34th Australasian Association for Engineering Education Conference
3 - 6 December 2023 – Gold Coast, QLD

Uncovering Students' Learning Pathways: A Process Mining Perspective

Abel Armas Cervantes[a]; Antonette Mendoza[a], and Ehsan Abedin[a,b]
The University of Melbourne[a], Monash University[b]
Corresponding Author Email: abel.armas@unimelb.edu.au

ABSTRACT

CONTEXT

This paper presents an approach to discovering students' pathways when accessing a Learning Management System (LMS), i.e., Canvas. These pathways reflect students' compliance with the subject design and/or alternate ways of learning. Discovering such routines can enable the early detection of students at risk of not achieving the intended learning outcomes, as well as informing academics about students' understanding of their progression in the subject. While LMSs report on aggregate data, they do not report on the order in which students follow the subject design. This information can reveal undesirable situations, such as students responding to the quizzes before completing the prerequisite activities (e.g., watching videos or completing the readings).

PURPOSE OR GOAL

The COVID-19 pandemic forced educators to rapidly innovate, including the introduction of emergency remote teaching and a greater reliance on online student resources. A challenge for educators, especially when many remote teaching formats appear likely to remain long-term, is gaining real-time insights into how students are performing and engaging with a subject. The approach presented here offers a way to produce data-informed insights by analysing student pathways through the LMS.

APPROACH OR METHODOLOGY/METHODS

The approach in this paper is based on *Process Mining*, an established family of tools and techniques that analyse data through a process lens. In process mining, data is represented by means of graphical models depicting student activities (e.g., opening an LMS page, watching a video, responding to a quiz) and the order between them. These models can be annotated with useful quantitative data regarding the frequency of such activities and their order of execution. The approach presented in this paper will show how these models can be extracted automatically from existing data captured by an LMS.

ACTUAL OR ANTICIPATED OUTCOMES

In this paper, first, we outline the pre-processing steps necessary to transform the LMS data into a suitable format for a process mining analysis. Second, we demonstrate how existing process mining operations and tools can be used to derive insights into students' learning pathways. Finally, to showcase the potential of our approach, real-life data extracted from an LMS is analysed.

CONCLUSIONS/RECOMMENDATIONS/SUMMARY

Process mining can be used to discover and analyse students' learning pathways. Existing tools and techniques in process mining can derive actionable insights from the data extracted in LMSs, which can benefit students and academics alike. The analytics extracted using a process lens can complement existing aggregate data analytics available in the LMSs.

KEYWORDS

Process mining, Learning analytics, LMS data, Student pathways.

Introduction

The COVID-19 pandemic had a profound impact on higher education, leading to a widespread adoption of Emergency Remote Teaching (ERT) practices to sustain educational activities during a times of crisis. Unlike traditional online or distance-based teaching approaches, ERT required a rapid and unplanned shift to remote instruction, resulting in a period of experimentation and innovation. As educators adapted to the challenges, they explored new technologies, pedagogical strategies, and teaching methodologies to facilitate effective remote learning experiences (Hodges et al., 2020; Rohani et al., 2023).

These unprecedented circumstances sparked a wave of creativity and flexibility, prompting the development and implementation of novel approaches to subject design and delivery. Educators embraced a variety of instructional formats, including asynchronous lectures, interactive multimedia content, virtual collaboration tools, and adaptive learning systems. In addition, as the pandemic gradually recedes, educational institutions are transitioning towards hybrid learning environments that combine elements of face-to-face instruction with online components. However, it is unclear, how effective these instructional formats are, and if students are following the intended subject designs. An analysis of this information can provide valuable insights into the benefits and drawbacks of different teaching methodologies and the ways in which they can be leveraged to enhance student learning outcomes (Juhaňák et al., 2019; Pardo et al., 2018).

Hybrid environments rely on online Learning Management Systems (LMSs), such as Canvas and Moodle, to manage learning material such as video lectures, readings, quizzes, and discussion boards. In addition to being platforms for content management, LMSs also record how students consume subjects' content and use available resources. This is rich data that can be analysed to discover students' learning pathways and extract insights to improve students' learning experiences. However, analysing such data and deriving meaningful insights is not trivial. Thus, this paper will attempt to answer the following questions:

- How can we discover students' learning pathways from the data recorded in an LMS?
- How can academics use insights gained in the analysis of students' learning pathways to improve teaching and learning practices?

This paper presents an innovative approach to analyse data captured in an LMS (i.e., Canvas) through a process lens. To do so, this research uses Process Mining (van der Aalst, 2012), a well-established family of tools and techniques in the area of Business Process Management and Analytics, that has been less utilised in the teaching and learning realm. By applying process mining techniques to the data captured in LMS platforms, the presented study will analyse the actions executed by students (e.g., accessing course materials, participating in online discussions, watching video lectures, and watching recorded seminars) to identify areas of improvement, and effective practices that can be leveraged to enhance teaching and learning outcomes. This approach will complement existing LMSs analytics, which report on aggregate data (e.g., the number of times a user has visited a page or the number of participations per page) but do not report on the order in which students follow the subject design. This new information can make undesirable situations evident, e.g., students responding to the quizzes before completing prerequisites, such as watching videos, completing the readings, or attending live lectures.

Using this unique process mining approach, we can shed light on the intricacies of the implicit educational processes within LMS platforms. This research shows how technology-mediated learning environments can optimize teaching practices and enhance student engagement through the interpretation of students' learning pathways. The findings of this study have practical implications for educators, instructional designers, and policymakers, as they seek to improve educational practices. Ultimately, we consider our approach shows how the effective utilization of the data captured in LMS can contribute to the ongoing efforts to shape the future of education.

The next section provides a review of the relevant literature about learning analytics and process mining in education. Then, we introduce the approach proposed in our study, alongside with the analysed dataset. In the findings section, we present the primary results and key discoveries derived from our research. These findings are subsequently discussed in terms of their practical implications and potential applications, as well as their significance for future research endeavours.

Background

In the present era, particularly after the COVID-19 pandemic, the prevalence of web-based educational systems has undergone a significant surge. This surge has led to a substantial accumulation of extensive data from a variety of sources, exhibiting diverse formats and levels of granularity (Romero & Ventura, 2017).

New educational environments, such as blended learning, virtual/enhanced environments, mobile/ubiquitous learning, and game-based learning, generate a vast amount of valuable data about students' learning pathways. However, manually analysing this abundance of information is impractical. Indeed, as mentioned in (Baker, 2015), one of the foremost challenges faced by educational institutions today is the exponential growth of educational data and the ability to derive meaningful insights from it. Different approaches and methods have emerged to leverage educational data to enhance both education and the field of learning science (Berland et al., 2014). These approaches can provide a better understanding of students' behaviours in online learning environments, such as LMSs (Juhaňák et al., 2019). The next section discusses two of these approaches: Learning Analytics (LA) and Process Mining in Education (PME).

Learning analytics

Learning Analytics (LA) refers to the process of gathering, analysing, and interpreting data pertaining to learners and their learning environments (Lang et al., 2017). LA research has primarily focused on achieving several key objectives, including the support of instructional strategies and the identification of at-risk students to facilitate effective interventions. Furthermore, LA has been instrumental in recommending appropriate reading materials and learning activities to students, as well as assessing their learning outcomes (Elmoazen et al., 2023; Romero & Ventura, 2020).

In LA, various approaches have been employed, improved, or introduced to facilitate data analysis and interpretation (Elmoazen et al., 2023). These include machine learning, social network analysis and process and sequence mining (Romero & Ventura, 2020). The early stages of LA research primarily focused on developing algorithms for predicting students' success and identifying at-risk students (Ifenthaler & Yau, 2020). Within virtual labs, LA techniques have been applied in various approaches to examine the effectiveness of using virtual labs for acquiring essential skills and competencies. For instance, Govaerts et al. (2012) employed the Student Activity Meter (SAM) to visualize students' performance using multiple metrics. These metrics were then presented in a comprehensive dashboard with dimensional filtering capabilities, providing a holistic view of students' activities and progress.

Other authors have utilized interaction data to analyse students' engagement in online learning environments. This includes statistical analysis of students' interactions, such as the amount of time spent, distribution of time-on-task per student, and examination of different user configurations (Elmoazen et al., 2023; Heikkinen et al., 2023; Ifenthaler & Yau, 2020). These studies aim to gain insights into students' behaviours and patterns of interaction, ultimately enhancing our understanding of their learning experiences (Elmoazen et al., 2023).

Process mining in education

Process mining (PM) is a family of tools and techniques that aims to extract valuable insights and knowledge from event logs (van der Aalst, 2012). PM originated as a technique to analyse event logs captured by information systems when executing business processes; however, PM has been used in a wide range of contexts including education.

Educational Process Mining (EPM) is a nascent area within the field of Educational Data Mining (EDM). By leveraging data recorded in educational environments, EPM aims to make unexpressed knowledge explicit and extract insights from educational settings (Bogarín et al., 2018). It involves the analysis of event logs and data generated by LMSs to discover, monitor, and improve educational processes (Romero et al., 2016). For instance, Schoor and Bannert (2012) applied process mining techniques to map social regulatory processes in computer-supported collaborative learning, demonstrating its usefulness in capturing and analysing collaborative activities. In another study, Bannert et al. (2014) focused on self-regulated learning and utilized PM methods to analyse qualitative data obtained through think-aloud protocols. Such work found that PM can provide valuable insights into self-regulated learning processes. Maldonado-Mahauad et al. (2018) analyse data collected in a Massive Open Online Course (MOOC). The authors use a PM method and hierarchical clustering techniques to identify common learning tactics and strategies. Their results show that it is possible to identify distinct levels of self-regulated learning from the collected data. Other works have applied PM techniques to specific learning activities. For instance, Juhaňák et al. (2019) analysed students' behaviour while taking a quiz. The authors uncovered unique patterns providing insights into students' quiz-taking strategies within the LMS environment.

The current paper shows how PM can be used to discover and analyse students' learning pathways from the data collected by an LMS. The data analysed in this paper originated from a subject implementing flipped classroom. Different from existing work (e.g., van der Aalst (2012), Bogarín et al. (2018), Juhaňák et al. (2019); Romero and Ventura (2020)), our approach demonstrates that it is not necessary to have data stemming from fully online subjects or MOOCs to apply PM. In fact, we show that PM could be useful if it was integrated as part of the analytics toolset of an LMS.

Approach

The approach presented in this paper uses Process Mining (PM) to analyse the data recorded in an LMS. Process mining adopts a process lens to analyse event logs generated by information systems. These event logs record footprints of process executions describing – among other information – activities performed by process participants and the time when such activities were performed. There are four main operations in process mining: model discovery, conformance checking, variants analysis and process enhancement. Model discovery aims to generate graphical models representing the process captured in the log. Conformance checking compares the expected and observed behaviour to find undesirable deviations. Variants analysis compares different versions of a process to find discrepancies between them, and process enhancement improves a process with information extracted from the log.

In PM, an event log contains information about events, which are activity instances performed during the execution of a process. A sequence of events representing the execution of the process from beginning to end is called a trace. Thus, an event log contains a set of traces. For example, Fig. 1 depicts the model of a trace with five events (boxes), where the arrows denote their order of execution. This small example can describe the expected students' learning pathway in a week: the student watches a video lecture, reads suggested material, attends the live zoom seminar, completes the workshop exercise, and answers a weekly quiz. The model in Fig. 1 could be the result of a model discovery operation. In PM, this type of models is commonly known as Directly Follows Graph (DFG), and it contains nodes and edges, where the nodes represent activities and edges define the order between these activities.

Figure 1. Model example of an ideal student' learning pathway

Each event in an event log requires three key pieces of information: *Case ID, Activity name* and *Timestamp*. Events with the same *Case ID* represent a trace, and these events are ordered according to their *Timestamp*, which represents the moment when the event was created or completed. As described earlier, an event is an instance of an activity, hence there can be many events with the same *Activity name* in a single trace (e.g., if a student watched a video lecture more than once).

Process mining using LMS data

In this paper, the data provided by the LMS administrators[1] consisted of records of transactional data for each user (e.g., students, tutors, and lecturers), and each data entry represented a visit to a page in an LMS module. An entry contains information about the URL of the visited page, Timestamp, user ID, session ID, internet browser used, etc. In this paper, we focus only on three elements: URL of the page visited, Timestamp and user ID. These elements are sufficient to define the three key pieces of information necessary for PM: *Activity name, Timestamp* and *Case ID*. Each data entry recorded in the LMS is transformed into an event log event as follows:

Activity name: To obtain an activity name that is meaningful, it is necessary to look at the content of the page at the URL captured in the data entry. For example, the URL can be for a page with a video lecture, a workshop exercise, a quiz, or a page with the seminar recording.

Timestamp: This is taken from the LMS data, which represents when the URL was visited.

Case ID: Case ID segments the data into traces, and its definition depends on the type of analyses to be performed. For example, all the actions performed by a student during the whole semester can define a trace, then the user ID can be the *Case ID*. This segmentation would describe from beginning to end the way students progressed during the semester. Another possibility is to assign the same *Case ID* to all the activities performed during a single session, where a session is from the moment when the student logs in into the LMS until the moment when the student logs out. This segmentation would describe what students do every time they log into an LMS module. Another possibility is to track the progress of the student per week.

Thus, the *Case ID* can be defined as a combination of “user ID + Week number”. In this segmentation, all events executed within a week by a student will have the same *Case ID*.

Table1 shows an example of an event log for process mining. This event log can be used to discover the model in Fig. 1.

Table 1. Event log example

Case ID	Activity name	Timestamp
1	Watch Video lecture	Feb 30, 2023, 15:00
1	Complete Reading	Feb 30, 2023, 16:00
1	Attend Live Zoom Seminar	Mar 1, 2023, 11:00
1	Complete Workshop Exercise	Mar 1, 2023, 15:00
1	Answer Weekly Quiz	Mar 1, 2023, 16:00

Traces represent different executions of the process, which can capture alternative orders between activities. In the context of an LMS, students can visit the Canvas pages in different ways, e.g., they could attempt to answer the quiz before watching the video lecture or attend the workshop without reading the material. This heterogeneity in the observed traces may lead to very complex DFG models. Figure 2 depicts two DFGs representing many different traces, such models can be automatically generated with a process mining tool such as Disco[2] or Apromore[3]. Note that such DFGs contains two additional nodes representing the start (green node) and end of the process (red node). Additionally, the nodes and edges are annotated with information about the frequency. Frequency of nodes represents the number of times an activity was executed, and frequency of edges represents the number of times a particular order between activities was observed; this information about frequency also corresponds to the colour of the boxes and thickness of the edges, where the darker the colour (thicker the edge), the more frequent the activity (order between activities).

Results and Findings

To show the potential of the approach, we analysed a semester’s worth of data collected in the LMS module of a master’s subject. The subject was part of the Master of Information Systems at

[1] The LMS data was available only after an ethics application had been approved. The data was provided after the final marks had been released and it was anonymised prior its analysis.
[2] https://fluxicon.com/disco/
[3] https://apromore.com/

the University of Melbourne and had 155 enrolments. During the data preparation, only data entries related to the subject's content were kept (e.g., video lectures, quizzes, and workshops).

Data entries related to support resources were filtered out, for example, pages with information about student support services or IT support. The subject adopted a flipped classroom approach, and it was expected that students would watch pre-recorded videos prior the live seminar (live lecture), and they would attend a workshop after the seminar. The collected data was suitable for a PM analysis as a great part of the learning activity was expected to happen in the LMS.

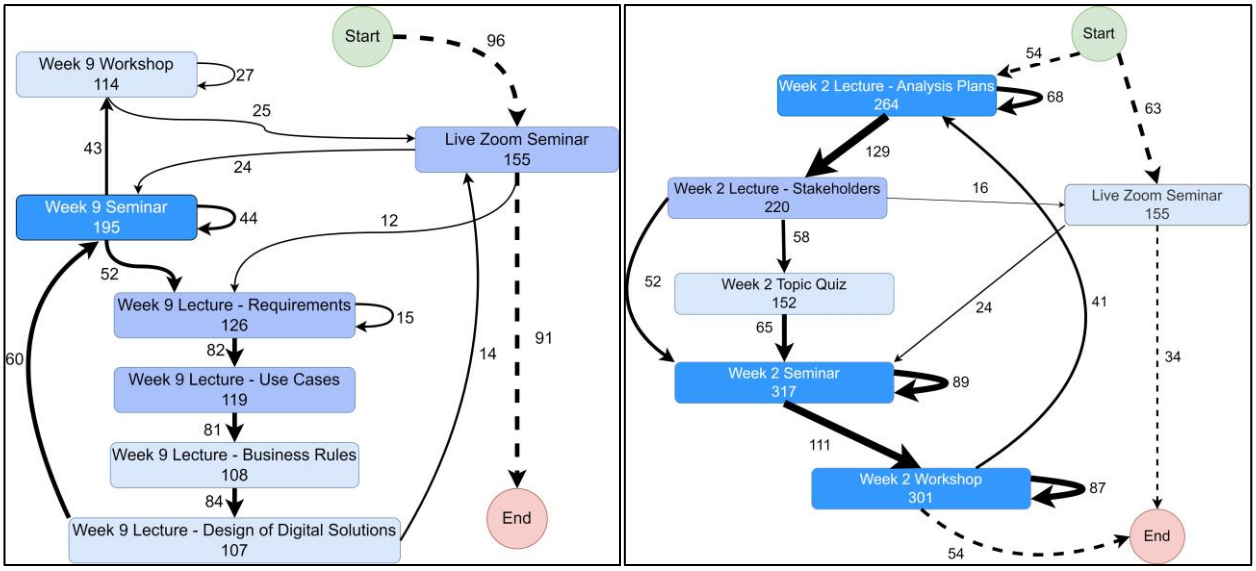
When parsing the LMS data into an event log for PM, each data entry was transformed into an event, where the *Case ID* was defined as a combination of *user ID + Week number*, the *Activity Name* was the name of the resource at the URL in the LMS data such as video lectures, quizzes, seminars, sample exams, workshops and the *Timestamp* was kept as it was recorded in the data entry in the LMS. This resulted in an event log with a total of 2066 traces, which represents the 155 students accessing the LMS data for 16 weeks (the semester consists of 12 teaching weeks, but some students also accessed the LMS module during the exam period). In the event log, we named the activities as follows. Lectures are pages containing pre-recorded videos that students needed to watch before the live zoom seminars. Live zoom seminars represent live lectures where lecturers guided students through exercises related to the content of the lectures. Seminars represent the pages where the slides for the live lectures and the live zoom seminar recordings were uploaded (the recording was uploaded after the live lecture). Workshops are pages containing exercises for the workshop session, in this subject there was a 1-hour workshop per week. Finally, quizzes and sample exam are self-explanatory. Due to low student participation, quizzes were removed after the first half of the semester.

Table 2 shows the 10 most-visited resources. Activities such as "Week 1 Seminar", "Week 2 Seminar", and "Week 9 Seminar" show higher frequency, indicating that students visited the corresponding pages several times. The relative frequency indicates the number of times an activity was visited, as a percentage of the total number of visits recorded. As mentioned above, seminar pages contain the slides of the week and the recordings of the live lectures. While high participation is expected from the first two weeks of the semester where the outline of the subject and details of the assignments were discussed, the content delivered during week 9 was particularly complex and, in fact, the content of this week was majorly changed in the next edition of the subject.

Table 2. Top 10 most-frequent activities

Activity	Frequency	Relative Frequency
Week 1 Seminar	778	3.22%
Week 2 Seminar	734	3.03%
Week 9 Seminar	681	2.82%
Week 8 Lecture - Business Process Models	645	2.67%
Week 2 Lecture - Analysis Plans	642	2.65%
Week 12 Sample Exam	622	2.57%
Week 7 Lecture - Personas	609	2.52%
Week 2 Lecture - Stakeholders	549	2.27%
Week 7 Lecture - Journey Map	540	2.23%
Week 8 Lecture - Future state analysis	527	2.18%

Given that week 9 was highly visited, and to demonstrate the power of PM in making meaning out of the log from LMS, we used week 9 as one instance to be analysed. In this analysis, only events executed during week 9 were kept and, using these events, a model was discovered (Fig. 2 (a)). These visualisations can be automatically generated with any of the available process mining tools (e.g., Disco or Apromore). Please note that, existing tools offer functionality to hide infrequent edges and focus on the most-common pathways. The models in Fig.2 show the most-common pathways for readability purposes, hence the sum of edge frequencies may differ from activity frequencies (the complete models with all edges become unreadable). Figure 2 (a) shows that 96 students – almost 62% of the cohort – started their week by attending the Live Zoom Seminar and did not watch the pre-recorded video lectures that laid the foundations for the exercises discussed during the live session. While not visualised in Fig. 2, the data shows that only 11% of the students watched the videos before the live zoom seminar. For comparison purposes, Fig. 2 (b) shows the model representing the pathways of the students during week 2. The model shows that 40% of the students attended the live seminar without watching the pre-recorded video lectures, while 35% of students were compliant with the instructions and watched them.



(a) **(b)**

Figure 2. The main process map of students' pathways **on LMS**

As can be seen in the Fig. 2, students followed different learning pathways. For instance, in Fig. 2 (b) some students followed a more linear learning pathway – e.g., Start → Week 2 Lecture – Analysis Plans → Week 2 Lecture - Stakeholders → Week 2 Topic Quiz → Week 2 Seminar → Week 2 Workshop → End). On the other hand, some other students show a more cyclical approach by repeating some activities – e.g., Week 2 Lecture – Analysis plans → Week 2 Lecture – Stakeholders → Week 2 Seminar → Week 2 Workshop, and again going back to Week 2 Lecture – Analysis plans.

Discussion

Using process mining, we were able to uncover students' learning pathways in the LMS. We observed that although students had their own unique way of participating in a subject, some students also tended to follow a similar progression of activities, visible as paths formed by the thicker edges in a DFG. For instance, Fig.2 (a) shows that several students watched the pre-recorded video lectures in a certain order. One noteworthy observation is the fluctuation in student activity across different weeks of the course. As mentioned previously, weeks 1, 2 and 9 were highly visited by students. This suggests that these weeks may have featured topics of interest, such as the start of the semester, or challenging topics, such as the content of week 9.

By using students' pathways and analysing the sequence of activities they performed, we can identify their learning path through the subject. For example, we could easily detect that only 62% of the students did not watch the video lectures before the live seminar. This can help identify activities that students have not completed and are essential for understanding concepts and content delivered in subsequent seminars or workshops.

Our analysis reveals disparities in student learning behaviour with specific topics and workshops. For instance, video lectures on "Business Process Models" in Week 8, "Analysis Plans" in Week 2, and "Personas" in Week 7 have garnered considerable attention. The relevance, practicality, or complexity of these topics may have played a key role in attracting students' interest and encouraging their revision. Please note that there may be limitations in the information captured in the LMS. For example, students may study in groups using one student's LMS account to access the content, hence only the activity of a single student may be captured in the data.

Implications:

The implications of these findings are significant for educational practitioners and researchers alike. By visualising students' learning pathways on LMS, educators can design more effective and engaging learning experiences in following editions of a subject. For instance, the insights gained in the analysed data corroborated the lecturer's impression about the complexity of the content of week 9, which was completely redone for the next edition of the subject. Other insights can support data-informed decisions about the allocation of resources, time, and effort to topics and activities that generate higher student interest and participation, or that are more complex for students. For example, using our approach, it is easy to detect the video lectures that are being re-watched several times, which may indicate that students are struggling to understand the concepts presented in such videos.

Another advantage of our approach is that it can help detecting and visualising students' adherence to suggested learning pathways. Violations to the suggested pathways can be detected in real-time, as LMS is constantly recording data. Using this information, educators can then focus on reminding students what is the suggested learning pathway and optimizing the structure of the course, ensuring a smooth progression of knowledge acquisition. Finally, the presented approach can also help detecting underperforming students who are at risk of not achieving the intended learning outcomes. For example, students who are neither watching the lectures nor watching the seminar recordings. Identifying such students in a timely manner can inform educational practitioners when is necessary to intervene and assist these at-risk students.

Limitations and future research:

While the quantitative analysis of log data provides valuable insights, it is important to acknowledge the limitations of this study. The log data alone lacks contextual understanding of students' motivations, challenges, or learning experiences. Supplementing the quantitative findings with qualitative data, such as surveys or interviews, could provide a more comprehensive understanding of student engagement, allowing for a richer analysis of factors influencing their behaviours. A possible limitation in the applicability of the approach is related to the data captured in the LMS. The data analysed in this paper was suitable for this analysis because all content of the subject was distributed online, which may not be suitable when most of the learning happens outside the system.

As future research, we will investigate the use of interventions based on the LMS data to detect and help students at risk of not achieving the intended learning outcomes. Currently, we are developing a process mining tool that can be integrated into an LMS to complement the current analytics toolset. Another promising avenue for future research is the analysis of subject designers, lecturers or tutors to understand and improve processes related to subject design.

References

Baker, R. (2015). Big data and education. *New York: Teachers College, Columbia University*.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning*, *9*, 161-185.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, *19*, 205-220.

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1), e1230.

Elmoazen, R., Saqr, M., Khalil, M., & Wasson, B. (2023). Learning analytics in virtual laboratories: a systematic literature review of empirical research. *Smart Learning Environments*, *10*(1), 1-20.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 869-884).

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2023). Supporting self-regulated learning with learning analytics interventions–a systematic literature review. *Education and Information Technologies*, *28*(3), 3059-3088.

Hodges, C. B., Moore, S., Lockee, B. B., Trust, T., & Bond, M. A. (2020). The difference between emergency remote teaching and online learning.

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, *68*, 1961-1990.

Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, *92*, 496-506.

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (2017). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research New York.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179-196.

Pardo, A., Bartimote, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., & Mirriahi, N. (2018). Ontask: Delivering data-informed, personalized learning support actions.

Rohani, N., Gal, K., Gallagher, M., & Manataki, A. (2023). Discovering students’ learning strategies in a visual programming MOOC through process mining techniques. Process Mining Workshops: ICPM 2022 International Workshops, Bozen-Bolzano, Italy, October 23–28, 2022, Revised Selected Papers,

Romero, C., Cerezo, R., Bogarín, A., & Sánchez-Santillán, M. (2016). Educational process mining: A tutorial and case study using moodle data sets. *Data mining and learning analytics: Applications in educational research*, 1-28.

Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(1), e1187.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355.

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*(4), 1321-1331.

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

Copyright statement