Bus Inf Syst Eng
https://doi.org/10.1007/s12599-024-00901-7

RESEARCH PAPER

# Enhancing Personalized Learning Through Process Mining

## A Taxonomy and Design Patterns Approach

**Thiemo Wambsganss · Anuschka Schmitt**

Received: 7 September 2023 / Accepted: 3 September 2024


**Abstract** Technology-mediated learning offers new possibilities for individualizing learning processes in order to discover, monitor, and enhance students' learning activities. However, leveraging such possibilities automatically and at scale with novel technologies raises questions about the design and the analysis of digital learning processes. Process mining hereby becomes a relevant tool to leverage these theorized opportunities. The paper classifies recent literature on individualizing technology-mediated learning and educational process mining into four major concepts (purpose, user, data, and analysis). By clustering and empirically evaluating the use of learner data in expert interviews, the study presents three design patterns for discovering, monitoring, and enhancing students' learning activities by means of process mining. The paper explains the characteristics of these patterns, analyzes opportunities for digital learning processes, and illustrates the potential value the patterns can create for relevant educational stakeholders. Information systems researchers can use the taxonomy to develop theoretical models to study the effectiveness of process mining and thus enhance the individualization of learning processes. The patterns, in combination with the taxonomy for designing and analyzing digital learning processes, serve as a personal guide to studying, designing, and evaluating the individualization of digital learning at scale.

**Keywords** Process mining · Learning processes · Literature review · Taxonomy · Technology-mediated learning · Learning analytics · Design patterns

## 1 Introduction

The gold standard for helping learners to acquire and comprehend knowledge is a personal human educator. This is because human educators can recognize learners' gaps, anticipate needs, suggest suitable learning materials, or provide individualized feedback. However, such individualized learning is only common in small classes, as such learner-centered teaching approaches are hardly scalable to large classrooms (Gupta and Bostrom 2013; Huang et al. 2021). Therefore, individual support and personal recommendations for students in their learning processes remain a pending challenge that has not yet been solved in many contemporary learning scenarios (e.g., Kulik and Fletcher 2016). By *learning scenario*, we mean a certain course, class, or lecture which an educator or a machine is giving to students to help the students reach certain learning outcomes (e.g., understanding and applying a programming language). High schools and universities struggle to offer this kind of individual support due to financial and organizational constraints (Seaman et al. 2018). As Winkler et al. (2021) state, *"the growing number of [large] classroom sizes in high schools and vocational schools, mass lectures at universities with more than 100 students per lecturer, and massive open online courses (MOOCs) with more than 1000 participants make individual interaction with a teacher or tutor even more difficult"* (Winkler et al.

Accepted after three revisions by Daniel Beverungen.

T. Wambsganss (✉)
Institute of Digital Technology Management, Bern University of Applied Sciences, Brückenstrasse 73, 3005 Bern, Switzerland
e-mail: thiemo.wambsganss@bfh.ch

A. Schmitt
London School of Economics and Political Science, Houghton St, London WC2A 2AE, UK
e-mail: a.schmitt2@lse.ac.uk

Published online: 28 October 2024

2021, p. 1). Several studies have revealed that this lack of individualized support leads to procrastination, low learning outcomes, high dropout rates, and dissatisfaction with the overall learning experience, which can ultimately widen the gap among students (Eom et al. 2006; Brinton et al. 2015; Hone and El Said 2016; Huang et al. 2021).

In response to the lack of individual student guidance, there is a steady growth of technology-mediated learning (TML) information systems in education, such as the learning management systems Canvas or Moodle. These systems both facilitate the traditional teacher route to personalized learning and enable the more learner-centered route, where learners take a more active role in their educational journey (Betts and Rosemann 2022). Extending this concept further, the integration of machine learning and artificial intelligence into these platforms proposes a machine or augmented route that offers unprecedented opportunities for personalized learning experiences at scale (Betts and Rosemann 2022). TML thus represents a solution for providing more individualized learning support and TML systems offer the potential to analyze the students' learning process and identify potential gaps that can be addressed through individualized learning recommendations (Gupta and Bostrom 2009, 2013). According to Dahlstrom et al. (2014), 99% of U.S. colleges and universities organize their learning scenarios in a standard learning management system. Taking a TML perspective*, a learning process* is a sequence of single learning activities students perform to reach a certain learning goal (e.g., analyzing a historical text, and discussing its meaning with peers to understand the complexity of a certain historical event).

Both the organization of courses in a system and the embedding of exercises in intelligent tutoring systems (i.e., Kulik and Fletcher 2016) or computer-supported collaborative learning tools (i.e., Dillenbourg et al. 2009) are expected to continuously grow to a market size of $336.98 billion by 2026 (Bogarín et al. 2018; Syngene Research LLP 2019; Romero and Ventura 2020). However, a learning management system that only organizes courses does not necessarily provide individualized learning experiences to learners (Huang et al. 2021). It is important to capture learner behavior and preferences while deploying and managing learning materials and activities (Nguyen et al. 2020b; Huang et al. 2021). Canvas or Moodle offer promising opportunities to capture data about learning activities. For example, rich traces are captured in *event logs* – a combination of data points based on a timestamp, an event ID, and an activity (e.g., starting or ending event of a particular exercise at a certain time), textual data (e.g., written essays) or measured learning outcomes (e.g., a student's skill level after taking a quiz) (Nguyen et al. 2020b; Cerezo et al. 2020; Han et al. 2021). This opens promising potentials to discover individual learning activities at different granularity levels, anticipate learner needs, and create individualized learning experiences (Nguyen et al. 2020b; Han et al. 2021).

Even though TML systems in education are omnipresent, the potential of leveraging the created learner data for monitoring, discovering, and enhancing the students'learning has not been tapped (e.g., Nguyen et al. 2020a). Process mining techniques provide an effective way to use learner data to improve education by combining the benefits of learning analytics with process modeling and analysis. For example, process mining can be employed to enhance the identification of students at risk of dropping out or underperforming by analyzing their interaction patterns as learning processes in online learning platforms such as Moodle or Canvas. By examining sequences of activities and their frequencies, educators can pinpoint deviations from successful learning paths and thus receive a more transparent analysis compared to contemporary non-process focused monitoring techniques. Originating as a sub-discipline of data mining, process mining adds a process-oriented perspective to the analysis of data in business processes (van der Aalst 2012; vom Brocke et al. 2021). While usual learning analytics primarily focuses on data dependencies and pattern predictions of single-learner activities (Romero and Ventura 2020), process mining can consider existing event logs beyond a single activity (Bogarín et al. 2018; Söllner et al. 2018; Juhaňák et al. 2019; Cerezo et al. 2020). We refer to this learner-centered process viewpoint to analyze and improve digital learning activities as *educational process mining.* Based on the previously mentioned event logs, educational process mining can provide additional benefits for designing and analyzing learning processes based on data by crossing the boundaries of single events or tasks and enriching the analysis from a learner's process perspective (Söllner et al. 2018; Roth 1970).

While process mining is a well-established field of research and the theoretical benefits of educational process mining appear to be clear, the empirical exploration of applications and their practical implementation is not as straightforward (Ghazal et al. 2018; Rogiers et al. 2020). Thus, our understanding of educational process mining lacks a conceptualization of relevant and varied learning scenarios, including process mining's implications for the different, relevant stakeholders, including educators, institutions, and learners. This gap highlights the need for a detailed exploration of educational process mining for the different *stakeholders* (e.g., educator, institution, or learner), the *technology* (e.g., the applied discovery algorithm), the *pedagogical context* (e.g., in which educational domain or for which learning objective and task educational process mining is analyzed) and its effect on individual *learning outcomes* (e.g., Okoye 2019; Cerezo et al. 2020). An interdisciplinary information systems viewpoint is important to systematically design, analyze, and compare the various configurations of process mining that extends to the particular application domain of education (Sidorova et al. 2008; O'Neill et al. 2011; Matook and Brown 2017; vom Brocke et al. 2021). This research aims to advance the field by aggregating knowledge about the dimensions and characteristics of educational process mining (Nickerson et al. 2013), enabling more effective discovery, monitoring, and enhancement of student learning processes. Our study seeks to address this need by answering the following research question (RQ) on the design and analysis of digital learning processes through process mining:

RQ1 What are the dimensions and characteristics of designing and analyzing digital learning processes with educational process mining?

In a subsequent step, we are interested in equipping information systems researchers and educational practitioners with useful information and tools to translate knowledge on educational process mining into tangible design steps of leveraging digital learning data. According to Schoonderwoerd et al. (2022), design patterns are proven solutions for recurring problems that make complex domain knowledge accessible and applicable for non-domain experts (e.g., providing educators or educational designs with support to design and analyze digital learning processes). Design patterns have been proven to be a feasible way to communicate design knowledge on IT artifacts in general (e.g., Dickhaut et al. 2022), as well on digital learning processes, specifically (e.g., Weinert et al. 2021). We aim to build upon the design pattern paradigm to address the challenges of designing and analyzing learning process with educational process mining and pose the following, subsequent question:

RQ2 What are the design patterns for using data to discover, monitor, or enhance students' learning processes?

Consistent aggregation of the characteristics and dimensions of the design and analysis of learning processes based on educational process mining will help researchers and practitioners to systematically use data to discover, monitor, and enhance students' learning paths. The design patterns, derived from our taxonomy and informed by expert interviews, enable educational designers to enhance learning processes within TML environments. Our research emphasizes the need for stringent data regulation to ensure the ethical use of digital educational data. Overall, our findings contribute to improved design, analysis, and evaluation practices in digital learning.

## 2 Conceptual Background

### 2.1 Concepts and Applications of Data and Process Mining

Since the 1980s, the application use cases and relevance of data mining in business have substantially increased (Chen et al. 2012). We follow the definition of Bissantz and Hagedorn (2009, p. 118), who define data mining as *"the extraction of implicitly available, non-trivial, and useful knowledge from large, dynamic databases with relatively complex structures."* By facilitating better monitoring, identification, and analysis of raw business data, data mining aims to automate data preparation and analysis (Bissantz and Hagedorn 2009). Process mining intends to help organizations improve their understanding of processes by means of the identification of variants of a business process, identification of noncompliant behavior, performance of relevant insights, or assurance of the quality of discovered process model (van der Aalst et al. 2012, vom Brocke et al. 2021).

The application of data mining in the educational domain referred to as educational data mining evolved as a result of the availability of educational data collected by TML environments (Lemay et al. 2021). A related concept is learning analytics (see Fig. 1 for an overview and proposed demarcation of related concepts). Learning analytics aims to discover useful insights into a user's learning behavior from large educational data sets and activities, that is, the discovery of information (i.e., what is the current sentiment of the learner), generation of learning progress (i.e., what is the learner's relative and absolute progress) or understanding learner behavior (i.e., what are the personality and learning routines like) (Nguyen et al. 2020a). Hence, learning analytics embodies a learner-centered perspective (e.g., employ analytics to inform or empower instructors or learners about learner gaps or preferences). Learning analytics applications focus on evoking data-based interventions for learners to enhance their learning behavior (Chatti et al. 2012). Methods of learning analytics include classifications and predictions to monitor, individualize, and enhance learning behavior, the visualization of learning data and insights, social network analysis to identify relationships between learners, and optimization of individual learning activities (Chatti et al. 2012). Regarding the data sources, learning analytics can process various large data sets that are generated by TML environments and also other sources, such as social networks or physiological data.

Educational process mining can be viewed as a subfield of learning analytics, which uses process mining techniques to add a process-centered perspective on learning behavior (Bogarín et al. 2018; Juhaňák et al. 2019; Cerezo

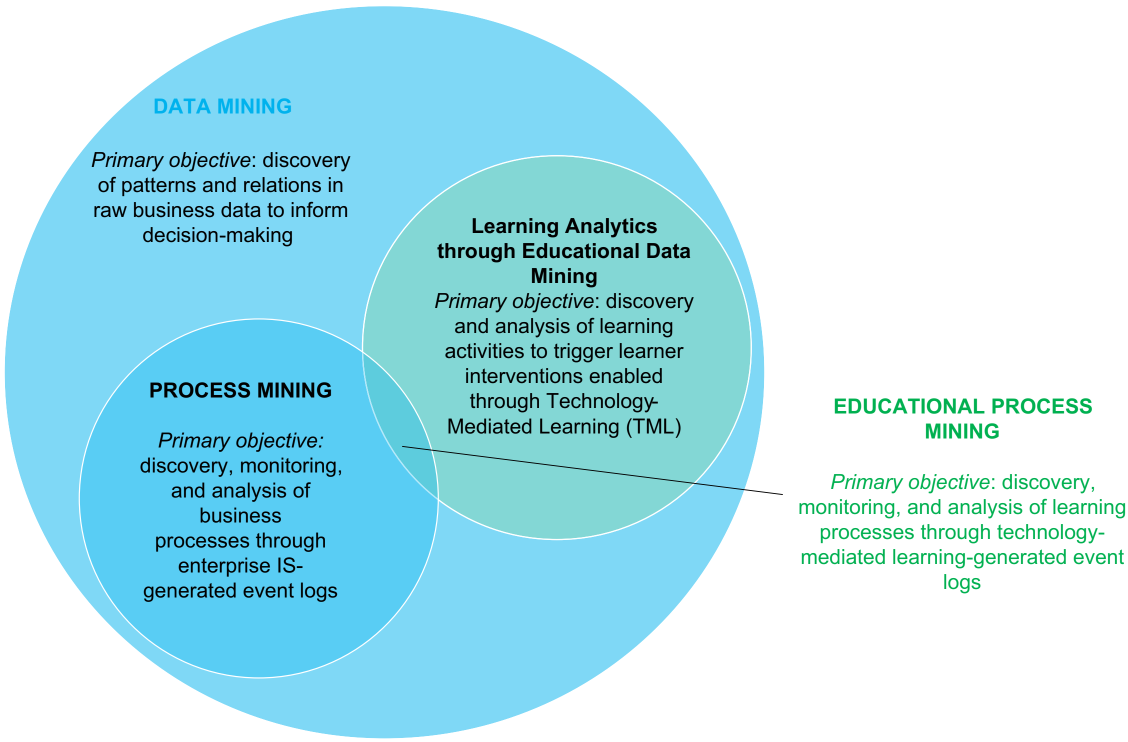


**Fig. 1** Overview and proposed demarcation of related concepts of educational process mining

et al. 2020). More specifically, educational process mining aims to analyze log data from TML environments to monitor, discover, and enhance learning processes (Cerezo et al. 2020). Its unique characteristic is that it specifically deals with the analysis of educational data logs and, therefore, aims to better understand learning processes, that is, learning process discovery, social network discovery, process monitoring, process analysis, or the enhancement of processes and flows. Instead of analyzing single data points of the learner's journey, educational process mining's process-oriented perspective assists in uncovering the entire end-to-end learning process and the dependencies among the activities in the process, that is, the learner's success or other measures of success.

In the following, we refer to the three major educational process mining activities of *discovering, monitoring,* and *enhancing* learning processes and activities. Consequently, educational process mining uses analytical methods for educational data. The academic discourse around educational process mining is increasing because this subfield of learning analytics offers a new, holistic, process-centered perspective on learning. Such a dynamic perspective is beneficial for individuals, educators, and organizations alike. The focus on learner processes during their educational development is a pedagogically beneficial perspective and has been well described in literature (Roth 1970; Söllner et al. 2018). Thus, educational process mining is a technology that could bring back the learning-process-centered perspective developed by Roth (1970) and thus offers a process-centered perspective in the analysis of learning paths. Figure 2 illustrates an architecture of relevant scenarios in practice.

User data is generated automatically as a learner interacts with the learning environment (such as the system in Fig. 2) through clicks, chat protocols, review comments, or specific learner content, such as written student texts. These event logs are then used to fuel the learning process model, which is used to discover, to conform, or enhance the system. With the continued expansion of TML, an enormous amount of student-centered data can enrich the pedagogical embedding itself (e.g., Wambsganss et al. 2021) through personalization, recommendations (Nguyen et al. 2020b), and formative feedback (Hattie and Timperley 2007). As well it can also support the educators and institutions who monitor and evaluate learning processes by extracting information from the models and acting on the findings.

Three basic types of process mining can be distinguished (van der Aalst 2012; Bogarín et al. 2018):

- *Process discovery* modeling and visualizing student learning processes, for example, to track a student's individual learning journey or the curricular path a student takes.
- *Conformance checking* determining whether an observed learning process conforms to a pre-defined

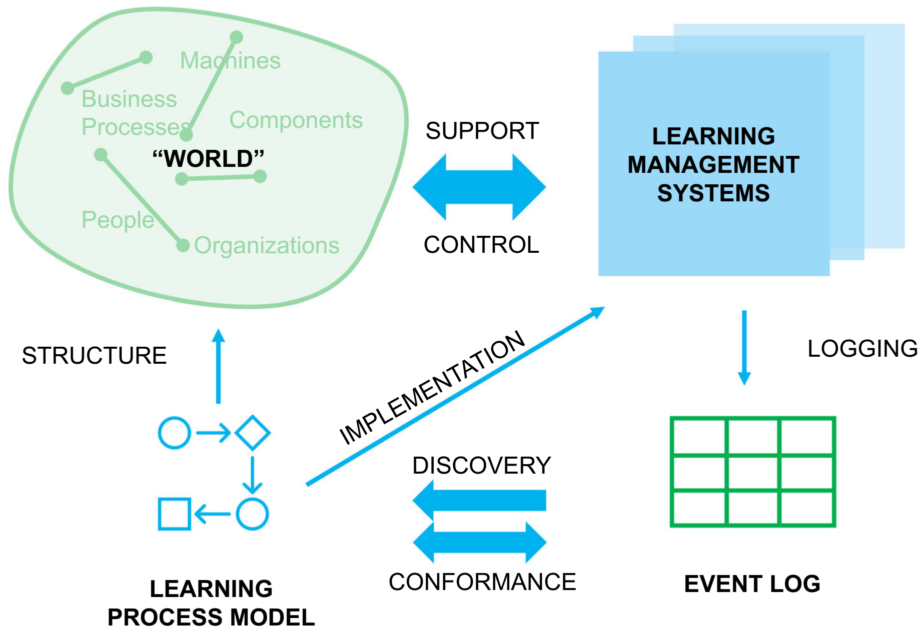


**Fig. 2** Overview of an architecture of educational process mining scenarios according to Bogarín et al. (2018)

learning process model in order to, for example, identify weaker students (outliers) or assess compliance with guidelines and prescriptions.

- *Process enhancement* extending a given learning process model using information extracted from a specific event log related to the same process, for example, to detect bottlenecks or provide students with adaptive feedback on their process.

Data mining and process mining, and more specifically, their subfields of learning analytics and educational process mining methods, can be used in the educational domain to enrich learning scenarios with data-driven value. However, this requires an understanding of the anatomy of TML (a general overview can be found in Gupta and Bostrom 2009). With their call for TML research 20 years ago, Alavi and Leidner (2001) defined TML as "*an environment in which the learner's interactions with learning materials, peers, and/or instructors are mediated through advanced information technology*" (Alavi and Leidner 2001, p.2). As Gupta and Bostrom (2009) noted, TML includes "*all the elements of a social-technical system: technology and learning techniques, process, actors, actions, and outcomes*" (Gupta and Bostrom 2009, p.3). Scholars in different disciplines have used the terms e-learning, online learning, distance learning, technology-enhanced learning, or IT-supported learning synonymously (Gupta and Bostrom 2009). TML is based on the underlying assumptions that a particular learning context influences the embedded learning structures (DeSanctis and Poole 1994).

In this paper, we follow a constructivist understanding of learning, which describes how knowledge is always constructed by the learner and can be influenced by experiences (Vygotsky 1980). In this vein, the learning experience is influenced by a learning method. The framework of TML focuses on the learning process of a learning activity. By definition, a learning process describes the sum of the individual activities a student performs. According to Gupta and Bostrom (2009), it describes how a student interacts with and adapts the learning method structures. A learner's process is a complex phenomenon and includes variables such as cognitive processes and interactions related to the learning methods. Gupta and Bostrom (2009) define learning outcomes as the pivotal measurement of the success of a learning process in TML. Accordingly, learning outcomes are "*the target assessment or measures used to determine the achievement of learning goals*" (Gupta and Bostrom 2009, p. 691), e.g., in a certain individual learning scenario. With "individual learning scenario", we mean a certain pedagogical set-up where a learner receives adaptive feedback on her own learning progress, for example, from a human educator (or artificial tutoring system) that recognizes learner gaps, anticipates needs, suggests suitable learning materials, and provides individualized feedback.

Educational process mining can contribute to different elements of the TML framework. It can have a structural impact (e.g., an impact on learning method structures as a technology that directly influences the learning process of students) or a process impact (e.g., by using it to better measure appropriation or learning outcomes).

### 2.2 Related Work on Technology-Mediated Learning and Learning Processes

Literature on TML has mostly focused on the implementation and analysis of personal and adaptive learning environments, e.g., based on adaptive formative feedback (Hattie and Timperley 2007; Ifenthaler and Gibson 2019). Here, perspectives in learning analytics, such as educational process mining, can enhance the adaptivity or the possibility of individual feedback and recommendations in TML environments (Bogarín et al. 2018). Nevertheless, literature has in the past mostly approached the use of process mining for educational purposes from a technical perspective. For example, Mouchel et al. (2023) used process mining to better understand the learning and revision behavior of students who received intelligent writing support (Mouchel et al. 2023). Ludwig et al. (2024) applied a sequence mining approach to behavioral process data to predict problem-solving success of students, which ultimately allows instructors to better intervene with personalized scaffolding. Another example is He et al. (2021), who applied different process mining techniques, such as Fuzzy miner or pMiner, to 122,167 event logs from 527 undergraduate students extracted from the learning management system Moodle in order to track students' self-regulated learning patterns in response to a formative assessment (He et al. 2021). Also, research in information systems and educational technology has motivated research and studied the effect of various learning scenarios and configurations based on educational process mining (e.g., Johnson et al. 2019; Cerezo et al. 2020). This includes, for example, the investigation of the impact of metacognitive prompts on self-regulated learning (Engelmann and Bannert 2019), the investigation of student adherence to a recommended course path (Cameranesi et al. 2017), and the investigation of process-based feedback during medical training (Lira et al. 2019). The expanding number of interdisciplinary studies on educational process mining emphasizes the importance of a holistic understanding of its design, capabilities, and potential repercussions (Bogarín et al. 2018; Costa et al. 2020).

Still, research is distributed over a wide range of socio-technical and technological viewpoints, resulting in a critical lack of an integrated perspective. For instance, in their systematic literature review, Ghazal et al. (2018) did not derive specific educational process mining traits and dimensions and instead concentrated on technical factors. Bogarín et al. (2018) have presented a systematic summary; however, they did not use a clear methodological approach, such as a systematic literature review (Webster and Watson 2002; vom Brocke et al. 2015). Costa et al. (2020) only considered process mining literature on Moodle, making the review inapplicable to other learning scenarios. Assessing the potential of educational process mining for individualized learning scenarios and deriving suitable design characteristics requires an integrative perspective based on TML and the potentials described before. Here, classifications of educational process mining are beneficial for developing tailor-made use case applications for different stakeholders (e.g., educators, institutions, or learners). Several scholars have highlighted the lack of shared information about the embedding of process mining in various TML scenarios, as well as the lack of insights into the involved users (e.g., educator, institution, or learner), the technological context (e.g., the discovery algorithm used), or the pedagogical structure (e.g., in which educational domain or for which pedagogical task process mining is analyzed (Ghazal et al. 2018; Rogiers et al. 2020). In fact, present reviews fall far short of providing a complete and solid structure for educational process mining applications. A systematic classification of educational process mining scenarios from a TML would allow researchers to more effectively design, evaluate, compare, and theorize how different technological embeddings of the young field of process mining impact student learning outcomes in a specific learning scenario and task. A TML perspective would allow the classification of a certain information system into key aspects (user, task, structure, and technological standpoint), which results in varied configurations and outputs (Bostrom and Heinen 1977; Gupta and Bostrom 2009). We want to fill this knowledge gap by deriving a unique taxonomy that aids decision-making when developing, designing, and evaluating relevant scenarios and applications, as well as by specifying the links of educational process mining characteristics to the outcome of a learning scenario. Educational process mining can serve as a key for creating and establishing individualized learning scenarios and help information systems researchers and educators design and analyze digital learning processes more effectively. Here, the design and analysis of learning processes refers to the systematic creation and evaluation of educational strategies and methods to enhance student learning experiences and outcomes. This involves developing educational interventions (design) and assessing their effectiveness and impact on learning (analysis).

## 3 Research Methodology

To answer RQ1, we systematically classify the objects of interest for designing and analyzing learning processes with educational process mining (Sect. 4). Therefore, we follow a taxonomy development process (Nickerson et al. 2013; Kundisch et al. 2021) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-

Analyses, Mother et al. 2009) (Steps 1, 2, and 3 in Fig. 3). RQ2 is answered based on the result of the systematical literature review from RQ1, by clustering the resulting empirical papers about the design and analysis of learning processes with educational process mining and collecting additional empirical data from field interviews to dive deeper into the design patterns of recurring challenges (Steps 4 and 5 in Fig. 3).

### 3.1 Step 1: Database Creation Through a Systematic Literature Review

We conducted a literature review to identify relevant material as the foundation for the systematic construction of a taxonomy according to Webster and Watson (2002), vom Brocke et al. (2015), and PRISMA from Mother et al. (2009). Based on recent literature reviews on educational process mining (e.g., Bogarín et al. 2018; Costa et al. 2020), we identified different keywords that researchers used to describe process mining in the educational domain.Based on the keywords, we created the following search string*: [“Process Mining” OR “Workflow Mining” OR “Task Mining”] AND [“Education” OR “Learning Analytics” OR “Training” OR “Skill Development” OR “Student” OR “Teaching” OR “Learner” OR “Pedagogic” OR “University”]*. We selected two major areas for educational process mining research: information systems and educational technology. They cover a sizable portion of the literature on our topic of interest. To find relevant studies that applied process mining in educational scenarios, we applied the search strings to the following six databases: AISeL, EBSCO, ScienceDirect, ProQuest ABI Inform, IEEE Xplore, and ACM Digital Library (see Fig. 4).

The process of applying the four steps of PRISMA (Mother et al. 2009), namely *Identification, Screening, Eligibility,* and *Inclusion*, is presented in Fig. 4, along with justification for the inclusion and exclusion of studies in each step.

We identified 3028 studies from six databases in January 2024. After the initial screening of titles, abstracts, and keywords and removing of duplicates, we counted N = 2944 studies. Next, we reviewed the studies to ensure that they met the selection criteria and applied to the subject of our study. We eliminated studies that did not mention process mining or that used process mining in an area other than education. Many manuscripts were rejected because they represented a distinct study scope. Several studies, for example, addressed mineral mining or machine learning and process mining outside of an educational environment and were thus excluded from our sample. This filtering yielded 77 empirical studies for review that referenced using process mining in education throughout their research. Following that, a forward and backward search

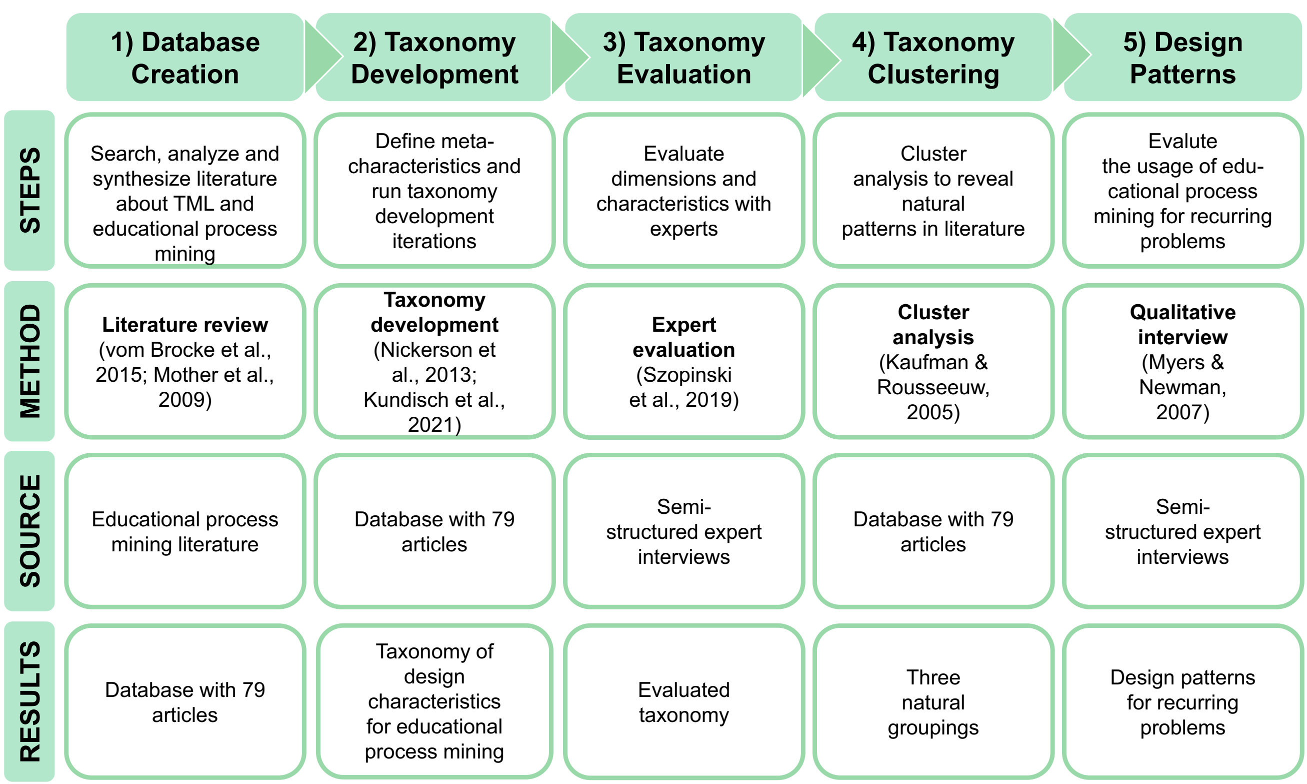


**Fig. 3** Overview of our five consecutive research steps

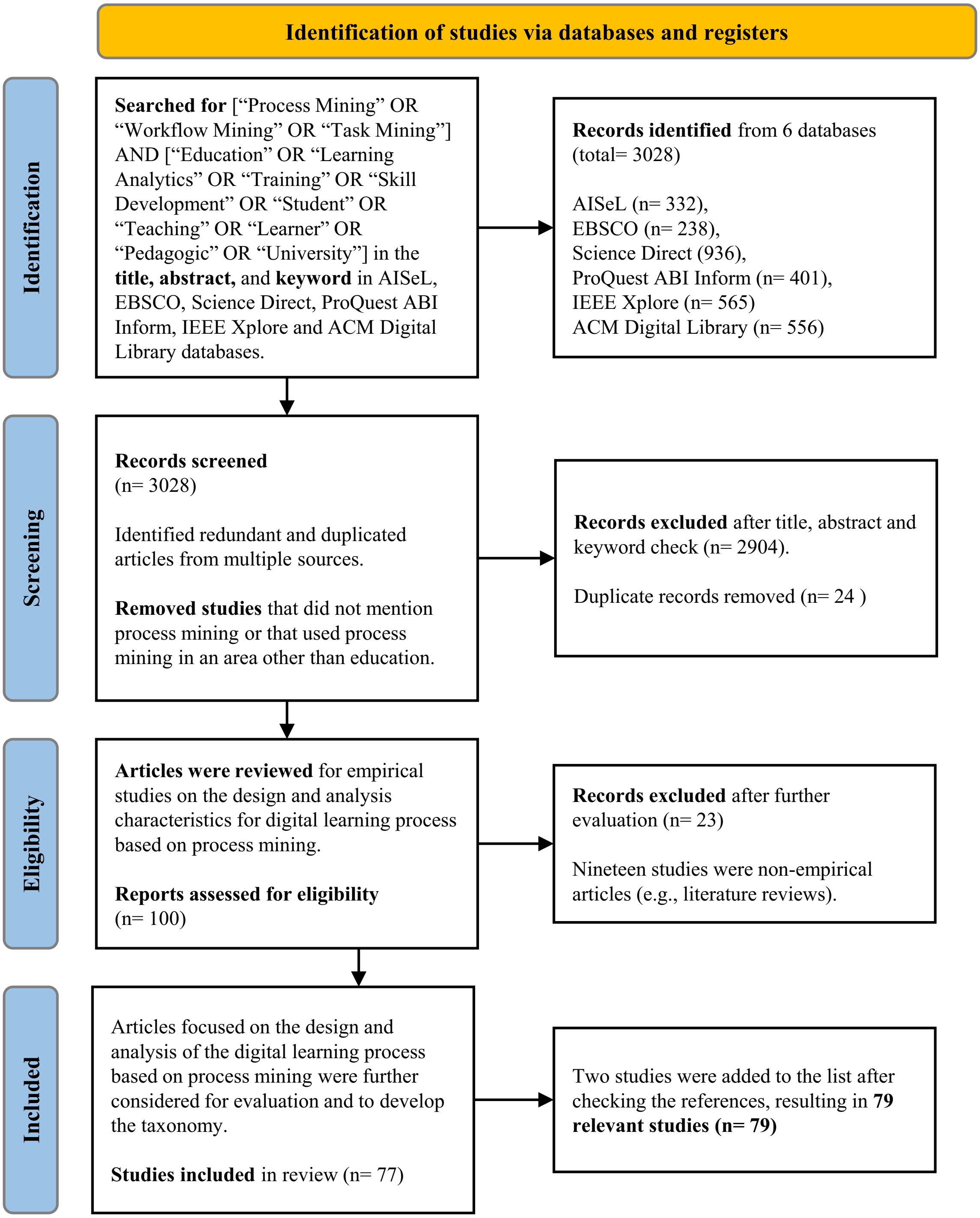


**Fig. 4** Systematic review of articles using PRISMA (Mother et al. 2009)

was performed following Webster and Watson (2002). Two publications were added to the list after checking the references, resulting in 79 relevant papers.

### 3.2 Step 2: Taxonomy Development

Our goal was to offer a holistic yet nuanced framework on educational process mining. We used the taxonomy development method of Nickerson et al. (2013), which has been used in several previous information systems studies (e.g., Singh and Varshney 2020). Their approach to constructing taxonomies is iterative and systematic and is based on theoretical underpinnings (deduction) and empirical data (induction). We constructed several dimensions and characteristics based on the published research regarding process mining in education and the empirical evidence of certain meta-attributes. The purpose of a meta-characteristic is to *systematically identify design and analysis elements for a digital learning process based on educational process mining applications to better identify, design, and compare process mining applications in educational contexts*. To do so, we looked at educational process mining design characteristics from a TML perspective (Gupta and Bostrom 2009) to form a holistic contribution to the current knowledge of process mining in the specific domain of education.

#### *3.2.1 Paper Analysis*

The 79 relevant publications were examined from a concept-centric standpoint using an abductive technique. We created a list of master codes and descriptions to reflect various situations to consolidate the findings from recognized educational process mining investigations. These master codes and their development are depicted in Table 1. We followed the TML standpoint for the codes to identify design characteristics of process mining applications (Alavi and Leidner 2001; Gupta and Bostrom 2009) (i.e., user, task, technology, and structure). Hence, we started with the definitions and dimensions of user, task, technology, and structure provided by Alavi and Leidner (2001) as well as Gupta and Bostrom (2009). The iterative procedure necessitated numerous rounds of coding of the discovered publications. The procedure began with three researchers independently coding a subset of ten randomly selected articles. We described the situation in which process mining was utilized and defined the distinct design elements based on the employed technology, the learning environment, the user, and the overall pedagogical framework for each of these ten studies. We held a workshop to explore how to incorporate process mining design elements across studies, which resulted in a unique set of characteristics and descriptions. During the subsequent rounds, one researcher coded 25 articles based on the previous list and definitions. Following that, three researchers and two practitioners from a process mining software provider convened to discuss the findings. If the coding was ambiguous, the process mining characteristics and descriptions were debated and revised until a consensus was established. We added process mining qualities to our list in each cycle, based on the TML dimensions and descriptions, until all the articles were coded.

#### *3.2.2 Taxonomy Iterations*

Nickerson et al. (2013) provide many subjective and objective criteria, also known as ending conditions, that a taxonomy must meet at the end of the iterative taxonomy-building process. To identify when to halt the iterative process, we developed the following ending criteria (EC).

(A) At least one object (educational process mining scenario) is categorized under each dimension's characteristic.

(B) No new dimension or characteristic was added in the previous iteration.

**Table 1** Taxonomy development iterations and master codes based on Nickerson et al. (2013)

| Iteration | Approach | Taxonomy | EC met |
|---|---|---|---|
| 1 | Conceptual-to-empirical | $T_1$ = {Task (intended timeframe, interaction duration, focus of analysis, area of implementation), Structure (learning context, learning outcome, cognitive domain, application domain), User (main end user, objective user measurement, subjective user measurement), Technology (process mining type, tool, discovery algorithm, data collection interface, visualization, data format)} | A, C |
| 2 | Empirical-to-conceptual | $T_2$ = {Task (area of implementation, primary learning task), Structure (learning context, primary learning outcome, cognitive domain, application domain), User (main end user), Technology (process mining type, data collection interface, visualization of process mining results, data input)} | A, C |
| 3 | Empirical-to-conceptual | $T_3$ = {Task (area of implementation, learning task), Structure (learning context, learning outcome), User (intended main end user, involved learners), Technology (data input beyond event logs, data collection interface, process mining type, analysis beyond process mining, output representation)} | A, C |
| 4 | Empirical-to-conceptual | $T_4$ = {Task/Structure (area of implementation, educational level, learning outcome, learning task), User (intended main end user, involved learners), Technology (data input beyond event logs, data collection interface, process mining type, analysis beyond process mining, output representation)} | A, C |
| 5 | Empirical-to-conceptual (expert interviews) | $T_5$ = {Purpose (application focus, learning mode, learning outcome, learning task), User (intended main end user, learning context), Data (data input beyond event logs, data collection interface), Analysis (process mining type, analysis beyond process mining, output representation)} | A, B, C, D |

(C) Dimensions and traits are distinct and do not reoccur.
(D) The taxonomy classifies every known object.

The final taxonomy should satisfy all the finishing requirements. We began with a conceptual-to-empirical cycle and then moved on to four empirical-to-conceptual cycles. For example, for the first conceptual-to-empirical cycle (iteration 1), we started with the theoretical concepts (conceptual) of TML (Alavi and Leidner 2001; Gupta and Bostrom 2009) (i.e., task, structure, user, and technology) and mapped them to the first empirical papers (empirical) retrieved from our database. In the next cycle (iteration 2), we then conducted an empirical-to-conceptual cycle where we refined (e.g., changed, removed, or added) the concepts based on the additional literature from our database. Table 1 depicts the progression of our taxonomy. The fifth iteration presents the results from our expert-interview evaluation further explained in the next section. The column "Taxonomy" depicts the master codes. We categorized each of the 79 studies five times until all ending conditions were fulfilled.

To sum up, we reviewed and organized research suggesting that an individual's learning process can be characterized along four main dimensions: purpose, user, data, and analysis. The purpose and user categories refer to design options that, for example, a lecturer might take when creating digital learning processes for TML scenarios. They focus on learning objectives and learning task design aspects (also known as learning activities) in a specific application context of the digital learning process. The human-centered viewpoint on the social embedding of the pedagogical TML scenario is included in the purpose and the user dimension. In contrast, the data and analysis dimensions refer to analysis options of data from digital learning processes and focus on data prerequisites and suitable evaluation procedures. Data and analysis cover the technical embedding and analysis of process mining features, the sorts of data analyzed, and how the analysis results are conveyed (Bostrom and Heinen 1977; Gupta and Bostrom 2009). The distinction among these four key dimensions improves the taxonomy's applicability in terms of cluster analysis of the reviewed studies and the resulting design patterns. To that end, we strive for a precise and unambiguous explanation of the various classes for each dimension to enable robust categorization.

### 3.3 Step 3: Taxonomy Evaluation Based on Semi-structured Interviews

To assure the quality of our taxonomy, we evaluated it using the five listed criteria: *comprehensibility, conciseness, robustness, extendibility,* and *explanatory power* (Nickerson et al. 2013). We conducted semi-structured interviews with three experts in the analysis of digital learning processes with educational process mining and three experts in the design of TML environments following the taxonomy evaluation guidelines of Szopinski et al. (2019) (see Table 2). Both groups represent experts in the domain and potential end users of our taxonomy. The shortest interview was 30 min, while the longest lasted 85 min. The interview guideline included 18 open-ended questions based on the five listed assessment criteria. Before the interviews, we emailed the final copies of our taxonomy, the meta-characteristic, and exemplary process mining scenarios to the interviewees. We asked interviewees to remark on and provide ideas for modifying and refining the taxonomy during the interviews. We redesigned the taxonomy after obtaining some ideas for dimension categorization and other minor improvements from the experts, which resulted in the final taxonomy. Following the expert interviews, several significant changes were made to the taxonomy. Originally, the taxonomy dimensions were categorized as T4 (see Table 1): *Task/ Structure*, *User*, and *Technology*. A majority of interviewees gave us detailed feedback on the explanatory power of our dimensions and characteristics. For example, interviewees 1–5 suggested using more self-explaining terms for educators and researchers to understand the overall structure of the taxonomy given our meta-characteristics. Hence, based on interviewees 3 and 4 suggestions we renamed and reorganized the dimension of *Task/ Structure* to *Purpose* to encapsulate a wider scope, including the application focus and learning mode. This change reflects the feedback that the taxonomy should capture the broader objectives and instructional methods involved in educational process mining. Or, based on the suggestion of interviewees 1 and 2, we split the original *Technology* dimension into *Data* and *Analysis*. This change was driven by the need to highlight the importance of data quality and sources more prominently and to better reflect the variety of analytical approaches used in process mining.

These changes were made to ensure the taxonomy is both comprehensive and practically applicable, addressing the identified gap between theory and practice. This redesigned taxonomy, informed by expert insights, better captures the complexities and varied aspects of educational process mining.

### 3.4 Step 4: Taxonomy Cluster Analysis

We conducted a cluster analysis on the 79 systematically identified studies on learning process design and analysis (Kaufman and Rousseeuw 2005). Our goal was to find clusters to better comprehend the use of certain educational process mining application groups. We constructed a binary data matrix indicating the presence (1) or absence (0) of

**Table 2** Expert panel for taxonomy evaluation

| No | Function | Organization | Expertise in |
|---|---|---|---|
| 1 | Educational designer | University | Teaching process mining use with SAP ERP (enterprise-resource-planning) |
| 2 | Senior software engineer | Celonis SE | Technical development of process mining applications and scenarios |
| 3 | Researcher | University | AI-supported learning and pedagogical design |
| 4 | Researcher | | Research on TML and pedagogical design |
| 5 | Senior educational designer | Celonis SE | Process mining education and design of learning scenarios |
| 6 | Educational designer | University | Process optimization, lecturer six sigma MOOC edX |

specific educational process mining characteristics identified from literature across different studies. These characteristics, aligned with our taxonomy's dimensions, and the studies, ordered alphabetically by first author, formed the rows and columns of the matrix, respectively. The database consisted of the final eleven dimensions and 45 characteristics from our final taxonomy (see Sect. 4 and Fig. 5). Using Ward's algorithm (Ward 1963), known for its efficacy with smaller datasets, we conducted an agglomerative clustering with a Euclidean metric to find natural groupings and patterns in the data. We discovered three natural clusters of learning scenarios based on educational process mining. These clusters lay the empirical foundation for formulating the patterns for designing and analyzing the digital learning process with process mining.

## 3.5 Step 5: Interview Study and Design Patterns

In the fifth step, we conducted semi-structured expert interviews to dig deeper into the solution patterns for recurring problems in designing and analyzing digital learning processes based on educational process mining. Equipped with the resulting dendrogram, the clustered data, and the taxonomy of learning process design and analysis, we interviewed eleven experts in digital learning design analysis. We applied the guidelines of Myers and Newman (2007) for qualitative interviews for rigorous execution and analysis. We conducted expert sampling following Bhattacherjee (2012) to select suitable interview partners. We chose experts who are experienced educators or hold a role in an educational institution (e.g., program coordinator) and conduct research on the design and analysis of TML learning scenarios. Table 3 provides an overview of our interviewees (E1–E11). We conducted 11

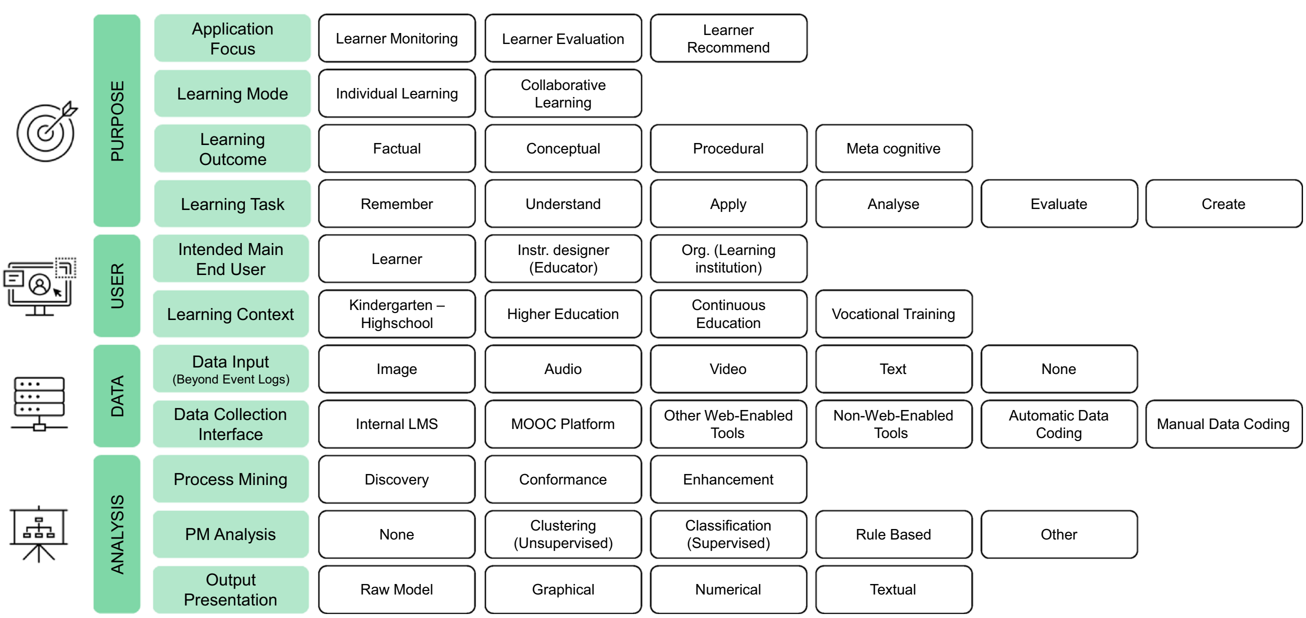


**Fig. 5** Taxonomy for designing and analyzing digital learning processes with process mining

interviews from September until October 2022. The interviews lasted between 31 and 57 min.

Our interview guideline consisted of 28 questions, asking the experts about their general experience in TML, including how they plan, design, conduct, monitor, and improve digital learning scenarios. Moreover, we specifically asked questions according to each dimension and characteristic of our taxonomy (without showing our taxonomy) and about the requirements of solutions to recurring problems in the design and analysis of the digital learning process based on our cluster analysis. In this vein, we dug deeper into the nuances of each cluster and derived the design patterns presented in Sect. 5 (see Fig. 6, 7, and 8). The transcriptions of the interviews were evaluated using qualitative content analysis. Hence, the data were coded, and abstract categories were formed. The coding was performed using open coding to form a uniform coding system during the valuation. We derived several topics in each interview and aggregated the most common ones, resulting in five common topics for the design and analysis of the digital learning process. The codes and topics were "recurring challenges" of educational processes mining (e.g., discovering course-taking behavior or predicting outcomes to provide students with ongoing feedback), "influential factors and requirements" to solve the challenges (e.g., data quality or legal requirements), "exemplary success stories" (e.g., tracing learning behavior to prevent student dropouts or modeling student outcomes to provide feedback) and "design and evaluation steps" to implement the solutions (e.g., concrete steps to discover, monitor, and enhance learning scenarios with educational process mining). Based on these codes, we derived three patterns for designing and analyzing the digital learning process that build on our taxonomy and informed based on a field evaluation with eleven experts in the domain of TML. To structure the design patterns, we follow a similar table format of Dickhaut et al. (2022) and Weinert et al. (2021), since this provides a validated structured, and effective way of presenting the design patterns to different use groups. The categories are based on our codes and

**Table 3** Overview of expert interviews

| Expert # | Position | Expertise | Experience | Time in minutes |
|---|---|---|---|---|
| E1 | Researcher | Design of digital interactive learning scenarios | 3 years of teaching and research experience in Vocational training design | 57 |
| E2 | Researcher | Pedagogical theory for the design of TML scenarios | 1 year of teaching experience, 2 years of research experience on socio-technical system design from a pedagogical point of view | 37 |
| E3 | Researcher and Educator | General research on TML in information systems | 1 year experience in teaching, 1 year experience in research on TML | 34 |
| E4 | Researcher | Design of digital learning process for information systems education | 3 years of experience in teaching, 3 years of experience in information systems research | 35 |
| E5 | Researcher and teaching assistant | Data analysis of behavioral student data | 8 years of experience as a teaching assistant, 3 years learning analytics research | 40 |
| E6 | Senior researcher and educator | Data analysis, design, and evaluation of digital learning processes | 10 years of teaching experience, 10 years of research experience on human–computer interaction and educational technology | 42 |
| E7 | Senior researcher and educator | Information systems researcher with a focus on education and learning analytics | 8 years of experience in teaching, 12 years of experience in the design and analysis of TML scenarios | 47 |
| E8 | Senior researcher and educator | Information systems researcher with a focus on education and learning analytics | 9 years of experience in teaching, 10 years of experience in research on the design of digital learning scenarios and scaffolding | 24 |
| E9 | Senior researcher and educator | Process mining of behavioral data | 5 years of experience in teaching, 7 years of experience in research in data mining and analytics | 26 |
| E10 | Program coordinator (undergraduate level) | Program coordination and organization of digital learning at scale at the university level | 4 years of experience in the coordination and evaluation of digital learning scenarios | 31 |
| E11 | Program coordinator (graduate level) | Design science research and program coordination at the university level | 4 years of experience in information system research, 5 years of experience in teaching | 49 |

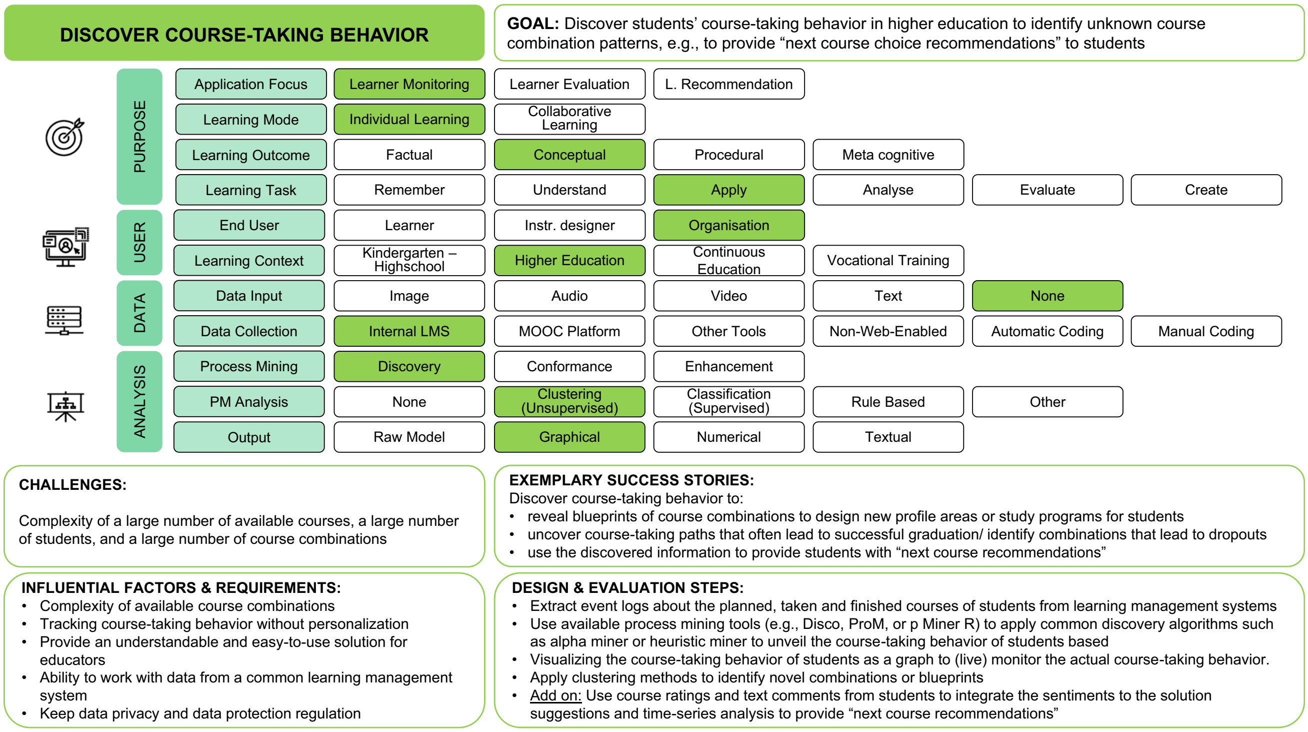


**Fig. 6** Design pattern "discover course-taking behavior" with process mining

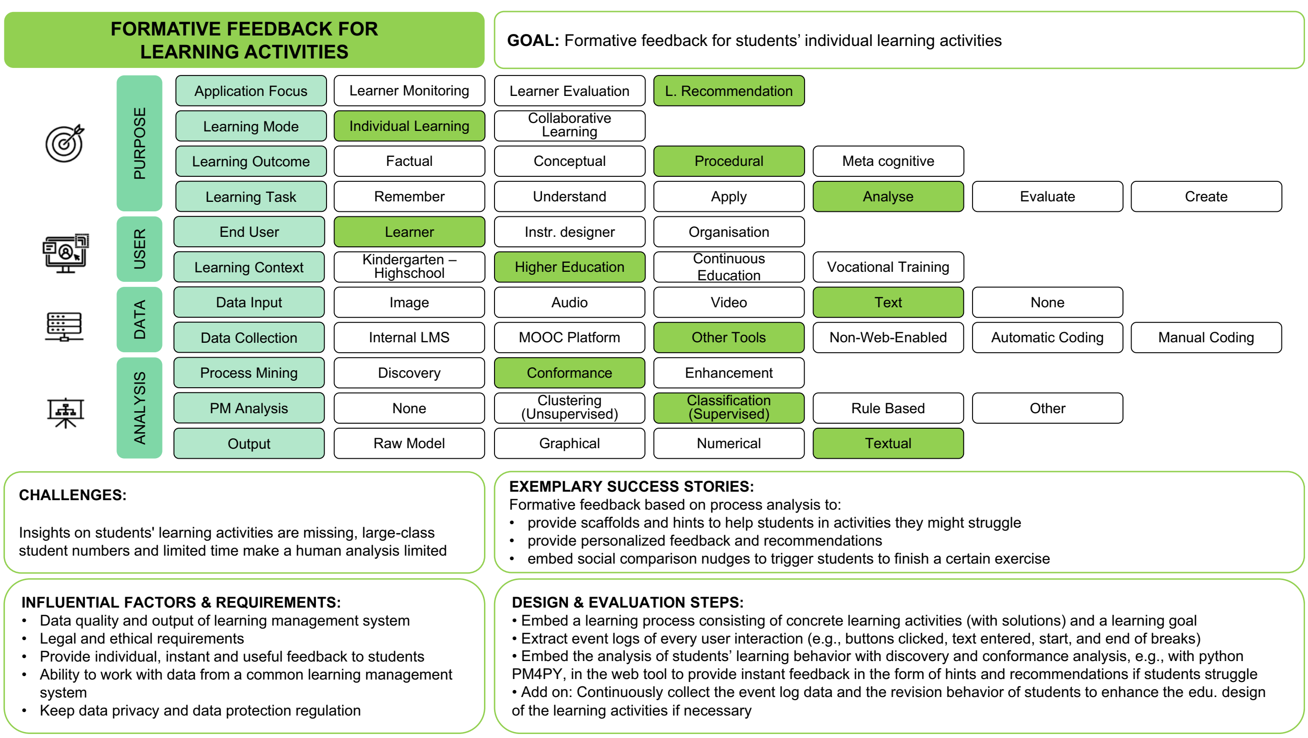


**Fig. 7** Overview of design pattern "formative feedback for learning activities" with process mining

include "name", "goal", "pattern based on taxonomy", "challenges", "influential factors & requirements", "exemplary success stories", and "design and evaluation steps".

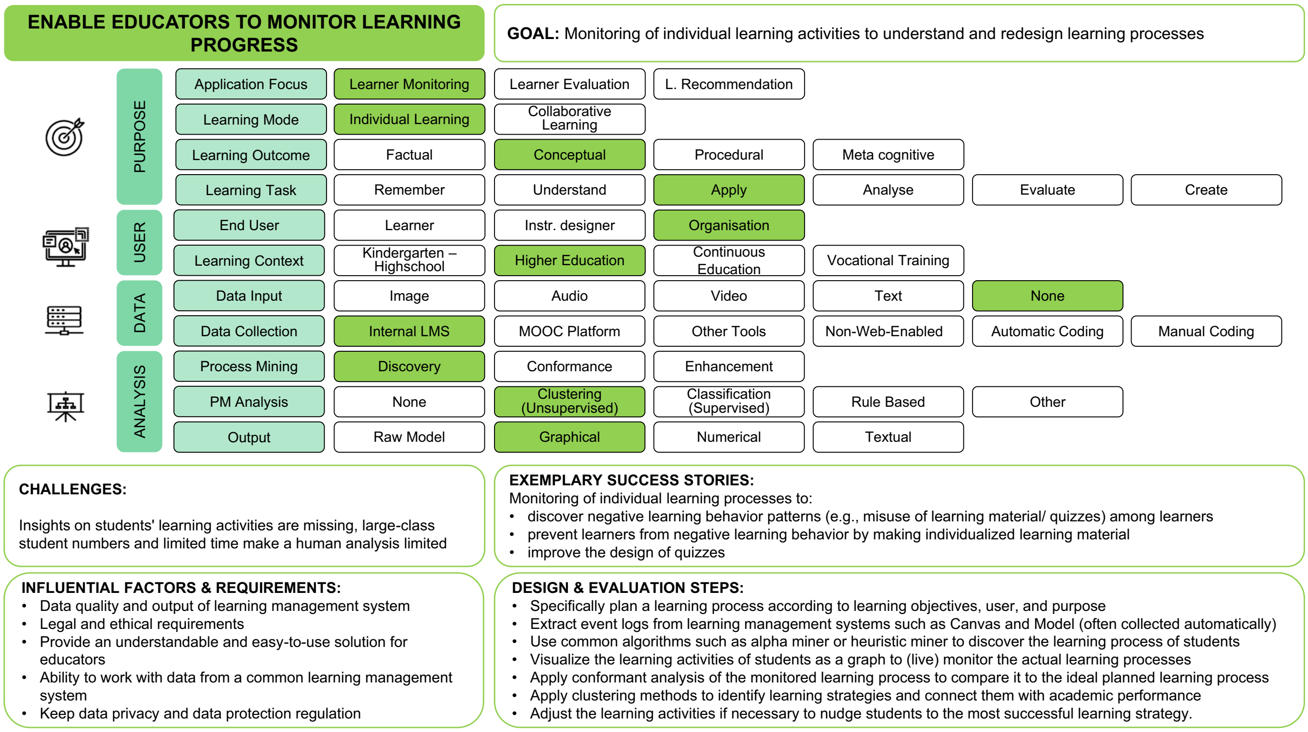


**Fig. 8** Overview of design pattern "enable educators to monitor learning progress" with process mining

## 4 Taxonomy of Digital Learning Process Design and Analyses

In this section, we present the results of our taxonomy development process. We derived a taxonomy represented as a matrix that provides clear classifications of related characteristics and dimensions for designing and analyzing digital learning processes with process mining (see Fig. 6). An important element of our taxonomy and theorizing is that valuable elements may be associated with the design and analysis of student learning processes. TML informs the taxonomy and patterns in the next section, adding an important aspect to the discussion of how a digital learning scenario can be designed and orchestrated.

### 4.1 Dimension 1: Purpose of Designing and Analyzing Digital Learning Processes

Dimensions referring to the design options of digital learning processes are concerned with the ***purpose*** that a digital learning process should achieve. In this light, we distinguish between a*pplication focus, learning mode, learning outcome, and learning task.*

The *application focus* dimension outlines the learning behaviors and preferences that an instructor hopes to reveal and analyze using an educational process mining application. *Learner monitoring* refers to situations in which an instructor examines a learning process to acquire insights. An example is Uzir et al.'s (2020) use of educational process mining to monitor whether learners comprehend the offered learning techniques. In that sense, educational process mining can offer valuable feedback on learner development, oftentimes on an aggregated group level. *Learner evaluation* refers to situations in which educational process mining provides information on and assesses individual learning processes (e.g., bad task performance, clarification issues). In that sense, the evaluation looks back on an individual's past performance or identified issues in the past process to create options for process-based feedback. Lira et al. (2019) provide an example by investigating process-based feedback during medical training. An application of process mining in the domain of *learner recommendations* is presently not addressed by the educational literature but would define a system that provides learners with actionable suggestions on how to proceed with their learning process (e.g., based on scaffolding, see Winkler et al. 2021). Hence, different from a learner evaluation purpose, a learner recommendation looks forward into the digital learning process before it is completed. The purpose is to provide learning-process-integrated recommendations.

The *learning mode* dimension outlines the (scope of the) unit of analysis of digital learning processes in more detail. We discovered that application scenarios of digital learning processes that are suitable for educational process mining focus on *individual learning*, such as when learners study on their own and simply engage with learning resources (Saint et al. 2020) or *collaborative learning*, which occurs when learners interact, collaborate, and coordinate learning activities with other learners (Schoor and Bannert 2012). This distinction is seen as a component of the learning mode dimension. The *learning outcome* dimension refers to the knowledge type that should be trained or created while completing various activities of a digital learning process. Therefore, this dimension recognizes that educational process mining is usable for digital learning processes that focus on *factual*, *conceptual*, *procedural,* and *metacognitive knowledge.* The characteristics are based on Krathwohl (2002), who defined factual (knowledge) as knowledge of terminology and precise information (e.g., Cerezo et al. 2020), conceptual (knowledge) as an understanding of theories and models, and procedural (knowledge) as subject-specific skills and techniques (e.g., Lira et al. 2019). Metacognitive (knowledge) refers to both strategic and cognitive knowledge, such as debating abilities (Wambsganss et al. 2021). The *learning task* dimension refers to the learning objectives that learners should achieve when following a digital learning process. We employed characteristics influenced by Krathwohl's (2002) classifications to differentiate among different difficulty levels of a learning assignment in the form of learning objectives. We used Krathwohl's concept and differentiated between *remembering, understanding, applying, analyzing, evaluating,* and *creating* as learning objectives that a learner can attain during a learning process (Krathwohl 2002). We assume that it is easier to capture data for simpler and objectifiable objectives, such as remembering and understanding.

### 4.2 Dimension 2: Users When Designing and Analyzing Digital Learning Processes

The second set of dimensions refers to design options of digital learning processes and is concerned with the ***user*** who will benefit from analyzing a digital learning process. Despite the field's infancy and the difficulty in precisely distinguishing and identifying the end user for each case at hand, three primary end users may be established. In this light, we consider the *intended main end user* and the *learning context in which the learning process is embedded.*

We opted to separate scenarios by the *intended main end user*, who will have the most insight into the information that educational process mining will provide. A *learner* would, in most circumstances, be the one to (1) generate the data of the learning process to be analyzed, (2) receive individualized reports and insights, and (3) use the data analysis output for adjusting their learning strategy. Having a *student* as the intended primary end user implies that the learner may utilize the insights to enhance their learning or course-taking process by receiving insights on their study input quality, learning performance, or study patterns over time (e.g., Cameranesi et al. 2017). The *instructional designer (educator)* characteristic refers to an end user who primarily benefits from the insights of analyzed learner data. Instructional designers may use such reports to enhance the instructional design or to intervene in real time to provide students with guided feedback and support them in adapting their learning strategy (e.g., Lira et al. 2019). The primary beneficiary of the *organization (learning institution) characteristic* is the provider of the educational environment. Learning institutions include universities, MOOC providers, vocational training schools, and continuous training platforms, among others. On the university level, this category of main end users refers to course coordinators, considering learning behavior across courses, years, and individual learners. On an open educational level, it refers to platform providers of MOOC courses. Process mining can be used to analyze MOOCs to reduce dropout rates, for example (e.g., Rizvi et al. 2018). Analyzing learner data in terms of learner course completion, fulfillment, and dropout behavior has the potential to support such users with valuable insights for their course-taking behavior. For this type of main user, analyzing data on an aggregated level (as opposed to the individual learner or course level) is most suitable and insightful.

Finally, the *learning context* explains the learning environment in which digital learning processes will be deployed. We distinguished between *kindergarten–high school* (e.g., Gomez et al. 2021), *higher education* (Engelmann and Bannert 2019), and *continuous education* (Ariouat et al. 2016), which includes workforce training and programs for personal improvement and *vocational training*. The learning context can have important implications regarding the type of data available, as well as the legal and ethical treatment of data (i.e., using minors' data).

### 4.3 Dimension 3: Data for Analyzing Digital Learning Processes

The third set of dimensions refers to analysis options of digital learning processes and is concerned with the ***data*** dimension (i.e., *data input* and *data collection interface).* This input dimension refers to the origin and characteristics of the to-be-analyzed data.

The *data input characteristics* refer to data prerequisites that need to be created as the basis for further educational process mining-based analysis. Although most applications that analyze digital learning processes use pre-existing event data, such as automatically collecting data through a system, we discovered other data, such as *audio* (Nguyen et al. 2021), *video* (Lira et al. 2019) or *text* (Mittal and Sureka 2014), from which event data can be manually or automatically derived or supplemented. *Image* data might potentially be used, although no studies presently use it.

The *data-collecting interface* specifies the tools used to collect the event data. Most applications that aim to analyze digital learning processes collect event data by using an *internal system* (Juhaňák et al. 2019) or a *MOOC platform* (Rizvi et al. 2018), *other web-enabled learning tools* (i.e., accessible and distributed through a web browser from any kind of device), *non-web-enabled tools* (software accessed as an application on specific devices, e.g., Doleck et al. 2016) and *automatic* and *manual data coding*, for example, through the coding of video data, describing the remaining cases.

### 4.4 Dimension 4: Techniques for Analyzing Digital Learning Processes

The fourth set of dimensions refers to analysis options of digital learning processes and is concerned with ***analysis techniques*** and the analysis output format. This dimension differentiates between *process mining type, analysis beyond process mining,* and *output presentation*. It covers the methods used to examine digital learning processes.

Although various techniques are possible in this area, we opted to focus on distinguishing between the fundamental *process mining types* employed and the type of analysis that goes beyond the standard process mining functions. Only applications that expressly indicated the usage of discovery and conformance were found in terms of process mining categories. *Discovery* employs process mining to build a process model by analyzing event log data. An example of this would be creating a process model of a learning process using an event log from a system (e.g., Rogiers et al. 2020). We discovered *conformance* in scenarios in which process mining is used to compare the mined model based on the event log to an existing process model. One example is checking adherence to course order suggestions (e.g., Cameranesi et al. 2017). Process *enhancement* was not expressly employed in any of the evaluated articles, yet its applications are conceivable, such as extending a learner's learning process model using information extracted from specific event logs.

*Analysis beyond process mining* refers to the many sorts of analyses that are done to either precede or extend typical process mining analysis. This excludes supplemental analyses, which are, in essence, independent of the process mining application. *Clustering* may be used with process mining to separate learners based on characteristics such as grades. Approaches based on *rules* can be utilized similarly. Furthermore, the classification may be used to forecast learner success based on previously mined processes. Other approaches resemble the combination of unsupervised and supervised learning or time-series analysis. *None* entails applications that did not employ analysis beyond process mining techniques.

Finally, the *output presentation* dimension specifies how the process mining findings are displayed to the intended main end user. The fundamental issue here is that nonexpert users require a greater level of abstraction of information to derive useful insights from the data. Beyond the identified model using process mining, the *raw model* implies no other form of presentation. A *graphical presentation* defines how findings are shown, for example, by graphs. In *numerical* presentation, results are displayed in the form of numbers or tables, such as the fitness scores of identified models. Finally, *textual* presentation outlines instances in which the system converts the knowledge obtained from process mining into readable information, such as suggestions or automatically created reports.

## 5 Patterns of Learning Process Design and Analysis with Process Mining

In this section, we present three design patterns of learning process design and analysis with process mining. These are the results from our cluster analysis and eleven expert interviews by applying our taxonomy. The design patterns are illustrated according to the format of Dickhaut et al. (2022) and Weinert et al. (2021) in Fig. 6, 7, and 8. We believe that the representation of the design patterns in a clear illustration enhances comprehensibility for interdisciplinary readers and adds to the potential practical impact of our research. To avoid redundancies we only briefly explain the patterns in the text with reference to exemplary quotes from our collected interview data. The patterns are organized in the categories "name", "goal", "pattern based on taxonomy", "challenges", "influential factors & requirements", "exemplary success stories", and "design and evaluation steps" and we believe they are self-explanatory to read from left to right and up to down.

### 5.1 Design Pattern 1: Process Mining to Discover Course-Taking Behavior at an Institutional Level

Discovering student course-taking behaviors in higher education or at MOOC platforms is a common challenge for educational institutions (e.g., Nguyen et al. 2021, 2020a; Cameranesi et al. 2017). A rich analysis of existing course-taking data offers opportunities to identify unknown course combination patterns and provide "next course choice recommendations" to students (Cameranesi et al. 2017). However, due to the growing complexity of multiple available courses, large numbers of students, and countless course combinations, traditional data analysis is not straightforward (Rizvi et al. 2018; Cameranesi et al. 2017). The use of educational process mining to discover and visualize course combinations based on event log data from taken, finished, or dropout courses provides a solution suggestion for this common problem. In our interviewee subset, two interviewees were experts who dealt with course-taking behavior at an institutional level (E10, E11 see Table 3). Both interviewees mentioned that use cases of analyzing course-taking behavior with educational process mining provide multiple opportunities, such as revealing blueprints of commonly taken course combinations to design new profile areas or study programs for students, uncovering course-taking paths that often lead to successful graduation, identifying course combinations that lead to dropouts, or using the discovered information to provide students with "next course recommendations." For example, E10 mentioned, "*we try to collect as much data on the [course] bidding and registration process of students to collect, visualize, and understand the different course-taking behavior of our students.*"

Also, in our sample of studies, one natural grouping of studies focused on analyzing course-taking behavior (15 studies, 22% in our cluster analysis). Rizvi et al. (2018), for example, utilized process mining to reduce dropout rates, and Cameranesi et al. (2017) analyzed course-taking behavior to find bottlenecks in a degree program. The pattern to solve this recurring problem for educators and educational designers is to extract event logs about students' planned, taken, and finished courses from learning management systems. Past researchers have used available process mining tools (majority of studies in our sample use either Disco, ProM, or pMinerR) to apply common discovery algorithms such as alpha miner or heuristic miner to unveil the course-taking behavior of students (Rizvi et al. 2018; Cameranesi et al. 2017). Afterward, the visualization of the individual course-taking paths of students as a graph can be used to monitor the actual behavior. As revealed in our expert interviews (e.g., E5, E6, and E7) and the empirical literature (e.g., Uzir et al. 2020), pattern one involves the application of clustering methods (e.g., hierarchical clustering) to identify novel combinations or blueprints of student course behaviors. Moreover, time series analyses are commonly applied to provide "next course recommendations" to students. For example, E5 mentioned: "*We mostly use clustering or time series analysis to track student behavior.*" Or E6 mentioned, "*we use t-tests between groups and potentially also machine learning algorithms to provide students recommendations after.*" Also in the literature, the study of Cameranesi et al. (2017) provides a successful use case application of our pattern. They monitored student course-taking behaviors with process mining discovery and conformance checking based on event logs to investigate if and how students take courses and what might lead to best practices (successful career) or worst-case practices (dropout). Figure 6 summarizes this first design pattern on discovering course-taking behavior at the institutional level.

### 5.2 Design Pattern 2: Process Mining to Provide Formative Feedback for Learning Activities

One of the larger challenges in TML lies in the missing insights on student learning activities due to large-scale or distance learning scenarios and the limited resources of educators to analyze individual learning processes at scale and provide students with needed formative feedback (e.g., Eom et al. 2006; Brinton et al. 2015; Hone and El Said 2016; Huang et al. 2021). Also, almost all of our expert interviewees mentioned that monitoring individual student learning activities to provide formative feedback in the learning processes is still a challenge in most digital learning scenarios (Hattie and Timperley 2007).

Educational process mining can solve this challenge by monitoring and analyzing learning processes. In our cluster analysis, a natural cluster of 20 studies (31%) used process mining to monitor student learning outcomes. Hence, the second design pattern treats the leveraging of interaction data to monitor and enhance the learning activities of students (for an overview, see Fig. 7). This was also a recurrent challenge and application in our interviews with experts as this exemplary quote of E5 depicts: "*When we monitor students' performance and behavior our goal is usually to support them with feedback, e.g., to help students meet deadlines or when students turn in assignments which do not make sense.*" Nevertheless, E5 also mentions the limitation of feedback in the same vein: "*Psychological problems of students are sometimes hard to support. There is not much you can do. I wish I could help students who struggle emotionally [with process mining and automated feedback].*"

A learning process designed with formative feedback based on analysis can enable educators to provide scaffolds and hints to help students in activities they might struggle with, provide personalized feedback, or embed social comparison nudges to trigger students to finish a certain exercise (Hattie and Timperley 2007; Winkler et al. 2021; Wambsganss et al. 2021). Our pattern suggests embedding a learning process in a digital learning tool (e.g., in a quiz, a conversational agent, or a game). It is important that the learning tool leads the student through a sequence of concrete learning activities toward a measurable learning goal. During the learning interaction, the educator can extract event logs of every user interaction (e.g., buttons clicked, text entered, starts and ends of breaks). The goal is to embed the analysis of student learning behavior with discovery and conformance analysis. For example, the Python library PM4PY can be used in the back end of the web tool to provide instant feedback in the form of hints and recommendations if students struggle to find a solution. Wambsganss et al. (2021) provide a successful use case application for pattern 2. They built a conversational agent as a web tool that tutored students through a learning process consisting of a persuasive writing exercise. The authors tracked the students' interaction data, including the written texts, and provided adaptive formative feedback to argumentative errors. The feedback helped students enhance their argumentative writing skills compared to a control group without adaptive formative feedback.

### 5.3 Design Pattern 3: Enabling Educators to Monitor Learning Progress

Formulating learning goals, designing learning activities that address the goals, and organizing both in a curriculum for a particular learning scenario are the main tasks of educators, according to our interviews. One method of monitoring and improving a learning scenario is to monitor individual learning activities to understand and redesign learning processes. A solution to this recurring and common problem could be analyzing student activity event log data to monitor learning activities, evaluate learning goals, and possibly redesign the learning process (Uzir et al. 2020). Almost half of the papers in our sample formed a natural cluster of studies for educational designers in higher education to monitor student learning processes for factual knowledge to understand the learning tasks. Hence, our third identified solution pattern suggests that educators plan a learning process according to learning objectives, user, and purpose (see Fig. 8). Kratzwohl et al. (2002) provide a taxonomy for efficiently setting the right learning goal and learning task. Most of our expert interviewees mentioned that this is their usual practice when starting to plan a learning scenario. For example, E7 mentioned: "*We collect everything. The more the better. We collect the click stream with time stamps (log data and trace data). When they watch a video when they stop,* etc..*) it's better to have more and filter it. I like doing time series, and pattern analysis, how they perform over time. You can compare them with learning outcomes, motivation, engagement,* etc.*"*

Next, educators need to extract event logs from learning management systems such as Canvas and Moodle and apply common algorithms such as alpha, heuristic, or inductive miner to discover the student learning processes. Through visualizing the learning activities of students as a graph, the educators can then (live) monitor the actual learning processes and compare them with the planned learning process to redesign their curriculum. Uzir et al. (2020) provide a successful use case for applying the third pattern. They used process mining techniques to monitor and cluster log data of 482 undergraduate students over a 13-week course to uncover learning strategies and improve their curriculum. The effort revealed four different learning strategies among the students that were correlated with academic performance. Based on the data analysis, educators established new learning activities for self-regulated learning. However, to be able to utilize the data for learning progress monitoring several interviewees mentioned critical issues about data privacy, ethics, and data quality (E.g., E5, E6, E7). For example, E7 mentioned with regards to learning progress monitoring: "*The problem is that the data is not always clean. E.g., the data was not correctly logged. When the data is not clean or actions are missing. E.g. if there is no log-out event, number two is little data (e.g., less than 100 students) this brings additional challenges for us.*"

## 6 Discussion

Educational process mining offers the opportunity to reap the benefits that data in educational scenarios provide (Gupta and Bostrom 2009; Bogarín et al. 2018; Juhaňák et al. 2019; Cerezo et al. 2020). Its potential to enhance the personalization of learning through adaptive feedback and scaffolding underlines the importance of a solid theoretical and practical framework. Our taxonomy, design patterns, and the research agenda outlined in the next section collectively contribute to this need by expanding the knowledge base of design characteristics crucial for embedding process mining in educational contexts. Based on the findings of this study, we discuss theoretical and practical implications and suggest future research avenues for educational process mining.

### 6.1 Theoretical Contributions

In this study, we conceptualized educational process mining as a new perspective and set of techniques to leverage the potential of learner data in order to improve the individualization of education at scale. This can be done by creating synergies between learning analytics and process mining.

First, our study offers a nuanced understanding of what must be considered when designing and analyzing digital learning processes based on student data in order to discover, monitor, and enhance individual learning based on process mining. We synthesized existing research, including TML, individualization of digital learning, and literature reviews by creating a taxonomy that structured and grouped design characteristics of educational process mining applications (Gupta and Bostrom 2009; Kundisch et al. 2021). Past literature has mostly approached the use of process mining for educational purposes from a technical perspective (e.g., Mouchel et al. 2023, Ludwig et al. 2024, He et al. 2021). Thus, our study contributes to and extends the theoretical foundation of educational process mining by establishing a taxonomy that captures the design characteristics of process mining applications for the specific context of educational scenarios from both a technological and socio-technological viewpoint. Informed by theoretical frameworks in the information systems field (Bostrom and Heinen 1977; Gupta and Bostrom 2009), the taxonomy offers a comprehensive framework to evaluate, design, and compare educational process mining applications in various educational scenarios that become distinguishable through the learning mode or the intended main user, for instance. We uncovered and categorized new dimensions and features that are part of a TML scenario and play a critical role in student learning success beyond the technical perspective of process mining. Specifically, this includes embedding and evaluating algorithmic approaches, as well as incorporating the learner and his or her activities from a pedagogical perspective. These elements are crucial within a pedagogical learning scenario and are instrumental in fostering student learning success. Integrating the TML perspective (Alavi and Leidner 2001), our work offers a holistic view of the role of process mining in education, thereby bridging the gap between technical potential and pedagogical application (Bostrom and Heinen 1977; Gupta and Bostrom 2009).

Second, our study aims to contribute to the understanding of learning analytics and TML in information systems research that could incorporate process mining and improve individual digital learning. Existing literature in the field of TML has largely focused on using digital learning processes for process discovery and conformity approaches (e.g., Rogiers et al. 2020; Cameranesi et al. 2017). Our interview results suggest that current learning scenarios may explain this pattern. Gaps in data traces, for instance, can complicate applying a conformance analysis. Such data traces also provide promising potential for tailor-made individualized course recommendations, well-balanced course allocation for high-quality teaching, or the extraction of new students (Cerezo et al. 2020; Han et al. 2021; Nguyen et al. 2020b).

Third, we used the concept of design patterns to illustrate three use cases of educational process mining. We view patterns as both guidance for data usage and analysis and an illustration of relevant contextual factors to consider (Dickhaut et al. 2023). The three design patterns in this paper suggest data requirements that can enable the realization of respective educational process mining solutions. At the same time, the design patterns also illustrate intended goals and potential challenges to consider in relation to the respective application. We thereby emphasize and challenge the underlying assumptions of effective educational process mining. Understanding and expressing the necessary requirements and potential challenges as well as offering space for solutions is important to build a cumulative body of design knowledge, to show the limits of generalizability, and to prepare the ground for realizing educational process mining applications.

## 6.2 Practical Contributions

From a practical perspective, our findings on educational process mining enable a more targeted and effective implementation, as well as the analysis and effective use of technology in education. Researchers and practitioners can more effectively design, evaluate, compare, and theorize how different technological embeddings of the young field of process mining impact student learning outcomes in a specific learning scenario and task, thanks to this systematic classification of learning scenarios.

Our taxonomy not only categorizes educational process mining applications but also serves as a practical tool for enhancing the design, delivery, and evaluation of education. It enables educators, administrators, and designers to make informed decisions that directly improve student learning experiences and outcomes. More specifically, the taxonomy aids the identification of patterns and anomalies in student learning behaviors. For instance, by analyzing the process data categorized under different educational scenarios, practitioners can pinpoint areas where students repeatedly encounter difficulties. This identification process is crucial for adapting instructional materials and interventions that directly address these learning gaps. In a similar vein, the taxonomy facilitates the ongoing monitoring of student engagement and progression through their educational activities. By providing a framework to compare students' actual learning paths against optimal process models, educators can intervene in real-time to offer support or adjustments to the course trajectory. This real-time monitoring is particularly beneficial in large-scale learning environments like MOOCs, where individual attention is challenging yet critical for student retention and success.

By enabling the detailed analysis of how different educational processes affect learning outcomes, our taxonomy also guides the design of more effective educational interventions. For example, the taxonomy can help institutions experiment with and refine various teaching methodologies, such as flipped classrooms or blended learning, by providing a structured way to assess their impact on student performance and engagement. The taxonomy thereby supports educational administrators and policy-makers make data-driven decisions about curriculum development, resource allocation, and student support services. By understanding the types of process mining applications that are most effective in various educational contexts, decision-makers can allocate resources more efficiently and develop policies that promote optimal learning environments.

The design patterns of this study are relevant to the educational technology area and associated applications from a practical perspective. The patterns, in combination with the taxonomy for the design and analysis of digital learning processes, serve as a personal guide to studying, designing, and evaluating the individualization of digital learning at scale. We argue that design patterns can provide an actionable space for practitioners to imagine potential use cases of educational process mining for their own scenarios. Particular emphasis is hereby placed on making explicit important underlying assumptions, such as associated data challenges that are fundamental to the effectiveness of educational process mining.

In particular, our three design patterns demonstrate the variety of potential use cases educational process mining offers to TML by improving the individualization of digital learning. Different stakeholders, including instructional designers (design pattern 3), educational organizations (design pattern 1), and individual learners (design pattern 2), can benefit from different interventions. Depending on the user, different types of analyses and output presentations must be considered when deploying process mining for learning analytics. Across the three design patterns, educational process mining offers great opportunities to the higher education domain and individual learning mode, with other levels and modes (e.g., collaborative learning) of education to be further explored as part of future research.

### 6.3 Limitations and Future Research

This study and its findings should be interpreted in the light of certain limitations. First, the taxonomy and proposed design patterns and usage implications depend on the literature and data we reviewed. Both qualitative and quantitative empirical data about process mining systems in educational contexts are lacking, and obtaining those data is an overall research need. While much of the present research focuses on the theoretical aspects of process mining in education, few actual field assessments of process mining with deployed systems and users exist (e.g., Mouchel et al. 2023). More specifically, contemporary empirical research looks at process mining systems that may not be tied to real deployment environments (Bogarín et al. 2018; Juhaňák et al. 2019; Cerezo et al. 2020).

In this vein, we are aware that learning data alone may be insufficient to provide a comprehensive overview of learning activities (e.g., Baker and Hawn 2021). We acknowledge that other sources of data, such as tasks and assignments that occur outside of the TML through Moodle, for instance, should be considered to paint a more holistic and complete picture of data in a particular learning context. Also, not all learning activities are captured within digital platforms, which poses a limitation to the completeness and representativeness of the data (Dahlstrom et al. 2014). Data quality issues such as incompleteness, inaccuracy, or inconsistency can severely undermine the insights derived from process mining techniques (Baker and Hawn 2021). In educational contexts, the variability in data entry, the reliance on digital platforms for capturing student interactions, and the absence of standardized data collection protocols across different learning management systems further exacerbate these challenges. Consequently, the effectiveness of process mining in unveiling meaningful patterns and supporting pedagogical decisions may be compromised, necessitating robust data preprocessing and validation methods to ensure reliability and validity of the findings.

In a similar fashion, the comparison with other data sources, such as enrollment or systems data, highlights additional future research avenues (e.g., Dahlstrom et al. 2014). Incorporating such external, contextual data – often referred to as digital traces – can significantly enrich the process mining analysis by providing additional context to student behaviors. The integration of these external data sources requires careful consideration of prerequisites, such as data accessibility and privacy concerns, as well as a thorough understanding of the benefits they can bring to enhance the comprehensiveness of the analysis. Also, our study's applicability is limited by regional differences and variations in digital maturity among educational providers. The digital transformation of education is unevenly distributed, with significant disparities in the adoption of digital technologies and process mining capabilities across regions and institutions. These variations affect not only the availability and quality of data but also the relevance and applicability of process mining solutions. Consequently, our findings may not be universally applicable, necessitating further research that considers these regional differences and seeks to understand how process mining can be adapted to diverse educational contexts and levels of digital maturity.

While process mining offers unique insights into the flow of learning activities, it is but one of several tools available for analyzing educational data. As mentioned earlier, digital course directories and existing learning systems (e.g., for course registrations or student profiles) exist in practice and offer rich data traces in theory. Such systems provide complementary perspectives, capturing different facets of student engagement and academic progress. The reliance solely on process mining might overlook insights that could be derived from these other sources, suggesting a more integrated approach to data analysis that leverages the strengths of each system to provide a holistic understanding of student learning. Hence, with our study, we aim to establish process mining in education as an additional technique that leverages the experience and research on business process management for educational learning processes. Nevertheless, we believe this should be done in combination with other proven learning analytics methods rather than replacing them. While a large body of reviewed literature focuses on generic learning processes, practical training processes and course-taking sequences provide opportunities for additional empirical investigation, especially when coupled with process mining techniques. An interdisciplinary lens that allows for a multiplicity of data analysis tools offers novel research avenues. These include but are not limited to (1) investigating learner knowledge and skill levels, (2) personalizing systematic learning recommendations for students, and (3) considering learning scenarios in continuous and vocational education, such as MOOCs and practical training.

Exploring these research avenues also requires addressing more fundamental questions. While we have considered the potential of educational process mining as a point of departure, it is important to evaluate its appropriateness compared to other analysis methods and tools. These alternatives might be more suitable to address TML scenarios, for example, in terms of cost-efficiency considerations or questions of data availability. While process mining can offer deep insights into student learning processes, the financial and resource implications of adopting such technologies are not trivial. The costs associated with procuring, customizing, and maintaining sophisticated process mining software, alongside the need for training staff to effectively utilize these tools, can be prohibitive, especially for smaller or resource-constrained educational providers. This limitation underscores the importance of future research on conducting comprehensive cost–benefit analyses to ascertain the feasibility and potential return on investment of process mining initiatives in educational settings. It is necessary to mention and highlight ethical challenges that arise when using digital trace data, particularly in the context of education (Hakimi et al. 2021). As part of our reasoning, an important underlying assumption is that students have consented to the use of their data for TML. Informing and obtaining consent for digital trace data is an extensively discussed ethical challenge. According to Johnson (2019), new personal information might be inferred through already-gathered data. The aggregation and combination of data sources risk de-anonymizing individuals. Similarly, in the context of educational process mining, data might be reused and decontextualized to answer certain questions, thereby potentially developing proxies for certain variables (i.e., gender, race, age) even though those are not explicitly collected. Regulatory considerations also pose a significant limitation to the application of process mining in education. The management and analysis of student data are subject to a complex landscape of privacy laws and regulations, which vary significantly across jurisdictions. Compliance with the General Data Protection Regulation (GDPR) in the European Union, for instance, requires meticulous attention to how student data is collected, processed, and stored. Educational institutions must navigate these regulations carefully, ensuring that process mining activities are conducted in a manner that respects students' privacy rights and complies with all applicable legal requirements, adding another layer of complexity to the adoption of these technologies. The question of data ownership and consent must be discussed in consideration of the unintended impact educational process mining may have on student learning and social development. Future studies, as well as organizations deploying educational process mining, are required to define privacy notions, such as the extent to which the use of personal data should be accepted by individual learners and the society at large. When turning towards the commercialization of student data, sharing such data with technology vendors also introduces novel issues of inadequate security controls (i.e., Russel et al. 2018). While extant research has explored ethical issues and related solutions for digital trace data in education, research bodies in this realm are fragmented and lack consideration of the education of younger children and informal learning scenarios outside the traditional classroom (Hakimi et al. 2021). This becomes particularly important considering the variety of online learning formats and the increasing availability of various target groups. We urge future research to look further into the ethical challenges of educational process mining and consider the unique affordances of the educational setting when discussing ethical and societal implications.

Table 4 summarizes avenues for future research along our discussions around (1) general questions about the usefulness and use of educational process mining, (2) the analysis of digital learning based on our taxonomy, (3) the use of design patterns for educational process mining, and (4) ethical challenges that can arise in the context of educational process mining.

**Table 4** Avenues for future research and related questions

| Topic | Avenues for future research and related questions |
|---|---|
| Application of Educational Process Mining | How can we assess the usefulness of educational process mining compared to other tools and methods (e.g., simple queries against existing databases)? |
| | How can we ensure data quality and completeness, e.g. when using data sources outside of TML? |
| Analysis of Digital Learning | How can learning institutions best identify and leverage existing and potential data traces? |
| | How can process mining be used to enable individualized course recommendations, course allocation, and the extraction of new students in real-world applications? |
| | How can unique affordances of educational providers, for example, regarding digital maturity and regional requirements, be considered in educational process mining? |
| Design Patterns | How can we empirically validate the usefulness of a specific design pattern for the methodological approach of educational process mining? |
| | How can we embed educational process mining in contemporary learning scenarios through the application of the three design patterns? |
| | How can we empirically validate the usefulness of a specific design pattern for the practical impact for learning scenarios? |
| Ethical Challenges | How can user anonymity be assured, especially regarding the aggregation and combination of data sources? |
| | How can user data be protected against the commercialization of learning data? |
| | How do we protect younger learners' data and data inferred from learning scenarios outside the traditional classroom? |

## 7 Conclusion

Educational process mining is a crucial perspective for the design and analysis of digital learning processes as educational institutions strive to provide ever-personalized educational settings for their students and improve educational and administrative processes around course-taking and learning-related behaviors. In this paper, we set out the design and analysis characteristics for digital learning processes through the development of a taxonomy and design patterns on educational process mining applications. This research not only progresses the academic discourse surrounding educational process mining but also serves as a practical guide for its application in large-scale, TML environments.

The implications of our findings are twofold. Firstly, the refined taxonomy and the identified design patterns provide a structured approach that can guide educational designers and practitioners in enhancing digital learning scenarios. This framework helps in systematically leveraging process mining to discover, monitor, and enhance the learning processes of students, thereby promoting more effective and personalized educational experiences. Secondly, our research contributes to a more comprehensive understanding of the challenges and opportunities associated with educational process mining, which includes the need for strict data regulation to ensure ethical practices in handling and analyzing educational data. Overall, our study emphasizes the potential of educational process mining to facilitate significant advancements in digital education by enabling a deeper and more actionable understanding of student learning behaviors and educational processes. As educational technologies and methodologies continue to evolve (e.g., through generative Artificial Intelligence), the insights might play a crucial role in shaping the future of education, ensuring that learning environments are both effective and adaptable to the needs of all students.

**Funding** Open access funding provided by Bern University of Applied Sciences.



## References

Alavi M, Leidner DE (2001) Research commentary: technology-mediated learning – a call for greater depth and breadth of research. Inf Syst Res 12(1):1–10

Ariouat H, Cairns HA, Barkaoui K, Akoka J, Khelifa N (2016) A two-step clustering approach for improving educational process model discovery. In: IEEE 25th international conference on enabling technologies, pp 38–43

Baker RS, Hawn A (2021) Algorithmic bias in education. Int J Artif Intell Educ 32:1052–1092

Betts M, Rosemann M (2022) The new learning economy: thriving beyond higher education. Routledge

Bhattacherjee A (2012) Social science research: principles, methods, and practices. Textbooks Collection. 3

Bissantz N, Hagedorn J (2009) Data mining. Bus Inf Syst Eng 1(1):118–122

Bogarín A, Cerezo R, Romero C (2018) A survey on educational process mining. Wiley Interdiscip Rev Data Min Knowl Discov 8(1):1230

Bostrom RP, Heinen JS (1977) MIS problems and failures: a socio-technical perspective. Part I: the causes. MIS Q 1(3):17

Brinton CG, Rill R, Ha S, Chiang M, Smith R, Ju W (2015) Individualization for education at scale: MIIC design and preliminary evaluation. IEEE Trans Learn Technol 8(1):136–148

Cameranesi M, Diamantini C, Genga L, Potena D (2017) Students' careers analysis: a process mining approach. In: Proceedings of the 7th international conference on web intelligence, mining and semantics, Article 26

Cerezo R, Bogarín A, Esteban M, Romero C (2020) Process mining for self-regulated learning assessment in e-learning. J Comput Higher Educ 32(1):74–88

Chatti MA, Dyckhoff AL, Schroeder U, Thüs H (2012) A reference model for learning analytics. Int J Technol Enhanc Learn 4(5–6):318–331

Chen H, Chiang RHL, Storey VC (2012) Business intelligence and analytics: from big data to big impact. MIS Q 36(4):1165–1188

Costa J, Azevedo A, Rodrigues L (2020) Educational process mining based on moodle courses: a review of literature. In: Proceedings CAPSI 2020. https://aisel.aisnet.org/capsi2020/11

Dahlstrom E, Brooks CD, Bichsel J (2014) The current ecosystem of learning management systems in higher education: student, faculty, and IT perspectives. Research report. ECAR, Louisville. http://www.educause.edu/ecar

DeSanctis G, Poole MS (1994) Capturing the complexity in advanced technology use: adaptive structuration theory. Organ Sci 5(2):121–147

Dickhaut E, Li M, Janson A, Leimeister JM (2022) The role of design patterns in the development and legal assessment of lawful technologies. Electron Mark. https://doi.org/10.1007/s12525-022-00597-1

Dickhaut E, Janson A, Soellner M, Leimeister JM (2023) Lawfulness by design – development and evaluation of lawful design patterns to consider legal requirements. Eur J Inf Syst 33(4):441–468. https://doi.org/10.1080/0960085X.2023.2174050

Dillenbourg P, Järvelä S, Fischer F (2009) The evolution of research on computer-supported collaborative learning. In: Balacheff N et al (eds) Technology-enhanced learning: principles and products. Springer, Dordrecht, pp 3–19

Doleck T, Jarrell A, Poitras EG, Chaouachi M, Lajoie SP (2016) Examining diagnosis paths: a process mining approach. In: 2nd International conference on computational intelligence & communication technology, Ghaziabad. https://doi.org/10.1109/CICT.2016.137

Engelmann K, Bannert M (2019) Analyzing temporal data for understanding the learning process induced by metacognitive prompts. Learn Instruct. https://doi.org/10.1016/j.learninstruc.2019.05.002

Eom SB, Wen HJ, Ashill N (2006) The determinants of students' perceived learning outcomes and satisfaction in university online education: an empirical investigation. Decis Sci J Innov Educ 4(2):215–235

Ghazal MA, Ibrahim O, Salama MA (2018) Educational process mining: a systematic literature review. In: European conference on electrical engineering and computer science, EECS 2017, pp 198–203

Gomez MJ, Ruipérez-Valiente JA, MartínezKim PAYJ (2021) Applying learning analytics to detect sequences of actions and common errors in a geometry game. Sensors 21(4):1–16

Gupta S, Bostrom RP (2009) Technology-mediated learning: a comprehensive theoretical model. J Assoc Inf Syst 10(9):686–714

Gupta S, Bostrom R (2013) An investigation of the appropriation of technology-mediated training methods incorporating enactive and collaborative learning. Inf Syst Res 24(2):454–469

Hakimi L, Eynon R, Murphy VA (2021) The ethics of using digital trace data in education: a thematic review of the research landscape. Rev Edu Res 91(5):671–717. https://doi.org/10.3102/00346543211020116

Han J, Kim KH, Rhee W, Cho YH (2021) Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. Comput Educ 163:104041

Hattie J, Timperley H (2007) The power of feedback. Rev Educ Res 77(1):81–112

He Q, Borgonovi F, Paccagnella M (2021) Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. Comput Edu 166:104170

Hone KS, El Said GR (2016) Exploring the factors affecting MOOC retention: a survey study. Comput Educ 98:157–168

Huang N, Zhang J, Burtch G, Li X, Chen P (2021) Combating procrastination on massive online open courses via optimal calls to action. Inf Syst Res 32(2):301–317

Ifenthaler D, Gibson D (2019) Opportunities for analytics in challenge-based learning. In: Tlili A, Chang M (eds) Data analytics approaches in educational games and gamification systems. Springer, Singapore, pp 55–68

Johnson OA, Ba Dhafari T, Kurniati A, Fox F, Rojas E (2019) The clearpath method for care pathway process mining and simulation. In: Daniel F et al (ed) Business process management workshops. BPM 2018. Springer, Cham, pp 239–250

Juhaňák L, Zounek J, Rohlíková L (2019) Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. Comput Hum Behav 92:496–506

Kaufman L, Rousseeuw PJ (2005) Finding groups in data: an introduction to cluster analysis. Wiley, New York

Krathwohl DR (2002) A revision of Bloom's taxonomy: an overview. Theor Pract 41(4):212–218

Kulik JA, Fletcher JD (2016) Effectiveness of intelligent tutoring systems: a meta-analytic review. Rev Educ Res 86(1):42–78

Kundisch D, Muntermann J, Oberländer AM, Rau D, Röglinger M, Schoormann T, Szopinski D (2021) An update for taxonomy designers – methodological guidance from information systems research. Bus Inf Syst Eng 64(4):421–439

Lemay DJ, Baek C, Doleck T (2021) Comparison of learning analytics and educational data mining: a topic modeling approach. Comput Educ Artif Intell 2:100016

Lira R, Salas-Morales J, Leiva L, de la Fuente R, Fuentes R, Delfino A, Nazal CH, Sepúlveda M, Arias M, Herskovic V, Munoz-Gama J (2019) Process-oriented feedback through process mining for surgical procedures in medical training: the ultrasound-guided central venous catheter placement case. Int J Environ Res Publ Health 16(11):1877

Ludwig S, Rausch A, Deutscher V, Seifried J (2024) Predicting problem-solving success in an office simulation applying N-grams and a random forest to behavioral process data. Comput Edu 218:105093

Matook S, Brown SA (2017) Characteristics of IT artifacts: a systems thinking-based framework for delineating and theorizing IT artifacts. Inf Syst J 27(3):309–346

Mittal M, Sureka A (2014) Process mining software repositories from student projects in an undergraduate software engineering course. In: 36th International conference on software engineering, ICSE Companion. ACM, pp 344–353

Mother D, Liberati A, Teztlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. https://doi.org/10.1371/journal.pmed.1000097

Mouchel L, Wambsganss T, Mejia-Domenzain P, Käser T (2023) Understanding revision behavior in adaptive writing support systems for education. Proceedings of the 16th international conference on educational data mining, 445–452. https://doi.org/10.5281/zenodo.8115766

Myers M, Newman M (2007) The qualitative interview in IS research: examining the craft. Inf Organ 17:2–26

Nguyen A, Gardner L, Sheridan D (2020a) Data analytics in higher education: an integrated view. J Inf Syst Educ 31(1):61–71

Nguyen A, Tuunanen T, Gardner L, Sheridan D (2020b) Design principles for learning analytics information systems in higher education. Eur J Inf Syst 30(5):541–568

Nguyen H, Lim KY, Wu LL, Fischer C, Warschauer M (2021) "We're looking good": social exchange and regulation temporality in collaborative design. Learn Instruct 74:101443

Nickerson RC, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22(3):336–359

O'Neill S, Scott M, Conboy K (2011) A Delphi study on collaborative learning in distance education: the faculty perspective. Br J Educ Technol 42(6):939–949

Okoye K (2019) A systematic review of process modelling methods and its application for personalised adaptive learning systems. J Int Technol Inf Manag 27(3):23–46

Rizvi S, Rienties B, Rogaten J (2018) Temporal dynamics of MOOC learning trajectories. In: First international conference on data science, e-learning and information system, Article 39. https://doi.org/10.1145/3279996.328003

Rogiers A, Merchie E, van Keer H (2020) Opening the black box of students' text-learning processes: a process mining perspective. Front Learn Res 8(3):40–62

Romero C, Ventura S (2020) Educational data mining and learning analytics: an updated survey. Wylie Interdiscip Rev. https://doi.org/10.1002/widm.1355

Roth H (1970) Pädagogische psychologie des Lehrens und Lernens. Schroedel, Hannover

Russell NC, Reidenberg JR, Martin E, Norton T (2018) Transparency and the marketplace for student data. SSRN Electron J. https://doi.org/10.2139/ssrn.3191436

Saint J, Whitelock-Wainwright A, Gasevic D, Pardo A (2020) Trace-SRL: a framework for analysis of microlevel processes of self-regulated learning from trace data. IEEE Trans Learn Technol 13(4):861–877

Schoonderwoerd TAJ, van Zoelen E, van den Bosch K, Neerincx MA (2022) Design patterns for human-AI co-learning: a wizard-of-Oz evaluation in an urban-search-and-rescue task. Int J Hum Comput Stud. https://doi.org/10.1016/j.ijhcs.2022.102831

Schoor C, Bannert M (2012) Exploring regulatory processes during a computer-supported collaborative learning task using process mining. Comput Hum Behav 28(4):1321–1331

Seaman JE, Allen IE, Seaman J (2018) Higher education reports. Babson Survey Research Group

Sidorova A, Evangelopoulos N, Valacich JS, Ramakrishnan T (2008) Uncovering the intellectual core of the information systems discipline. MIS Q 32(3):467–482

Singh N, Varshney U (2020) IT-based reminders for medication adherence: systematic review, taxonomy, framework and research directions. Eur J Inf Syst 29(1):84–108. https://doi.org/10.1080/0960085X.2019.1701956

Söllner M, Bitzer P, Janson A, Leimeister JM (2018) Process is king: evaluating the performance of technology-mediated learning in vocational software training. J Inf Technol 33(3):233–253

Syngene Research LLP (2019) Global e-learning market analysis 2019

Szopinski D, Schoormann T, Kundisch D (2019) Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: 27th European conference on information systems (ECIS), Stockholm, Uppsala

Uzir NAA, Gaševic D, Jovanovic J, Matcha W, Lim LA, Fudge A (2020) Analytics of time management and learning strategies for effective online learning in blended environments. In: 10th international conference on learning analytics and knowledge, pp 392–401. https://doi.org/10.1145/3375462.3375493

van der Aalst W (2012) Process mining. Commun ACM 55(8):76–83

vom Brocke J, Simons A, Riemer K, Niehaves B, Plattfaut R, Cleven A (2015) Standing on the shoulders of giants: challenges and recommendations of literature search in information systems research. Commun Assoc Inf Syst 37(1):205–224

vom Brocke J, Jans M, Mendling J, Reijers HA (2021) A five-level framework for research on process mining. Bus Inf Syst Eng 63(5):483–490

Vygotsky LS (1980) Mind in society: the development of higher psychological processes. Harvard University Press

Wambsganss T, Kueng T, Soellner M, Leimeister JM (2021) ArgueTutor: an adaptive dialog-based learning system for argumentation skills. In: 2021 CHI conference on human factors in computing systems, Article 683. https://doi.org/10.1145/3411764.344578

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58(301):236

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26(2):13–23

Weinert T, Benner D, Dickhaut E, Janson A, Schoebel S, Leimeister JM (2021) Supporting digital education processes through interactive gamified learning videos – how innovative learning videos can increase motivation and learning success. HMD 58:1483–1503. https://doi.org/10.1365/s40702-021-00798-w

Winkler R, Söllner M, Leimeister JM (2021) Enhancing problem-solving skills with smart personal assistant technology. Comput Educ. https://doi.org/10.1016/j.compedu.2021.104148