Please quote as: Wambsganss, T., Schmitt, A., Mahning, T., Ott, A., Soellner, S., Ngo, N. A., Geyer-Klingeberg, J., Nakladal, J. & Leimeister, J. M. (2021). The Potential of Technology-Mediated Learning Processes: A Taxonomy and Research Agenda for Educational Process Mining. 2021. - International Conference on Information Systems (ICIS). - Austin, Texas, USA.

The Potential of Technology-Mediated Learning Processes: A Taxonomy and Research Agenda for Educational Process Mining

Completed Research Paper

Thiemo Wambsganss
University of St.Gallen
St.Gallen, Switzerland
Carnegie Mellon University
Pittsburgh, USA
thiemo.wambsganss@unisg.ch

Anuschka Schmitt
University of St.Gallen
St.Gallen, Switzerland
anuschka.schmitt@unisg.ch

Thomas Mahnig
University of St.Gallen
St.Gallen, Switzerland
thomas.mahnig@student.unisg.ch

Anja Ott
University of Kassel
Kassel, Germany
anja.ott@uni-kassel.de

Sigita Soellner
Sapoway
Kassel, Germany
sigitai.andrulytei@gmail.com

Ngoc Anh Ngo
Celonis SE
Munich, Germany
n.ngo@celonis.com

Jerome Geyer-Klingeberg
Celonis SE
Munich, Germany
j.geyerklingeberg@celonis.com

Janina Nakladal
Celonis SE
Munich, Germany
j.nakladal@celonis.com

Jan Marco Leimeister
University of St.Gallen
St.Gallen, Switzerland
University of Kassel
Kassel, Germany
janmarco.leimeister@unisg.ch

Abstract

Educational process mining (EPM) offers new possibilities to discover, monitor, improve, or predict students' learning processes using data about their learning activities captured in technology-mediated information systems (IS). Although EPM has recently attracted considerable research interest, there is still limited shared knowledge about the distinctive design characteristics of EPM from an integrative perspective. To address this *gap, we conducted a systematic literature review to identify EPM characteristics. Building on a technology-mediated learning perspective, we develop a taxonomy that classifies EPM characteristics into four major categories (i.e., purpose, user, input, analysis). We evaluate and refine our taxonomy with ten domain experts, identified three clusters in the reviewed literature, and derived six archetypes of EPM scenarios based on our categorization. Finally, we formulate a novel research agenda to guide researchers in systematizing and synthesizing research on different technological embeddings of EPM in a students' learning process.*

Keywords*:* Educational Process Mining, Technology-Mediated Learning, Socio-Technical Systems, Taxonomy

Introduction

Individual support and personal recommendations for students in their learning processes are still a pendent challenge and have not yet been not solved in many pedagogical scenarios today (e.g., Kulik and Fletcher 2016). According to predominant constructivist learning theories, students need individual tutoring to learn effectively (Vygotsky 1980). However, educational institutions, such as high schools or universities, struggle to offer this kind of individual support due to financial and organizational constraints. Large classroom sizes at high schools, mass lectures at universities with more than 100 students per lecturer, and the recent rise in massive open online courses (MOOCs) with more than 1000 participants impede the individual interaction between students and educators (Seaman et al. 2018). The current Covid19-pandemic and the resulting governmental lockdowns have amplified this effect. Consequently, technology-mediated, distance-learning scenarios have become a reality for many students and educators. Several studies have revealed that this lack of individualized support leads to low learning outcomes, high dropout rates, and dissatisfaction with the overall learning experience (Brinton et al. 2015; Eom et al. 2006; Hone and El Said 2016). Against this backdrop, the steady growth of *information systems (IS)* in education, such as the learning management systems (LMS) Canvas or Moodle, might represent a solution avenue for providing more individual learning support and a better analysis of the students' learning process (Gupta and Bostrom 2009). According to Dahlstrom et al. (2014), 99% of U.S. colleges and universities organize their pedagogical scenarios in standard LMS. Not only the organization of courses in LMS but also the embedding of exercises in intelligent tutoring systems (ITS, i.e., Kulik and Fletcher 2016) or computer-supported collaborative learning tools (CSCL, i.e., Dillenbourg et al. 2009) are expected to continuously grow during the upcoming years to a market size of $336.98 billion by 2026 (Bogarín et al. 2018; Romero and Ventura 2020; Syngene Research LLP 2019). This integration of learning activities in IS offers the opportunity to capture data about both learning activities and learning processes at different granularity levels. A learning process describes the different activities that a student is (or is not) attempting to reach a certain learning outcome (i.e., Roth, 1970). IS in education offer the potential to capture rich traces from these processes, such as event logs (e.g., starting or ending event of a particular exercise at a certain time), textual data (e.g., written essays), or comments on learning content, or measured learning outcomes (e.g., a student's skill level after taking a quiz). This data can be leveraged by learning analytic (LA) algorithms to discover, monitor, improve, or predict a student's learning processes, e.g., through simple scaffolded interventions (e.g., providing instructions based on current skill levels, i.e., Winkler et al., (2020)), or course recommendations (i.e., decision support on course choices). In fact, the mining of processes based on digital traces has gained momentum in several domains, e.g., business processes (van der Aalst et al. 2007), healthcare (Johnson et al. 2019; Rojas et al. 2016), and education (Cerezo et al. 2020). According to Bogarín et al. (2018), educational process mining (EPM) combines the benefits of educational data mining (EDM) and process modelling and analysis. Process mining (PM) itself is a sub-discipline of data mining, adding a process-oriented viewpoint to the analysis of data (Van Der Aalst 2012). While LA primarily focuses on data dependencies and pattern predictions of single activities (Romero and Ventura 2020), EPM adds a learner-centered process viewpoint by looking at event logs beyond a single activity (Bogarín et al. 2018; Cerezo et al. 2020; Juhaňák et al. 2019). This brings additional benefits for data analysis by crossing the boundaries of single events or tasks and enriches the analysis from a learner's process perspective, which is also deeply rooted in didactical design research (i.e., Roth, 1970).

In recent years, a multitude of research has emerged in different disciplines, most prominently in IS, *human—computer interaction* (HCI), and educational technology, that has motivated and investigated the effect of different scenarios and configurations of EPM (e.g., Johnson et al., 2019; Cerezo et al., 2020). This includes, among others, the analysis of the effects of metacognitive prompts for self-regulated learning (Engelmann and Bannert 2019), the examination of the adherence of students to a recommended course path (Cameranesi et al. 2017), and the investigation of process-based feedback during medical training (Lira et al. 2019). The growing number of interdisciplinary studies on EPM highlights the necessity to better understand the design, capabilities, and potential implications of EPM from a holistic perspective (Bogarín et al. 2018; Costa et al. 2020). However, several authors, e.g., Ghazal et al. (2018) and Rogiers et al. (2020), have stated that there is a lack of shared knowledge about the embedding of PM in different learning scenarios and insights for the different personas (e.g., educator, institution, or learner), the technology (e.g., the applied discovery algorithm), or the pedagogical context (e.g., in which educational domain or for which learning task PM is analyzed). Most studies that investigate the potential of EPM focus on a specific task in a certain pedagogical domain and its effect on selected outcomes or user perceptions (e.g., Okoye, 2019; Cerezo et al., 2020). This eventually leads to a fragmented literature base and sometimes contradictory research results (Bogarín et al. 2018; Ghazal et al. 2018). However, for a nascent research field, especially, such as EPM, an integrative viewpoint would be of utmost importance to systematically design, analyze, and compare the different configurations of PM from different aspects (e.g., from a technological, learning context or task-based perspective) to form an impactful research stream (Nickerson et al. 2013). A consistent knowledge aggregation on the different characteristics and dimensions of EPM from a holistic perspective will help researchers and practitioners to systematically design, compare, and evaluate new or existing PM applications. Although initial literature reviews on EPM have emerged over the past years (e.g., Bogarín et al., 2018; Ghazal, Ibrahim and Salama, 2018; Okoye, 2019; Costa et al., 2020), the research is dispersed in a multitude of socio-technical perspectives, resulting in an acute shortage of an integrative perspective. For instance, in a systematic literature review, Ghazal et al. (2018) indicate the need for a systematic classification of EPM but do not derive distinct characteristics and dimensions of EPM across domains. In this regard, IS research can offer a promising viewpoint for classifying a certain IS from a technology-mediated learning perspective (Gupta and Bostrom 2009) into relevant elements (e.g., learning context, user characteristics, and technical standpoint), which ultimately yield different configurations of the technological embedding and the outcomes for different stakeholders (Bostrom and Heinen 1977). Consequently, a systematic classification of empirical studies on EPM scenarios taking this perspective would enable researchers to more effectively design, evaluate, compare, and theorize how different technological embeddings of the young field of PM impact the students' learning outcomes in a specific pedagogical scenario and task. Hence, this paper focuses on the following research questions (RQ):

***RQ1**: What are the dimensions and characteristics of educational process mining (EPM) from an integrative technology-mediated learning perspective?*

***RQ2:** What patterns of usage of these characteristics emerge across empirical studies that help us to identify clusters and archetypes of EPM applications?*

To answer these questions, we develop a taxonomy of PM for education. We follow the rigorous taxonomy development framework as outlined by Nickerson et al. (2013). Based on five iterations, we classify and organize dimensions and characteristics embedded in 66 research publications. According to recommendations provided by ten experts from research and practice familiar with PM in different domains, we evaluate and revise our taxonomy regarding structure and content. Moreover, we perform a cluster analysis and derive three distinct clusters and six archetypes of EPM applications. Finally, we discuss the results and derive implications for a future research agenda of EPM to guide researchers not only in IS, but also in HCI and educational technology.

Theoretical Background

Technology-mediated Learning and Learning Processes

Following constructivist learning theories, students need individual learning environments and tutoring to learn effectively (Vygotsky 1980). Meaningful feedback by an instructor not only helps students to reach a certain learning outcome but also avoids unfortunate attrition (Hattie and Timperley 2007; Kellogg et al. 2014). Therefore, several scholars promote a *"process perspective of learning artefacts"* to better understand how students learn and identify opportunities for learning support interventions (Romero and Ventura 2020; Schumacher and Ifenthaler 2021). In fact, considering learning as a process of activities has a long history in didactical research (e.g., by Martens 2003; Roth 1970). Roth 1970 distinguished every learning process into six core activities: *motivation, difficulty, solution, practice, rehearsal, and transfer*. Also, in IS research, the learning process embedded around technology-mediated learning (TML) forms a prominent literature stream (Gupta and Bostrom 2009; Janson et al. 2020; Winkler and Söllner 2018). 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 – by definition - "*all the elements of a social-technical system: technology and learning techniques, process, actors, actions, and outcomes*" (Gupta and Bostrom 2009, p.3). Based on these and other learning process perspectives, literature has designed teaching methods, such as adaptive feedback via prompts (Hattie and Timperley 2007) or tutoring concepts via pedagogical conversational agents (Wambsganss, Küng, et al. 2021). A famous concept to guide and facilitate the learning processes is the concept of scaffolding (Janson et al. 2020; Wood et al. 1976). Wood, Bruner, and Ross (1976) describe *scaffolding* as temporary instructional support for learners to overcome challenges within their zone of proximal development, adjusting the learners' individual learning processes and experiences. With its origin in social constructivist theory, scaffolding posits that the intersubjectivity between the instructional designer and the individual learner, as well as between the learners, is vital for learning (Vygotsky 1980; Wood et al. 1976). Four types and forms of scaffoldings to guide and facilitate the learning process in TML environments are defined in the literature: *procedural*, *metacognitive*, *conceptual*, and *strategic* scaffolds (Azevedo and Hadwin 2005; Sharma and Hannafin 2007). Scaffolds support learning processes by giving advice on how to use relevant methods and structures (Gupta and Bostrom 2009; Winkler et al. 2020), how to structure tasks, and how to self-monitor learning processes by providing cues or hints to complete a task (Janson et al. 2020). Therefore, Ifenthaler and Gibson (2019) urge for the implementation and further analysis of personal and adaptive learning environments on concepts such as scaffolding to improve individual, personalized learning as well as personalized and adaptive feedback whenever needed. EPM offers the potential to enhance the concept of scaffolding or individual feedback in TML environments (Bogarín et al. 2018). EPM might not only provide a new conceptual framework and technology for providing better learning support through scaffolding but may also enhance self-regulated and self-directed learning through new forms of self-monitoring and self-evaluation (e.g., through conformance checking or process enhancement). Therefore, the mining of the learning process can further help to build learning environments and provide a scaffold when needed.

Educational Process Mining

The most common concepts to process data from technology-mediated learning systems are EDM and LA (Romero and Ventura 2020). Whereas EDM is driven by data (automated adaption of computer systems), the focus is usually not on human interactions. LA aims to analyze educational data but is usually more focused on a learner-centered perspective, e.g., on how to employ analytics to inform or empower instructors or learners (Romero and Ventura 2020). Scholars of LA have called for new application tools and techniques to gain a deeper understanding of the student's learning process (e.g., Siemens 2013). Here, PM for education has evolved by leveraging techniques from EDM with the more learner-centered perspective of LA. Nevertheless, PM is still a relatively young research field, which has been mainly evolving from business process modelling (Van Der Aalst 2012). Due to rapidly increasing data recordings in organizational IS, such as ERP or CRM systems, PM has made it possible to discover, analyze, and improve business processes based on widely available event data. Events are activities or processes carried out by people, machines, and software, leaving trails in so-called event logs. Event logs are data points consisting of a time stamp, event IDs and certain meta data, e.g., the name of the event. They can be leveraged to manage, support, and improve business processes in organizations (Van Der Aalst 2012). The potential of PM has not only been called as one of the top strategic technology trends by Gartner (Panetta 2019) but has also been rapidly adopted by the industry, e.g., by PM start-ups such as Celonis[1], which have been quickly

[1] https://www.celonis.com

evolving to billion-dollar companies. Nevertheless, PM for education is still a rather nascent field. Since around 2009, researchers have aimed to apply PM to raw educational data (event logs of students' learning processes), which is left as a side product of the common educational IS, ranging from traditional LMS to the fields of intelligent tutoring systems (ITS, Kulik and Fletcher, 2016), smart personal assistants (Winkler and Söllner 2018), or computer-supported collaborative learning tools (CSCL) (Dillenbourg et al. 2009). User data is automatically generated throughout the learner's interaction with the learning environment, e.g., clicks, chat protocols, review comments, or even specific learner content, such as written student texts (Wambsganss, Küng, et al. 2021). With the further growth of e-Learning scenarios and TML, an enormous amount of student-centered data bears an attractive potential to not only enrich the pedagogical embedding itself (e.g., Wambsganss et al. 2021) through personalization, recommendations, or formative feedback but also the educators and institutions in monitoring and evaluating learning processes on different levels to extract information from the models and act upon the findings. We refer to PM in education as educational process mining (EPM). EPM is regarded as a subset of EDM and LA (Bogarín et al. 2018). Classical data mining techniques, such as classification, clustering, regression, or sequence mining, are of little use for the control-flow discovery since they are usually not process-centric (Van der Aalst and Weijters 2004; Bogarín et al. 2018). To allow for longer sequences of data based on several time events, a process perspective with PM has been proposed (Van der Aalst and Weijters 2004). Instead of analyzing single data points in the learner's journey, the process-oriented perspective of EPM helps uncover the full end-to-end learning process and the dependencies among the activities in the process, i.e., the learner's success or other measures of success. For EPM, three different basic types of PM can be distinguished (Van Der Aalst 2012; Bogarín et al. 2018):

- **Process discovery:** modelling and visualizing the learning process of students, e.g., in order to monitor a student's individual learning journey or the curricular path a student takes.
- **Conformance checking:** analyzing whether an observed learning process conforms with a pre-defined learning process model, e.g., in order to identify weaker students (outliers) or analyze conformance with guidelines and prescriptions.
- **Process enhancement:** extending a given learning process model based on information extracted from a specific event log related to the same process, e.g., to detect bottlenecks or provide students adaptive feedback on their process.

Classifications of Dimensions and Characteristics of Educational Process Mining

Several authors have emphasized that there is a lack of shared knowledge about the embedding of PM in different learning scenarios and insights on the involved users (e.g., educator, institution or learner), the technological context (e.g., the used discovery algorithm), or the pedagogical structure (e.g., in which educational domain or for which pedagogical task PM is analyzed, Ghazal et al. (2018) or Rogiers et al. (2020)). Therefore, scholars from different diciplines have conducted literature reviews to provide an overview or a conceptualization of EPM (e.g., Bogarín et al., 2018; Ghazal, Ibrahim and Salama, 2018; Okoye, 2019; Costa et al., 2020). However, no holistic classification of characteristics and dimensions of EPM from an interdisciplinary perspective exists (i.e., incorporating a socio-technical viewpoint), which could help researchers and practitioners design, compare, and evaluate (their) use-case for EPM. Classifications are useful to researchers and practitioners, as they enable the structuring of novel and complex domains, which is especially valuable for young and emerging research fields such as EPM. Classifications highlight the interrelationships between the different elements of a phenomenon transparently and coherently and indicate their respective theoretical bases (Bailey 1994; Schöbel et al. 2020). In fact, current reviews fall rather short on a comprehensive and robust structuration of EPM applications. Research is dispersed in a multitude of socio- and technical perspectives, resulting in a pressing shortage of an integrative perspective. For instance, in their systematic literature review, Ghazal et al., (2018) did not derive distinct characteristics and dimensions of EPM at all and focused primarily on technical aspects. The review of Bogarín et al. (2018) provides a structured overview; however, they did not follow a transparent methodologcial approach, e.g., a systematic literature review (vom Brocke et al. 2015; Webster and Watson 2002). Costa et al. (2020) only regarded literature on PM in the LMS Moodle, making it not applicable for other pedagogical scenarios. In this regard, we aim to follow a TML perspective based on the socio-technical system viewpoint since it allows the classification of a certain IS into relevant elements (user, task, structure, and technical standpoint), which ultimately yields different configurations and outcomes (Bostrom and Heinen 1977; Gupta and Bostrom 2009). To systematically classify objects of interest, we can refer to a taxonomy (Nickerson et al. 2013). Consequently, a systematic classification of EPM scenarios taking this perspective would enable researchers to more effectively design, evaluate, compare, and theorize how different technological embeddings of the young field of PM impact the students' learning outcomes in a specific pedagogical scenario and task. Therefore, we aim to address this literature gap by investigating a novel taxonomy that supports decision making in building, designing, and comparing EPM scenarios and applications and helps to specify the relationships of the EPM charateristics towards the outcome of a pedagogical scenario.

Research Methodology

To systematically classify the objects of interest of EPM, we followed a rigorous taxonomy development process, resulting in four distinct steps (Table 1):

	Step 1: Database Creation	**Step 2: Taxonomy Development**	**Step 3: Taxonomy Evaluation**	**Step 4: Taxonomy Application**
Steps	• Search for relevant papers in IS and educational technology • Analyze and synthesize literature about TML and EPM	• Define meta characteristic • Run taxonomy development iterations	• Evaluate dimensions and characteristics with experts based on quality criteria	• Cluster analysis to identify patterns of EPM characteristics and archetypes in literature
Method	Literature Review (vom Brocke et al. 2015), Qualitative Coding	Taxonomy Development (Nickerson et al. 2013)	Expert Evaluation (Szopinski et al. 2019)	Cluster analysis (Kaufman and Rousseeuw 2005)
Source	EPM literature	Existing literature reviews on EPM database	Semi-structured interviews with experts	EPM literature (identified in phase 1)
Results	Database with 66 articles	Taxonomy of design characteristics for EPM	Evaluated taxonomy	Three clusters and six archetypes

Table 1. Overview of our four consecutive research steps

Step 1: Database Creation Through a Systematic Literature Review

To identify relevant literature as the basis for the systematic development of a taxonomy, we conducted a systematic literature review according to Webster & Watson (2002) and vom Brocke et al. (2015). The overall scope of the conducted review can be defined along the dimensions of *process*, *source*, *coverage*, and *techniques* (vom Brocke et al. 2015). To establish the basis for the taxonomy development and conceptualization, we used a *comprehensive set of techniques* (i.e., keyword search, backward search, and forward search). To reach a high level of reproducibility and transparency of our research, we describe the detailed methodical steps that we followed. Based on recent literature reviews on EPM (e.g., Bogarín et al., 2018; Costa et al., 2020), we identified different keywords, which researchers used to describe PM in the educational domain. Based on the keywords, we built the following search strings*: "[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 identified three broad areas for deriving studies on EPM – *IS, HCI* and e*ducational technology* – as they cover a substantial share of the literature on our phenomenon of interest. To find relevant literature of studies that applied PM in educational scenarios, we applied the search strings to the following six databases: *AISeL, EBSCO, Science Direct, ProQuest ABI Inform, IEEE Xplore, and ACM Digital Library.*

Paper selection: The database search resulted in 2885 hits. Titles, abstracts, and keywords were screened to fit the abovementioned definition and application to the scope of our study. We excluded papers that did not refer to PM or that applied PM in another domain than education. Multiple papers were excluded due to a different research scope described in their abstract, e.g., several papers described mineral mining or machine learning and PM outside of an educational setting and thus were eliminated from our sample. This screening process resulted in 69 empirical papers that mentioned that they applied PM in education in the course of their study. After the elimination of all duplicates, 64 relevant papers were left. Afterwards, a forward and backward search was carried out according to Webster and Watson (2002). Through screening the references, two articles were added to the list, resulting in 66 relevant papers. An overview of the publication dates of our relevant papers on EPM applications is illustrated in Figure 1. Moreover, we found 16 additional papers that dealt with a conceptual viewpoint on EPM, e.g., through a literature review. These papers were not part of our empirical database; however, they were used to gain information for constructing the taxonomy in step 2.

Paper analysis: The 66 relevant papers were analyzed from a concept-centric perspective based on an abductive approach. To aggregate the insights from identified EPM studies, we developed a list of master codes and master code descriptions representing different EPM scenarios. Moreover, we initially identified design elements of PM applications based on a TML perspective from a socio-technical system viewpoint (Alavi and Leidner 2001; Bostrom and Heinen 1977; Gupta and Bostrom 2009) (i.e., user, task, technology, and structure) provided by the studies. This process was iterative and required multiple rounds of coding of the identified papers by three researchers. The process started with three of the researchers independently coding a subset of ten randomly chosen articles. For each of these ten studies, we listed the scenario in which PM was used and identified the different design characteristics based on the used technology, the learning context, the user, and the overall pedagogical structure. We conducted a workshop to discuss how to combine the design characteristics of PM across studies, which resulted in a distinct list of characteristics and descriptions. During the next iterations, one researcher coded a batch of 25 articles based on the list and definitions from before. Afterwards, a group of three researchers and two practitioners who work for a PM software vendor met to discuss the findings. If the coding was unclear, the PM characteristics, as well as the descriptions, were discussed and corrected until an agreement was reached. In each iteration, we added new characteristics of PM to our list according to the socio-technical dimensions and descriptions until all the papers were coded.

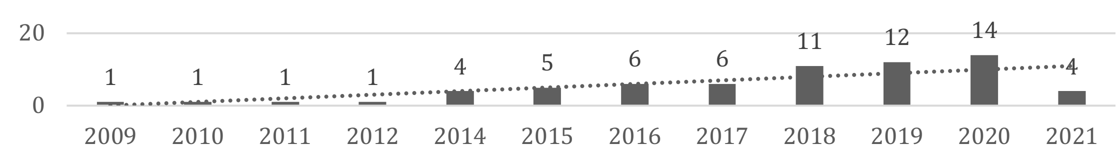

Figure 1. Overview of publication of year of relevant papers on EPM applications.

Step 2: Taxonomy Development based on Five Iterations

We aimed to provide a framework through the development of a comprehensive taxonomy. Therefore, we followed the method presented by Nickerson et al. (2013), which has been applied to several other studies in the IS field (e.g., Zierau et al. 2020). The method follows an iterative and structured process for developing taxonomies and is grounded on theoretical foundations (deduction) and empirical evidence (induction). By applying the method of Nickerson et al. (2013), we developed different dimensions and characteristics based on the published studies about PM in education we found in our literature review and the empirical evidence of specific meta-attributes. The development of a taxonomy usually starts with defining a specific phenomenon of interest, also called a meta-characteristic. The creation of all dimensions and characteristics should be based on contributing to this meta-characteristic. Our *meta-characteristic* is described by the aim *to systematically classify the design characteristics of EPM applications to enable the effective identification, design, and comparison of process mining applications in educational scenarios*. To do so, we looked at EPM design characteristics from a TML perspective with a socio-technical system viewpoint (Bostrom and Heinen 1977; Gupta and Bostrom 2009) to form a holistic contribution to the current knowledge of process mining in the specific domain of education.

Nickerson et al. (2013) suggest different subjective and objective criteria, also called ending conditions, which a taxonomy must fulfil after the iterative taxonomy development process. We defined the following ending conditions (EC) to determine when to terminate the iterative process.

A) At least one object (text feature) is classified under every characteristic of every dimension.

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

C) Dimensions and characteristics are unique and are not repeated.

D) Every known object (text feature) is classified in the taxonomy.

All of the ending conditions should be met by the final taxonomy. The taxonomy that we present in this paper has met all of the ending conditions. To the best of our knowledge, we believe that we have represented all of the PM applications in the educational domain. We first conducted a conceptual-to-empirical cycle, followed by four empirical-to-conceptual cycles. We inductively challenged the latest status of the taxonomy by classifying PM scenarios in education and revising the existing dimensions and characteristics accordingly. The development of our taxonomy is illustrated in Figure 2. In total, we classified each of the 66 studies in five iterations until all ECs were met.

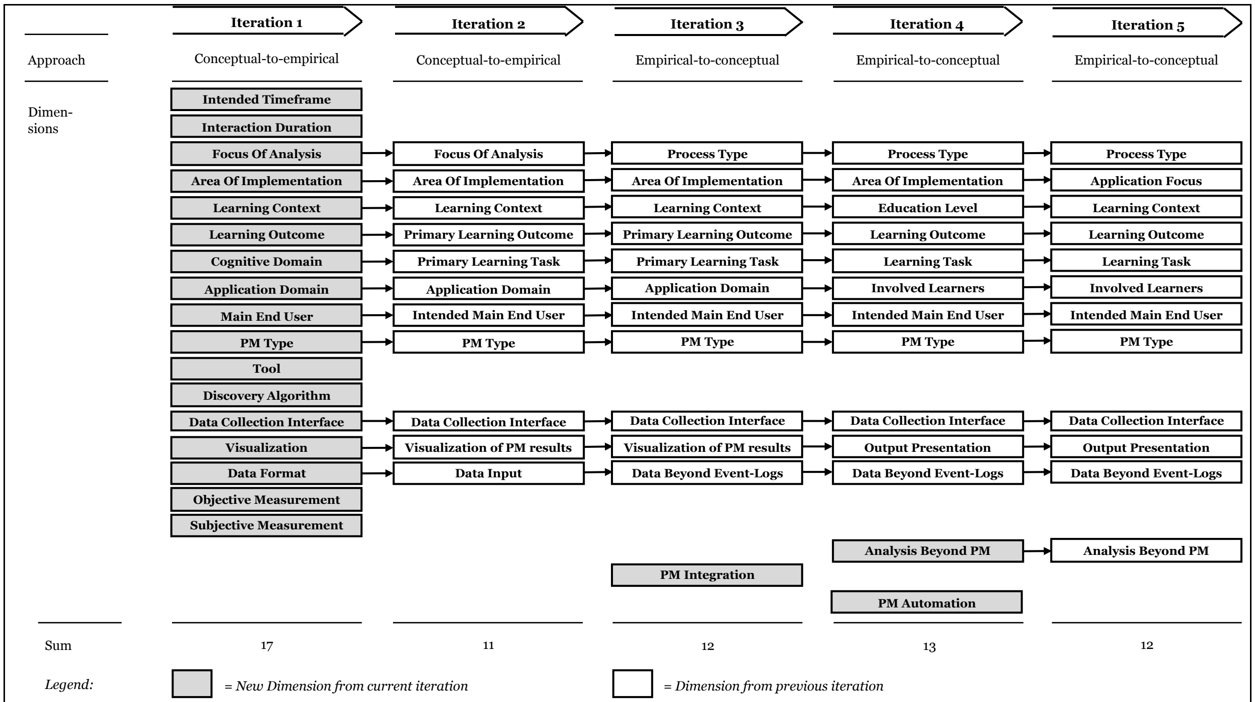

Figure 2. Taxonomy development iterations based on Nickerson et al. (2013).

Step 3: Taxonomy Evaluation based on Semi-Structured Interviews

To ensure the quality of our taxonomy, we assessed it against the following five criteria: *conciseness, robustness, comprehensibility, extendibility*, and *explanatory power* (Nickerson et al. 2013). Hence, to evaluate the taxonomy, we conducted semi-structured interviews with ten experts that either had expertise in EPM (five interviewees) or the design of technology-mediated learning environments (five interviewees), following the taxonomy evaluation suggestions of Szopinski et al. (2019). The shortest interview lasted 30 minutes, while the longest took 85 minutes. The interview guideline consisted of 18 open questions that were based on the five evaluation criteria. The final versions of our taxonomy, the meta-characteristic, and the exemplary PM scenarios were sent to the interviewees before the interviews. In the interviews, we asked the interviewees to comment and note any suggestions for revising and improving the taxonomy.

Concerning *understandability*, in general, the interviewees were able to discern the meaning behind the taxonomy but initially struggled with its structure. After receiving some suggestions regarding the categorization of dimensions, we restructured the taxonomy, which leads to generally positive remarks in later interviews. Regarding *conciseness,* most interviewees agreed that the taxonomy could differentiate between PM scenarios without overwhelming the reader. Multiple interviewees acknowledged that the subject matter was indeed somewhat complex, and that the removal of dimensions could lead to a loss of value. In terms of *robustness,* the interviewees generally agreed that the dimensions did not overlap and that the taxonomy could effectively differentiate between different scenarios. The inputs about *comprehensiveness* were mostly positive. There were some suggestions about additional objects, which were carefully considered regarding the *conciseness*. In the end, we restructured, for example, the *Learning Outcome* dimension and added the *Analysis beyond Process Mining* dimension to improve the *comprehensiveness*, which was well received. Lastly, focusing on *extendibility*, the interviewees did not see any reasons why the taxonomy could not be adapted and developed further, given its current state.

Step 4: Taxonomy Application for Cluster Analysis

Based on our final evaluated taxonomy, we aimed to provide further applications by identifying clusters of EPM design characteristics. Thus, we aimed to use cluster analysis as a descriptive, exploratory tool to identify natural patterns in data (Kaufman and Rousseeuw 2005). Therefore, we created a data matrix, $D(i,j)$, $1 <= i <= N$, $1 <= j <= M$, with the rows representing the N characteristics and the columns representing the M empirical studies. The matrix element $D(i,j)$ was set to 1 if publication j contained feature, *I* and *0* if not. This resulted in a binary NxM data matrix. Since the individual EPM characteristics were created from our systematic literature review, they were pre-ordered according to the dimensions of our taxonomy (see Table 2). For the publications in the columns, there was no equivalent "natural" order. They were simply organized in alphabetical order. We performed a clustering based on Ward's algorithm (Ward 1963) as an agglomerative clustering with a Euclidean metric since it is a proven method for accurate clustering with smaller data sets using the python-based API SciPy and Matplotlib for plotting the resulting matrix in a heat map. The result of hierarchical clustering is a total ordering of the objects in the form of a tree or dendrogram, the leaves of which define a similarity sequence. This sequence can be used to rearrange the columns of a matrix in such a way that the columns that are most similar to each other become directly adjacent. In other words: the publications that used similar sets of EPM characteristics were grouped closely together, which formed natural clusters. Based on the dendrogram, we identified three clusters and found six archetypes of EPM scenarios.

Taxonomy of Process Mining Design Characteristics for Education

In the following section, we present our consolidated version of the taxonomy after conducting five iterations and a revision based on the feedback from the expert interviews. All of the presented design dimensions define the relevant elements of a PM scenario in education according to the reviewed literature. Our unit of analysis for the taxonomy was a single EPM application. Therefore, we expect a researcher to iteratively apply our taxonomy when designing multiple EPM applications (e.g., a pedagogical scenario that implies two different learning objectives). Below, we will introduce the different dimensions and their characteristics as presented in Table 2. Overall, the design characteristics of an EPM application can be grouped into *four categories: purpose, user, input, and analysis*. *Purpose* and *user* incorporate the human-centered perspective around the social embedding of the pedagogical scenario of EPM. *Input* and *analysis* address the technological embedding of process mining characteristics, the analyzed data types, and how the result of the analysis is displayed (Bostrom and Heinen 1977). The differentiation among these four perspectives allowed an increase in the usability of the taxonomy regarding the cluster analysis of the reviewed papers and the derived research agenda. To that end, we strive for a precise and unambiguous description of the different classifications for each dimension to allow for a robust categorization.

	Dimensions	Characteristics					
Purpose	Application Focus	Learner's Monitoring		Learner's Evaluation		Learner's Recommendation	
	Involved Learners	Individual Learning			Collaborative Learning		
	Learning Outcome	Factual	Conceptual		Procedural	Metacognitive	
	Learning Task	Remember	Understand	Apply	Analyze	Evaluate	Create
	Process Type	Learning Process (in specific Learning Unit)		Practical Training Process		Course Taking Order	
User	Intended Main End User	Learner		Instructional designer (Educator)		Organization (Learning institution)	
	Learning Context	Kindergarten - Highschool	Higher Education		Continuous Education	Vocational Training	
Input	Data Input (Beyond Event Logs)	Image	Audio	Video	Text	None	

Analysis	Data Collection Interface	Internal LMS	MOOC Platform	Other Web-Enabled Tools	Non Web-Enabled Tools	Automatic Data Coding	Manual Data Coding	
	Process Mining Type	Discovery		Conformance		Enhancement		
	Analysis beyond Process Mining	None	Clustering (Unsupervised)	Classification (Supervised)	Rule Based	Other		
	Output Presentation	Raw Model	Graphical	Numerical	Textual			

Table 2. Taxonomy of process mining scenarios for education

Purpose: The first cluster of dimensions relates to the *purpose* of an EPM application. The *Application Focus* dimension describes the goal of a PM application. *Learner's Monitoring* applies to scenarios where a user looks at a learning process to gain specific insights from it. An exemplary research contribution would be the use of PM by Uzir et al. (2020) to foster the understanding of learning strategies. *Learner's Evaluation* defines a situation in which information derived from PM analysis is used to evaluate the process, for example, to facilitate process-based feedback, as studied by Lira et al. (2019). An application of PM in the realm of *Learner's Recommendation* is currently not covered by the literature in the educational domain but would describe a system that gives actionable recommendations to users on how to proceed with their learning process (e.g., based on scaffolding, see Winkler et al. 2020). Additionally, we recognized that application scenarios can be classified by either analyzing *individual learning*, e.g., Saint et al. (2020), or *collaborative learning*, e.g., Schoor and Bannert (2012). This differentiation is considered part of the *Involved Learners* dimension. Furthermore, the *Learning Outcome* dimension acknowledges that EPM scenarios can differ in the learning outcome that is generated by the process being analyzed. The characteristics are inspired by Krathwohl (2002), where *Factual (knowledge)* represents the knowledge of terminology, and specific details (e.g., Cerezo et al. 2020) and *Conceptual (knowledge)* encompasses the knowledge of theories and models, whereas *Procedural (knowledge)* refers to subject-specific skills and methods (e.g., Lira et al. 2019). *Metacognitive (knowledge)* is the strategic knowledge and the cognitive knowledge (e.g., argumentation skills (Wambsganss 2021) or empathy skills (Wambsganss, Niklaus, et al. 2021). Moreover, we used characteristics inspired by the same classifications of Krathwohl (2002) to make a distinction among various *Learning Tasks*. We followed Krathwohl's definition and distinguished between *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate, and Create,* as objectives that a learner can achieve during their learning processes (Krathwohl 2002). Lastly, for the *Purpose* category, the *Process Type* dimension describes which type of process is being analyzed specifically. For *Learning Processes (in a specific learning unit),* PM is used to analyze the learning path of students (e.g., Uzir et al. 2020). This includes learning material usage or navigation paths in LMS. Concerning *Course Taking Order*, PM is used to analyze the curriculum either as a whole or of a given learner (e.g., Cameranesi et al. 2017). An example of this would be course recommendations based on insights gained by analyzing the course paths of successful students. Scenarios with a focus on *Practical Processes* use PM to analyze a specific practical process during training or education. An example of this would be the analysis of a medical procedure or programming process (e.g., Mittal and Sureka 2014) for the purpose of giving procedural feedback.

User: The next set of dimensions concern the *user* of the system. As it is quite apparent that a learner would, in most cases, be the one to generate the analyzed process, thus giving no room for differentiation, we decided to differentiate scenarios by the *Intended Main End User,* who has the most insight to gain from the output generated by the EPM application. Despite the nascency of the field and the resulting difficulty to clearly distinguish and identify the end user for each case at hand, three main end users can be defined. Having a *Learner* as the intended main end user suggests that insights can be used by the learner to improve their own learning or course taking process (e.g., Cameranesi et al. 2017). The *Instructional Designer (Educator)* characteristic describes an end user who mainly benefits from the insights and can therefore improve the instructional design or intervene in real-time, for example, to give directed feedback to students (e.g., Lira et al. 2019). For the *Organization (Learning Institution)* characteristic, the provider of the educational environment is the main benefactor of the application. This applies to cases in which PM is used to analyze MOOCs to improve the dropout rate, for instance (e.g., Rizvi et al. 2018). Lastly, *Learning Context* describes the learning environment in which the scenario takes place. We distinguished between *Kindergarten - Highschool* (e.g., Gomez et al. 2021), *Higher Education* (Engelmann and Bannert 2019), *Continuous Education* (Ariouat et al. 2016), including workforce training and programs for personal improvement and *Vocational Training* in the cases of education leading to a career using a craft or trade.

Input: The *Input* cluster describes the origin of the analyzed data. We identified that although most applications use already existing event data for analysis, for example, by automatically collecting data through an LMS, some use *Data Beyond Event Logs,* such as *Audio* (Nguyen et al. 2021), *Video* (Lira et al. 2019), or *Text* (Mittal and Sureka 2014), from which event data can either be manually or automatically derived or supplemented. *Image* data could also be used, although it is not currently found in any study. The *Data Collection Interface* describes the system through which event data is being collected. In most applications, event data was either collected using an *Internal LMS* (Juhaňák et al. 2019) or a *MOOC Platform* (Rizvi et al. 2018) with *Other Web-Enabled Tools, Non Web-Enabled Tools* (e.g., Doleck et al. 2016), and *Automatic and Manual Data Coding,* e.g., through the coding of video data, describing the remaining cases.

Analysis: The *Analysis* describes the means through which the processes are being analyzed. Although a lot of different techniques are imaginable in this category, we decided to mainly distinguish between the basic *Process Mining Types* being applied as well as the type of analysis being used which goes beyond the traditional PM functions. Regarding *Process Mining Types*, only applications that specifically mentioned the use of *Discovery* and *Conformance* were identified. *Discovery* uses PM to construct a process model by analyzing an event log. An example of this would be to use an event log from an LMS to construct a process model of a learning process (e.g., Rogiers et al.2020). *Conformance* describes a scenario where PM is used to compare the mined model based on the event log to a preexisting model of the process. For example, to assess compliance with course order recommendations (e.g., Cameranesi et al. 2017). Process *Enhancement* was not explicitly used in any of the reviewed publications, though the uses for this are imaginable, for example, to extend a learner's learning process model. *Analysis beyond Process Mining* describes the types of analyses being used that either precede or extend the traditional PM analysis. This does not encompass supplementary analyses, which are, in essence, separate from the PM application. *Clustering* can be used in conjunction with PM, for example, to segment learners according to certain criteria like grades. *Rule-based* analysis approaches can be used in a similar manner. Furthermore, *Classification* has uses, for example, in order to predict learner success based on previously mined processes. Lastly, the *Output Presentation* dimension defines the means through which the process mining results are presented to the user. The main concern here is that non-expert users need a higher abstraction of information to gain usable insights from the data. *Raw Model* implies no further form of presentation beyond the discovered model by using PM. A *Graphical* presentation describes the further visualization of results, for example, through the use of graphs. *Numerical* presentation concerns the presentation of results in the form of numbers or tables, for example, portraying fitness scores of discovered models. Lastly, T*extual* presentation describes scenarios where the information gained using PM is translated by the system into readable information, for example, recommendations or automatically generated reports.

Clusters of Educational Process Mining Applications

Our objective was to identify clusters in our reviewed literature sample to better interpret the application of certain EPM application groups and to structure the white spots in research. After storing the empirical studies in our binary *NxM data matrix,* we rearranged columns based on *Ward's algorithm* (Ward 1963). Based on the resulting dendrogram, the algorithm distinguished between three mutually exclusive and collectively exhaustive clusters of application patterns in EPM studies (C1, C2, C3), which were distinguished by the *purpose*, the *user*, the *input, and the output* of each EPM application. Based on the clusters, we further identified six distinct, more nuanced, archetypes that represented specific instantiations of the respective clusters. Cluster 1 (C1, 15 out of 66 studies, 22%) represents EPM scenarios that focus on learning applications in MOOCs or overall course path scenarios. The main reason these scenarios fit together is that their outcomes mainly benefit the organization (learning institution), e.g., where PM is used to improve dropout rates (e.g. Rizvi et al. (2018)) or find bottlenecks in a degree. These two use cases generate two derived scenario archetypes, A1 and A2. A1 focuses on monitoring the course-taking behavior of students for the benefit of the educational organization to improve a degree, curriculum, or course guideline. The A2 scenario archetype encompasses the monitoring of the learning processes of individuals in the context of MOOCs to the benefit of MOOC providers or educators for improving the learning content or generally improving dropout rates. The second identified group of studies (C2, 31 out of 66 studies, 46%) comprises EPM applications, which exclusively focus on the learning processes *(Process Type*) of individual learning scenarios of only one involved learner, with an educator as the indented main end user. Data is analyzed to gain insights about learning strategies or learning material usage. About 90% of studies in this cluster focus on factual and conceptual knowledge learning outcomes with understanding being the most prevalent learning task. Usually, the data comes from internal LMS. The A3 derived scenario archetype thus concerns itself with the monitoring of individual learning processes in higher education scenarios using LMS usage data to benefit educators in gaining insights about learning strategies. The third cluster mostly contains educational scenarios using PM to research practical training processes. The learning outcome in this cluster is almost exclusively based on procedural knowledge. Learning tasks considered in this cluster are generally of a higher nature, especially considering the share of *create* tasks employing studies (45% of the cluster). The cluster also contains most of the collaborative learning scenarios (82% of collaborative tasks of whole sample) as well as all of the application scenarios evaluating the learner's process. The two main use cases in this sample are exemplified by A4 and A5. A4 describes scenarios monitoring practical training processes consisting of *create* learning tasks, which, in the sample, are mostly represented by studies analyzing software engineering projects. The A5 archetype comprises scenarios using PM to evaluate the *learner's process* in practical training processes, with procedural learning outcomes and apply learning tasks in continuous education. This includes, for instance, studies analyzing the process of trainees performing medical procedures. An additional archetype, A6, can be identified, which contains collaborative learning scenarios in higher education, which qualifies for this cluster by the use of higher learning tasks.

Cluster	Description	Percentage of studies	Archetypes
C1	Institutional Analyses	22%	• A1: Monitoring of course taking behavior for educational institutions • A2: Monitoring of individual learning processes in MOOCs for MOOC provider or MOOC educator
C2	Learning Process Mining for Educators	46%	• A3: Monitoring of individual learning processes for factual knowledge and understand learning tasks in higher education for educational designer
C3	Practical Process Analyses for procedural and collaborative learning	32%	• A4: Monitoring of practical training processes in software engineering for procedural learning outcomes • A5: Learner's process evaluation of practical learning processes for procedural learning outcomes and apply learning tasks in continuous education • A6: Monitoring of collaborative learning processes in higher education

Table 3. Identified EPM clusters (C1, C2, C3) and archetypes (A1-A6).

Discussion and Research Agenda

In this section, we aim to propose a preliminary research agenda that provides promising points for future research on EPM and illustrates how they can be positioned based on the dimensions and characteristics of our taxonomy. As our literature review, the taxonomy, and the cluster analysis emphasize, the identified dimensions enable a distinct perspective on studying EPM, which also relates to the theoretical perspective of TML based on the socio-technical viewpoint we aimed to apply (Bostrom and Heinen 1977; Gupta and Bostrom 2009). To provide further insights on PM for educational purposes, we seek to identify patterns of EPM applications by illustrating the frequency of design characteristics according to the characteristics we found in our study sample. Most applications did not apply EPM beyond process discovery. In fact, only 29% of the analyzed studies checked models for their conformance, and no studies specifically mentioned the use of enhancement techniques. Regarding the process type, a large extent of the research sample focused on general learning processes (70%). Course taking order (10%) and the analysis of the practical training process (30%) left room for further exploration. Moreover, based on the sample, EPM was almost exclusively used for the purpose of monitoring students to get general insights about learning processes (92%). An overall lack of scenarios actively confronting the learner with the gained insights and a lack of systems repeatedly and portably employing PM in the domain lead to applications using PM for evaluation purposes being very rare (8%) and applications using PM to facilitate systematic recommendations being non-existent (0%). Furthermore, most (73%) of the research was concentrated on analyzing processes in a *higher education* context. *Continuous education*, especially focusing on MOOCs and *practical training,* was a somewhat distant secondary context being researched in 21% of the papers. *Kindergarten – Highschool* rarely appeared, with 6% of the studies in the sample mentioning this specific context and *Vocational Training* scenarios being currently unexplored.

Taxonomy Dimension	Research Opportunities	Research Questions
Learning Purpose	• EPM scenarios focusing on learner support through actionable recommendations • PM used for specifically evaluating students' processes as well as course path analysis	• What are the implications of integrating EPM output in real life cases (over a longer period of time)?
User	• PM used in Kindergarten – Highschool or Vocational Training • PM used to analyze learners with certain deficiencies • PM scenarios specifically addressing learners themselves	• How do affordances regarding EPM analyses differ across specific end users and contexts? • How can learner perceptions be integrated in existing EPM? • How can an ethically desirable treatment of individual user data be ensured?
Input	• Exploration of different data sources supplementing or leading to event logs • Consideration of automatic coding of different data types (e.g., automatic coding of think aloud audio or video recordings of medical procedures)	• How can all relevant data input in learning systems be identified? • How can data input other than event logs be best integrated into EPM?
Analysis	• Exploration of different PM types as well as their interrelatedness • Consideration of complexity of educational learning processes • Analysis of techniques beyond traditional PM uses	• How do different PM types relate to each other (in practice)? • What are education-specific PM methodologies / techniques / algorithms? • What is the potential of deep learning techniques for predicting learner success? • How can PM output be best presented for a specific learning case at hand?

Table 4. Preliminary research agenda on educational process mining based on our taxonomy

A sharp rise in publications on PM in education over the past ten years points towards the relevance and technical feasibility of PM to significantly enhance learning-related processes and analyses. However, our taxonomy and deducted conclusions illustrate how the current research has arrived at some initial, theoretical contributions to PM in education while leaving several important research avenues neglected. As an overarching research gap, we see a lack of empirical evidence, both qualitative and quantitative, of PM systems in educational settings. While the current research body mostly discusses the potential of PM in education with theoretical contributions, actual evaluations of PM with deployed systems and users in the field are missing. More so, current empirical studies investigate PM systems uncoupled from the actual deployment context. In that sense, past data from educational systems are deployed for PM analyses without integrating findings back into the use context. There is a distinct lack of repeatable and portable applications of PM. As a result, the implications of PM analyses in the field of education are largely unknown and can only be hypothesized. Turning towards the individual clusters, several white spots in the research can be identified, for instance, regarding the PM types and the analyses conducted beyond a classic PM analysis. We see these white spots as a basis for future research, which we will elaborate on below. Building on the insights that were gained through our research and by linking our dimensions and characteristics to our theoretical background, we propose the following preliminary research agenda as presented in Table 4.

Contributions and Limitations

From a theoretical perspective, we can make the following contributions. First, we integrate the current literature, including literature reviews on EPM, by developing a new taxonomy that goes beyond the current classifications, structuring, and grouping of design characteristics of EPM applications from a TML perspective and a socio-technological viewpoint (Bostrom and Heinen 1977; Gupta and Bostrom 2009). With a common classification of EPM design characteristics assuming a holistic perspective, we provide a better understanding of what needs to be considered when identifying, designing and comparing PM applications in educational scenarios. By combining the dimensions and characteristics of our taxonomy, we can now assist researchers and practitioners in better assuming which design characteristics influence which learning outcomes. This provides more room for future research and more detailed insights on how to apply, embed or design EPM applications. Second, we identify and classify new dimensions and characteristics beyond the technical perspective of PM that are part of a pedagogical scenario and play an important role for the learning success of students. Thus, we combine the TML perspective in IS research (Bostrom and Heinen 1977; Gupta and Bostrom 2009) with the components and characteristics of EPM scenarios of different dimensions, which contributes to a holistic perspective on PM in education. Third, the different characteristics are categorized and clustered in three distinct groups (C1, C2, C3) and six corresponding archetypes. These clusters and archetypes will help researchers and practitioners gain an overview of existing research on EPM and find the corresponding gaps. We hope that our research findings offer a point of departure to further develop guidelines or frameworks for identifying the requirements towards PM (i.e., chosen analyses, process mining type) for a specific educational use case.

Given the immense growth of EPM and the potential of PM to enhance the individual learning of students through scaffolding or adaptive feedback interventions, further research on this topic is warranted. Such research will require a solid classification and a theoretical understanding. In this paper, we offer such an understanding by presenting our taxonomy, its application through a cluster analysis, and a research agenda. In that regard, we expand the knowledge base on design characteristics significant to how we can embed PM and enhance the individual's learning experiences. Moreover, we offer researchers a more nuanced perspective when studying EPM based on a preliminary research agenda. Therefore, we also contribute to the forming of the young and nascent research field of EPM. From a practical perspective, a common understanding of the design characteristics of EPM applications also gives rise to several insights for the educational technology field and the related applications. Our systematic classification of EPM scenarios enables researchers and practitioners to more effectively design, evaluate, compare, and theorize how different technological embeddings of the young field of PM impact the students' learning outcomes in a specific pedagogical scenario and task. For instance, at a basic level, our taxonomy determines several high-level design decisions an educational designer has to make when configuring a PM application to inform, monitor, or evaluate a student's learning process. Based on our taxonomy, the designers can now identify different kinds of EPM applications by combining different characteristics of EPM, depending on the pedagogical scenario, the target group, or what the EPM application is used for. Our research suffers from several limitations that provide avenues for future research. First, we only identified design characteristics of EPM applications based on scientific publications since our main objective was to categorize EPM characteristics from a socio-technical perspective. Therefore, future research may adjust and extend our taxonomy based on an in-depth analysis of real-life use-cases. Second, based on technological advancements, more design elements may need to be added in the future. However, based on our evaluation with the experts, we believe that our taxonomy resembles a sufficient state-of-the-art tool for analyzing design characteristics of EPM applications. Third, our research is neglecting an ethical consideration of the use of personal learner data. Future research could therefore shed light on the requirements for EPM based on ethical and legal standards, such as Wambsganss, Höch, et al. 2021.

Conclusion

In sum, our results provide deeper insights into the design characteristics of EPM applications and support researchers by systematizing and synthesizing research in how to design EPM applications, e.g., towards individual learning support of students in large-scale or distance-learning scenarios. We conducted five iterations of reviewing the literature, one is conceptually based in the current EPM classification literature and four being empirically grounded on the set of the 66 articles on EPM applications we identified through a systematic literature review in the field of IS and HCI. We evaluated the taxonomy with ten domain experts and performed a cluster analysis, where we identified three clusters and six archetypes. With our taxonomy and the related clusters, we provide possible research topics when designing or applying EPM. Therefore, researchers and practitioners can use the results of our study to derive individual EPM design characteristics and better research the influence of EPM on student learning outcomes.

Acknowledgments

We thank the Swiss National Science Foundation for funding parts of this research (200207).

References

Van der Aalst, W. M. P., Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., Alves de Medeiros, A. K., Song, M., and Verbeek, H. M. W. 2007. "Business Process Mining: An Industrial Application," Information Systems (32:5), Pergamon, pp. 713–732.

Van der Aalst, W. M. P., and Weijters, A. J. M. M. 2004. "Process Mining: A Research Agenda," Computers in Industry, Elsevier, pp. 231–244.

Van Der Aalst, W. 2012. "Process Mining," Communications of the ACM (55:8), pp. 76–83.

Alavi, M., and Leidner, D. E. 2001. "Research Commentary: Technology-Mediated Learning - A Call for Greater Depth and Breadth of Research," Information Systems Research (12:1), INFORMS Inst.for Operations Res.and the Management Sciences, pp. 1–10.

Ariouat, H., Hicheur Cairns, A., Barkaoui, K., Akoka, J., Khelifa, N., Hicheur, A., and Hicheur CAIRNS, A. 2016. A Two-Step Clustering Approach for Improving Educational Process Model Discovery A Two-Step Clustering Approach for Improving Educational Process Model Discovery., pp. 38–43.

Azevedo, R., and Hadwin, A. F. 2005. "Scaffolding Self-Regulated Learning and Metacognition - Implications for the Design of Computer-Based Scaffolds," Instructional Science, pp. 367–379.

Bailey, K. D. 1994. Typologies and Taxonomies: An Introduction to Classification Techniques, CA: Sage.

Bogarín, A., Cerezo, R., and Romero, C. 2018. "A Survey on Educational Process Mining," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (8:1), p. 1230.

Bostrom, R. P., and Heinen, J. S. 1977. "MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes," MIS Quarterly (1:3), p. 17.

Brinton, C. G., Rill, R., Ha, S., Chiang, M., Smith, R., and Ju, W. 2015. "Individualization for Education at Scale: MIIC Design and Preliminary Evaluation," IEEE Transactions on Learning Technologies (8:1), Institute of Electrical and Electronics Engineers, pp. 136–148.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," Communications of the Association for Information Systems (37:1), pp. 205–224.

Cameranesi, M., Diamantini, C., Genga, L., and Potena, D. 2017. "Students' Careers Analysis: A Process Mining Approach," in ACM International Conference Proceeding Series (Vol. Part F129475), Association for Computing Machinery, June 19.

Cerezo, R., Bogarín, A., Esteban, M., and Romero, C. 2020. "Process Mining for Self-Regulated Learning Assessment in e-Learning," Journal of Computing in Higher Education (32:1), pp. 74–88.

Costa, J., Azevedo, A., and Rodrigues, L. 2020. AIS Electronic Library (AISeL) Educational Process Mining Based on Moodle Courses : A Review of Literature Educational Process Mining Based on Moodle Courses : A Review of Literature.

Dahlstrom, E., Christopher Brooks, D., and Bichsel, J. 2014. "The Current Ecosystem of Learning Management Systems in Higher Education: Student, Faculty, and IT Perspectives."

Dillenbourg, P., Järvelä, S., and Fischer, F. 2009. "The Evolution of Research on Computer-Supported Collaborative Learning," in Technology-Enhanced Learning: Principles and Products, N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, and S. Barnes (eds.), pp. 3–19.

Doleck, T., Jarrell, A., Poitras, E. G., Chaouachi, M., and Lajoie, S. P. 2016. Examining Diagnosis Paths: A Process Mining Approach.

Engelmann, K., and Bannert, M. 2019. Analyzing Temporal Data for Understanding the Learning Process Induced by Metacognitive Prompts.

Eom, S. B., Wen, H. J., and Ashill, N. 2006. "The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Empirical Investigation*," Decision Sciences Journal of Innovative Education (4:2), Wiley, pp. 215–235.

Ghazal, M. A., Ibrahim, O., and Salama, M. A. 2018. "Educational Process Mining: A Systematic Literature Review," Proceedings - 2017 European Conference on Electrical Engineering and Computer Science, EECS 2017, IEEE, pp. 198–203.

Gomez, M. J., Ruipérez-Valiente, J. A., Martínez, P. A., and Kim, Y. J. 2021. "Applying Learning Analytics to Detect Sequences of Actions and Common Errors in a Geometry Game," Sensors (Switzerland) (21:4), MDPI AG, pp. 1–16.

Gupta, S., and Bostrom, R. P. 2009. "Technology-Mediated Learning: A Comprehensive Theoretical Model," Journal of the Association for Information Systems (10:9), Association for Information Systems, pp. 686–714.

Hattie, J., and Timperley, H. 2007. "The Power of Feedback," Review of Educational Research (77:1), pp. 81–112.
Hone, K. S., and El Said, G. R. 2016. "Exploring the Factors Affecting MOOC Retention: A Survey Study," Computers and Education (98), Elsevier Ltd, pp. 157–168. (https://doi.org/10.1016/j.compedu.2016.03.016).
Ifenthaler, D., and Gibson, D. 2019. Opportunities for Analytics in Challenge-Based Learning, Springer, Singapore, pp. 55–68.
Janson, A., Söllner, M., and Leimeister, J. M. 2020. "Ladders for Learning: Is Scaffolding the Key to Teaching Problem-Solving in Technology-Mediated Learning Contexts?," Academy of Management Learning and Education (19:4), George Washington University, pp. 439–468.
Johnson, O. A., Ba Dhafari, T., Kurniati, A., Fox, F., and Rojas, E. 2019. "The ClearPath Method for Care Pathway Process Mining and Simulation," in Lecture Notes in Business Information Processing (Vol. 342), Springer Verlag, September 9, pp. 239–250. (https://doi.org/10.1007/978-3-030-11641-5_19).
Juhaňák, L., Zounek, J., and Rohlíková, L. 2019. "Using Process Mining to Analyze Students' Quiz-Taking Behavior Patterns in a Learning Management System," Computers in Human Behavior (92), Elsevier Ltd, pp. 496–506.
Kaufman, L., and Rousseeuw, P. J. 2005. Finding Groups in Data : An Introduction to Cluster Analysis.
Kellogg, S., Booth, S., and Oliver, K. 2014. "A Social Network Perspective on Peer Supported Learning in MOOCs for Educators," International Review of Research in Open and Distance Learning (15:5), Athabasca University, pp. 263–289. (https://doi.org/10.19173/irrodl.v15i5.1852).
Krathwohl, D. R. 2002. "A Revision of Bloom's Taxonomy: An Overview," Theory into Practice (41:4), Taylor \& Francis, pp. 212–218.
Kulik, J. A., and Fletcher, J. D. 2016. "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review," Review of Educational Research (86:1), pp. 42–78.
Lira, R., Salas-Morales, J., Leiva, L., De La Fuente, R., Fuentes, R., Delfino, A., Nazal, C. H., Sepúlveda, M., Arias, M., Herskovic, V., and Munoz-Gama, J. 2019. "Process-Oriented Feedback through Process Mining for Surgical Procedures in Medical Training: The Ultrasound-Guided Central Venous Catheter Placement Case †," International Journal of Environmental Research and Public Health (16:11), MDPI AG.
Martens, E. 2003. Methodik Des Ethik- Und Philosophieunterrichts : Philosophieren Als Elementare Kulturtechnik, (1. Aufl.), Siebert.
Mittal, M., and 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 2014 - Proceedings, New York, NY, USA: Association for Computing Machinery, May 31, pp. 344–353.
Nguyen, H., Lim, K. Y., Wu, L. L., Fischer, C., and Warschauer, M. 2021. "'We're Looking Good': Social Exchange and Regulation Temporality in Collaborative Design," Learning and Instruction (74), Elsevier Ltd, p. 101443.
Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A Method for Taxonomy Development and Its Application in Information Systems," European Journal of Information Systems (22:3), Taylor & Francis, pp. 336–359.
Okoye, K. 2019. "A Systematic Review of Process Modelling Methods and Its Application for Personalised Adaptive Learning Systems," Journal of International Technology and Information Management (27:3), pp. 23–46.
Panetta, K. 2019. "Gartner Top 10 Strategic Technology Trends," Gartner.Com, pp. 1–4.
Rizvi, S., Rienties, B., and Rogaten, J. 2018. "Temporal Dynamics of MOOC Learning Trajectories," in ACM International Conference Proceeding Series, New York, NY, USA: Association for Computing Machinery, October 1, pp. 1–6.
Rogiers, A., Merchie, E., and van Keer, H. 2020. "Opening the Black Box of Students' Text-Learning Processes: A Process Mining Perspective," Frontline Learning Research (8:3), pp. 40–62.
Rojas, E., Munoz-Gama, J., Sepúlveda, M., and Capurro, D. 2016. "Process Mining in Healthcare: A Literature Review," Journal of Biomedical Informatics, Academic Press Inc., pp. 224–236.
Romero, C., and Ventura, S. 2020. Educational Data Mining and Learning Analytics: An Updated Survey.
Roth, H. 1970. "Pädagogische Psychologie Des Lehrens Und Lernens."
Saint, J., Whitelock-Wainwright, A., Gasevic, D., and Pardo, A. 2020. "Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning from Trace Data," IEEE Transactions on Learning Technologies (13:4), Institute of Electrical and Electronics Engineers Inc., pp. 861–877.

Schöbel, S. M., Janson, A., and Söllner, M. 2020. “Capturing the Complexity of Gamification Elements: A Holistic Approach for Analysing Existing and Deriving Novel Gamification Designs,” European Journal of Information Systems (29:6), pp. 641–668.
Schoor, C., and Bannert, M. 2012. “Exploring Regulatory Processes during a Computer-Supported Collaborative Learning Task Using Process Mining,” Computers in Human Behavior (28:4), Elsevier Science Publishers B. V. PUB568 Amsterdam, The Netherlands, The Netherlands , pp. 1321–1331.
Schumacher, C., and Ifenthaler, D. 2021. “Investigating Prompts for Supporting Students’ Self-Regulation – A Remaining Challenge for Learning Analytics Approaches?,” Internet and Higher Education (49).
Seaman, J. E., Allen, I. E., and Seaman, J. 2018. “Higher Education Reports - Babson Survey Research Group.”
Sharma, P., and Hannafin, M. 2007. “Scaffolding in Technology-Enhanced Learning Environments,” Interactive Learning Environments (15:1), Routledge , pp. 27–46.
Siemens, G. 2013. “Learning Analytics: The Emergence of a Discipline,” American Behavioral Scientist (57:10), SAGE PublicationsSage CA: Los Angeles, CA, pp. 1380–1400.
Syngene Research LLP. 2019. “Global E-Learning Market Analysis 2019.”
Szopinski, D., Schoormann, T., and Kundisch, D. 2019. “BECAUSE YOUR TAXONOMY IS WORTH IT: TOWARDS A FRAMEWORK FOR TAXONOMY EVALUATION,” in Proceedings of the 27th European Conference on Information Systems (ECIS).
Uzir, N. A. A., Gaševic, D., Jovanovic, J., Matcha, W., Lim, L. A., and Fudge, A. 2020. “Analytics of Time Management and Learning Strategies for Effective Online Learning in Blended Environments,” in ACM International Conference Proceeding Series, pp. 392–401.
Vygotsky, L. S. 1980. Mind in Society: The Development of Higher Psychological Processes, Harvard university press.
Wambsganss, T. 2021. “Designing Adaptive Argumentation Learning Systems Based on Artificial Intelligence,” in Conference on Human Factors in Computing Systems - Proceedings.
Wambsganss, T., Höch, A., Zierau, N., and Söllner, M. 2021. “Ethical Design of Conversational Agents: Towards Principles for a Value-Sensitive Design,” in Proceedings of the 16th International Conference on Wirtschaftsinformatik (WI),
Wambsganss, T., Küng, T., Söllner, M., and Leimeister, J. M. 2021. “ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills,” in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.
Wambsganss, T., Niklaus, C., Cetto, M., Söllner, M., Leimeister, J. M., and Handschuh, S. 2020. “AL : An Adaptive Learning Support System for Argumentation Skills,” in ACM CHI Conference on Human Factors in Computing Systems, pp. 1–14.
Wambsganss, T., Niklaus, C., Söllner, M., Handschuh, S., and Leimeister, J. M. 2021. “Supporting Cognitive and Emotional Empathic Writing of Students,” in The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021).
Wambsganss, T., Winkler, R., Söllner, M., and Leimeister, J. M. 2020. “A Conversational Agent to Improve Response Quality in Course Evaluations,” in ACM CHI Conference on Human Factors in Computing Systems.
Ward, J. H. 1963. “Hierarchical Grouping to Optimize an Objective Function,” Journal of the American Statistical Association (58:301), JSTOR, p. 236.
Webster, J., and Watson, R. T. 2002. “Analyzing the Past to Prepare for the Future: Writing a Literature Review.,” MIS Quarterly (26:2), xiii–xxiii.
Winkler, R., Hobert, S., Salovaara, A., Söllner, M., and Leimeister, J. M. 2020. “Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent,” in Conference on Human Factors in Computing Systems - Proceedings.
Winkler, R., and Söllner, M. 2018. “Unleashing the Potential of Chatbots in Education : A State-Of-The-Art Analysis . In : Academy of Management,” Meeting, Annual Chicago, A O M.
Wood, D., Bruner, J. S., and Ross, G. 1976. “THE ROLE OF TUTORING IN PROBLEM SOLVING,” Journal of Child Psychology and Psychiatry (17:2), Pergamon Press, pp. 89–100.
Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., and Leimeister, J. M. 2020. “The Anatomy of User Experience with Conversational Agents : A Taxonomy and Propositions of Service Clues,” in International Conference on Information Systems (ICIS)., pp. 1–17.