate. In the final part of the workshop, participants will work to produce a draft of the SoLAR statement to 'inform and guide practice'.

2 BACKGROUND OF OPEN SCIENCE AND SCHOLARSHIP

There are many definitions and conceptualizations of the umbrella term 'open science'. For example, Fecher and Friesike (2014, p. 17) identified five Open Science schools of thought: "The infrastructure school (which is concerned with the technological architecture), the public school (which is concerned with the accessibility of knowledge creation), the measurement school (which is concerned with alternative impact measurement), the democratic school (which is concerned with access to knowledge) and the pragmatic school (which is concerned with collaborative research)." While all these might have relevance for SoLAR, the democratic school is particularly relevant given its focus on the "principal access to the products of research". As argued by Fecher and Friesike (2014), when research artefacts (e.g., data) are openly available, other researchers are able to check and reproduce published findings, as well as fostering data mining and aggregation of artefacts from multiple data sets and papers, thereby enhancing generalizability and cross-validation across different contexts.

Open scholarship is a slightly broader concept and encourages researchers to share their knowledge and artefacts as early as possible in the research process with others (Burgelman et al., 2019). This approach of open scholarship is gaining momentum by The European Commission and "reflects the inclusion of the humanities in the equation as well as emphasizing the open input side to science in the form of open collaboration and active data and knowledge sharing" (Burgelman et al., 2019, p. 1).

Within SoLAR there could be a range of applications in open science and scholarship, including (in alphabetical order) sharing of algorithms, codes, coding schemes, data, experimental materials, model outputs (without disclosing the underlying data), surveys, synthetic datasets. By allowing SoLAR researchers to share (some or all of) their science and scholarship in an open manner there could be substantial opportunities for a) replication; b) generalization; c) robustness) and d) education of the SoLAR community at large (Brooks, 2021; Kuzilek et al., 2017; van der Zee & Reich, 2018). As argued by Nosek et al. (2015, p. 2) for science to progress it "needs both innovation and self-correction; replication offers opportunities for self-correction to more efficiently identify promising research directions."

At the same time there could be a range of (potentially) legitimate concerns about open science and scholarship, including (in alphabetical order) a) bias towards forms of research that might be easier to facilitate open science and scholarship; b) equity and equality challenges amongst the practitioners of open science and scholarship; c) ethical implications with respect to human subject research; d) intellectual property and legal challenges; e) reputational damage; f) time consuming. As argued by Gehlbach and Robinson (2021) in educational disciplines that use mainly quantitative and standardized approaches it might be easier to facilitate open science and scholarship, but this might bias reported research to quantitative research, while mixed method research or qualitative research might be more difficult to share artefacts due to privacy, ethical, and sample size concerns. In terms of b) as highlighted by one of the few large publicly available educational datasets, OU Analyse (Kuzilek et al., 2017), its widespread use by other researchers has highlighted some unexpected and potentially negative implications. For example, substantial differences in regional progression rates were identified (Rizvi et al., 2019) related to social and economic conditions rather than to educational provision.

In terms of ethical challenges, as urged by Korir et al. (2020) when artefacts are extracted from multiple sources and formats, and in particular when combined with social network and/or discourse artefacts, even if they are “appropriately” anonymized it might still be possible for participants to identify their peer learners. Such issues may not be something which individual researchers can control themselves, as most scholarly institutions have research overseen by ethics or review boards and these boards may not allow disclosure of artifacts. If open science methods were implemented poorly, such actions could bias research reporting towards those who are at institutions which are more willing to release underlying artifacts instead of those who are engaging in the most meaningful research.

Several researchers and organizations consider their artifacts of research as intellectual property, and by making these publicly available they might face legal challenges. For example, one could imagine students who failed a degree suing a university for not reacting to analytics that signaled a risk, a concrete example of John Campbell (2007) 's early call that there may exist an obligation of knowing in learning analytics (Fritz & Whitmer, 2019). In addition, intellectual property is regularly monetized by scholars and their institutions, whether in the form of software, algorithms, or survey instruments, and open release may be a disincentive to some members of the community.

The reputational damage for academics who have made errors in data collection, cleaning, analyzing, or reporting of results could be extremely detrimental to their career, and fear of cancelation for mistakes (such as appearing in Retraction Watch) may paralyze researchers from sharing otherwise strong work with the SoLAR community. Finally, making research open to others may involve substantial costs in terms of time spent on ethics approval, participant consent, data cleaning, anonymization processes, and sharing of analyses or software.

3 OBJECTIVES OF THE WORKSHOP

- Building on the notion of Nosek et al. (2015) that open access is a classical collective action problem, what could and should SoLAR do to encourage open science and scholarship?
- What prevents researchers in the learning analytics community from contributing to open science and scholarship?
- Gehlbach and Robinson (2021) argued, “Some open science practices may be readily adopted, some adapted, and some inappropriate”. Which learning analytics fall into these categories, and why?
- What could SoLAR do to make open science more attractive and relevant, while acknowledging that not all scholarship can be made available openly?

3.1 Contributions to workshop

Short position papers will be invited, using the LAK22 workshop template.

3.2 Sharing of outcomes of the workshop

Accepted papers will be published in the LAK22 Companion Proceedings. A summary of the workshop by the organizers will be shared with SoLAR with the goal of informing the society's policy on openness for LAK and JLA. For example, a statement will form the center of a poster presented (if accepted) at the conference, where it will be used as a stimulus for further discussion and knowledge building. This pairing of LAK workshop and poster to build knowledge together as a community has proved successful at LAK in the past (Clow et al., 2017; Ferguson & Clow, 2017). The organizers will prepare a poster as a result of the workshop to further engage with members of the SoLAR community.

3.3 Dissemination of outcomes

The outcomes of the workshop will be shared broadly within the SoLAR community by SoLAR exec members, which will help to inform a SoLAR position paper. The SoLAR community is invited to respond to both the outcomes of this workshop and the follow-up SoLAR position paper.

REFERENCES

Brooks, C. (2021). *Let's prioritize open data, analyses, and science in LAK and JLA!* Retrieved 11 October 2021 from https://groups.google.com/g/learninganalytics/c/CASQInkuy3c/m/VYZXC_qkAQAJ

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century [Perspective]. *Frontiers in Big Data, 2*(43). https://doi.org/10.3389/fdata.2019.00043

Campbell, J. P. (2007). *Utilizing Student Data within the Course Management System to Determine Undergraduate Student Academic Success: An Exploratory Study.* (Publication Number 3287222) Purdue University]. http://docs.lib.purdue.edu/dissertations/AAI3287222/

Clow, D., Ferguson, R., Kitto, K., Cho, Y.-S., Sharkey, M., & Aguerrebere, C. (2017). *Beyond Failure: The 2nd LAK Failathon Poster* LAK17 Learning Analytics & Knowledge Conference, Vancouver, British Columbia, Canada. http://oro.open.ac.uk/48994/

Fecher, B., & Friesike, S. (2014). Open science: one term, five schools of thought. *Opening science*, 17-47.

Ferguson, R., & Clow, D. (2017). *Learning analytics: avoiding failure*. Educause Review. Retrieved 31 July 2017 from https://er.educause.edu/articles/2017/7/learning-analytics-avoiding-failure

Fritz, J., & Whitmer, J. (2019). Ethical Learning Analytics: "Do No Harm" Versus "Do Nothing". *New directions for institutional research, 2019*(183), 27-38. https://doi.org/https://doi.org/10.1002/ir.20310

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist, 56*(2), 79-89. https://doi.org/10.1080/00461520.2021.1898961

Korir, M., Mittelmeier, J., & Rienties, B. (2020). Is mixed methods social network analysis ethical? In D. Froehlich, M. Rehm, & B. Rienties (Eds.), *Mixed Methods Approaches to Social Network Analysis* (pp. 206-218). Routledge.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset [Data Descriptor]. *Scientific Data, 4*, 170171. https://doi.org/10.1038/sdata.2017.171

Nosek, B., Alter, G., Banks, G., & et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425. https://doi.org/doi:10.1126/science.aab2374

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning; A decision tree based approach. *Computers & Education, 137*(August 2019), 32-47. https://doi.org/10.1016/j.compedu.2019.04.001

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open, 4*(3), 2332858418787466. https://doi.org/10.1177/2332858418787466