



OPEN ACCESS

EDITED BY

Leman Figen Gul,
Istanbul Technical University, Türkiye

REVIEWED BY

Luis Alex Alzamora De Los Godos Urcia,
Norbert Wiener Private University, Peru
Fazla Rabbi,
Arkansas State University, United States

*CORRESPONDENCE

Rosaine F. Semler
✉ rosainesemler@utfpr.edu.br

RECEIVED 11 December 2024

ACCEPTED 23 May 2025

PUBLISHED 18 June 2025

CITATION

Semler RF, Semler JR, Wehrmeister MA,
Southier LFP, Casanova D and Teixeira M
(2025) Educational process mining: literature
classification, gaps, and emerging
opportunities. *Front. Educ.* 10:1543761.
doi: 10.3389/feduc.2025.1543761

COPYRIGHT

© 2025 Semler, Semler, Wehrmeister,
Southier, Casanova and Teixeira. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Educational process mining: literature classification, gaps, and emerging opportunities

Rosaine F. Semler^{1*}, Jhonnatan R. Semler¹,
Marco A. Wehrmeister¹, Luiz F. P. Southier², Dalcimar Casanova²
and Marcelo Teixeira²

¹Graduate Program in Electrical Engineering and Industrial Informatics (CPGEI-CT), Federal University of Technology - Paraná (UTFPR), Curitiba, Brazil, ²Graduate Program in Electrical and Computer Engineering (PPGEEC-PB), Federal University of Technology - Paraná (UTFPR), Pato Branco, Brazil

Process Mining (PM) is a well-known approach for workflow analysis and has the *Educational PM* (EPM) as its education-oriented version. Despite promising applications, the EPM literature landscape is quite unclear in reporting the bridge between the existing tools, techniques, research groups, main frontiers, and, especially, main directions to guide future efforts. These gaps induce initiatives to be conducted empirically and disconnected from each other, preventing efforts from converging. This paper presents a *Systematic Literature Review* (SLR) that collects a reliable set of results on EPM and classifies their predominant profile and contributions. A total of 4,312 articles were identified, of which only 35 remained after removing duplicates and applying exclusion criteria. After peer review, 5 more articles were removed, and the references of the 30 articles were subjected to snowballing. This resulted in 28 more candidate articles, from which 14 remained after applying the exclusion criteria, and were joined with the other 30, totaling 44 articles. After closer and individual inspection, 28 articles remained to compose the final portfolio. They were then analyzed, and insights were provided based on their combined contributions, which allowed us to evidence the main gaps in EPM and how they could be fulfilled in future research. These findings can be used as a starting point for initiatives that aim to demarcate new frontiers of EPM.

KEYWORDS

process mining, educational process mining, intelligent decision making, literature review, conformance checking, process mining techniques

1 Introduction

Nowadays, a substantial amount of data is recorded across multiple domains every second. Semantically, most of these data are associated with events that occur in the real world and about which one aims to store information. As such, events offer rich opportunities for gathering insights to comprehend and possibly improve real phenomena. This movement drives the field of *Process Mining* (PM), which aims to uncover, verify, and enhance workflows using event data (Van Der Aalst, 2016). While traditional workflows typically depend on human interpretation and modeling, PM offers an automated construction mechanism that leads to workflow models automatically, by extracting and processing temporal details of event logs.

In the literature, PM has been extensively explored in specific domains such as healthcare, information and communication technology, logistics, and the industrial sector (dos Santos Garcia et al., 2019). In contrast, the application of PM in the educational context, the so-called *Educational Process Mining* (EPM), is still limited. Although some

studies address issues such as the detection of learning styles (Wang et al., 2019), course architecture (Salazar-Fernandez et al., 2021), and the analysis of student-teacher interactions in online environments (Domínguez et al., 2021; Hachicha et al., 2021; Liu et al., 2022; dos Santos Neto et al., 2022; AlQaheri and Panda, 2022), many cases focus on Data Mining and Learning Analytics techniques without considering the process perspective. Thus, EPM emerges as a promising alternative for analyzing and improving educational processes upon processing educational data logs.

When considering existing literature reviews on the topic of EPM, four articles were published between 2017 and 2022. Bogarín et al. (2018b) focuses on applying EPM to analyze interactions in learning environments; Sypsas and Kalles (2022) emphasizes the practical applicability of EPM; Ghazal et al. (2017) approaches classifications of methods and types of achievements; finally, Dutt et al. (2017) reviews the broader area of educational data mining, including EPM as a subset. Although some reviews have addressed the implementation of data mining in education, no comprehensive review has been found that systematically compiles and examines where process mining has been employed in education.

Given the lack of up-to-date reviews on EPM, this article presents a structured and interpretative synthesis of recent research 2018–2025(April) using a systematic literature review (SLR) methodology. From an initial pool of 4,312 articles, a rigorous selection process—comprising duplicate removal, exclusion criteria, peer review, and snowballing—resulted in a final portfolio of 28 relevant studies. They were then interpreted and used to map and classify the existing efforts and to connect their objectives, methodologies, and outcomes. Key patterns were identified in the field, such as the predominance of discovery-focused research, frequent use of Moodle datasets, reliance on Inductive and Heuristic Miner algorithms, and limited methodological validation. In addition, our results reveal underexplored areas like conformance, enhancement, and predictive modeling. This integrative analysis also captures the most active authors, institutions, venues, and techniques, offering not only a descriptive overview but also a critical foundation for future research and methodological advancement in EPM.

The manuscript is organized as follows: PM and EPM are presented in Section 2; Section 3 describes the research method, which is interpreted in Section 4; Section 5 answers the research questions which are the foundation for the literature classification proposal in Subsection 5.1.1; finally, Section 7 presents the final remarks.

2 Process mining

PM systematically analyzes data to provide insights, identify bottlenecks, anticipate problems, recommend countermeasures, and optimize processes (Van Der Aalst, 2016). For example, PM can be applied to track a student's progress through various subjects in an undergraduate course. Despite common steps, variations in processes often occur, resulting in deviations, repetitions, or multiple paths. Each unique instance of such a process is called a “case,” where different choices made by students, such as selecting various subjects, represent distinct cases.

The key elements in PM are activities (steps in the process), events (instances of activities executed at specific times), and event logs (records of these events). PM utilizes these logs to generate process models, detect deviations, and suggest improvements. It includes three main phases: discovery (creating process models from data), conformance checking (comparing actual processes to expected models), and enhancement (proposing optimizations). Additionally, PM can explore various perspectives like control-flow (activity sequences), organizational structure (resource roles), and temporal analysis (event frequency), providing a comprehensive view of process dynamics and enabling data-driven decision-making.

Figure 1 illustrates a typical PM pipeline. After data extraction, the process model is discovered, which supports conformance analysis, checking, predictions, and general knowledge discovery that can lead to process enhancement.

Each recorded event is formally represented as a tuple ($\pm\text{Event} = \pm\text{Activity}, \pm\text{Time}, \pm\text{Instance}$), where $\pm\text{Activity}$ denotes the educational action performed (e.g., submitting a quiz or accessing a forum), $\pm\text{Time}$ corresponds to the timestamp of the action, and $\pm\text{Instance}$ identifies the specific case (i.e., the student). By chronologically ordering the events associated with a case, it is possible to construct a trace $\pm\text{Trace}$ that characterizes the evolution of the learning process over time (Kaymakci et al., 2010).

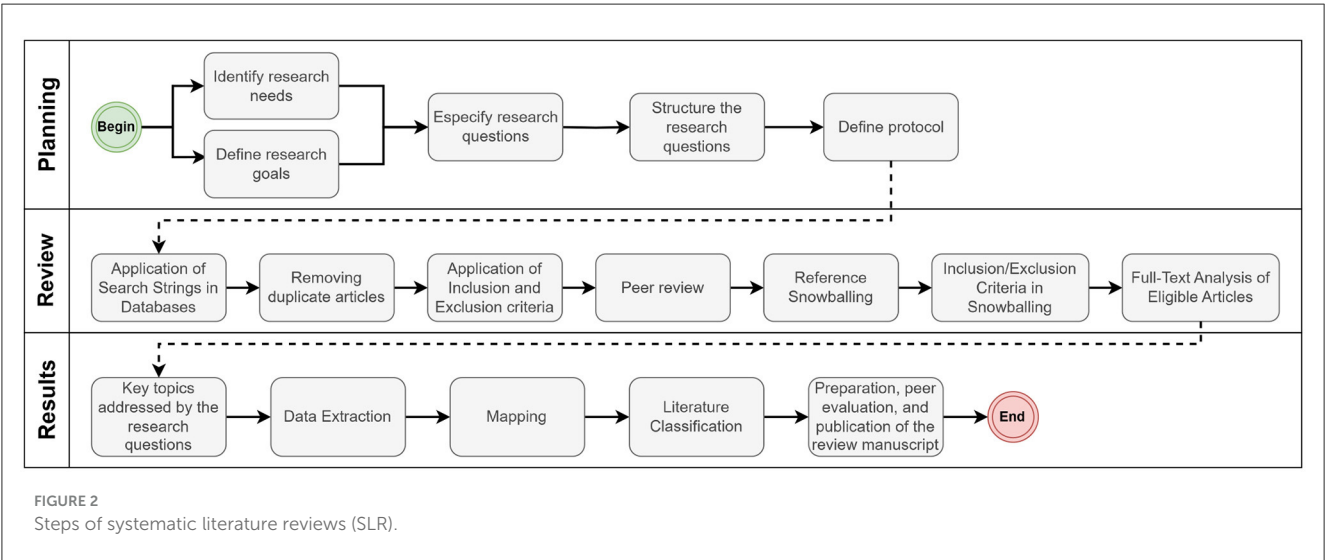
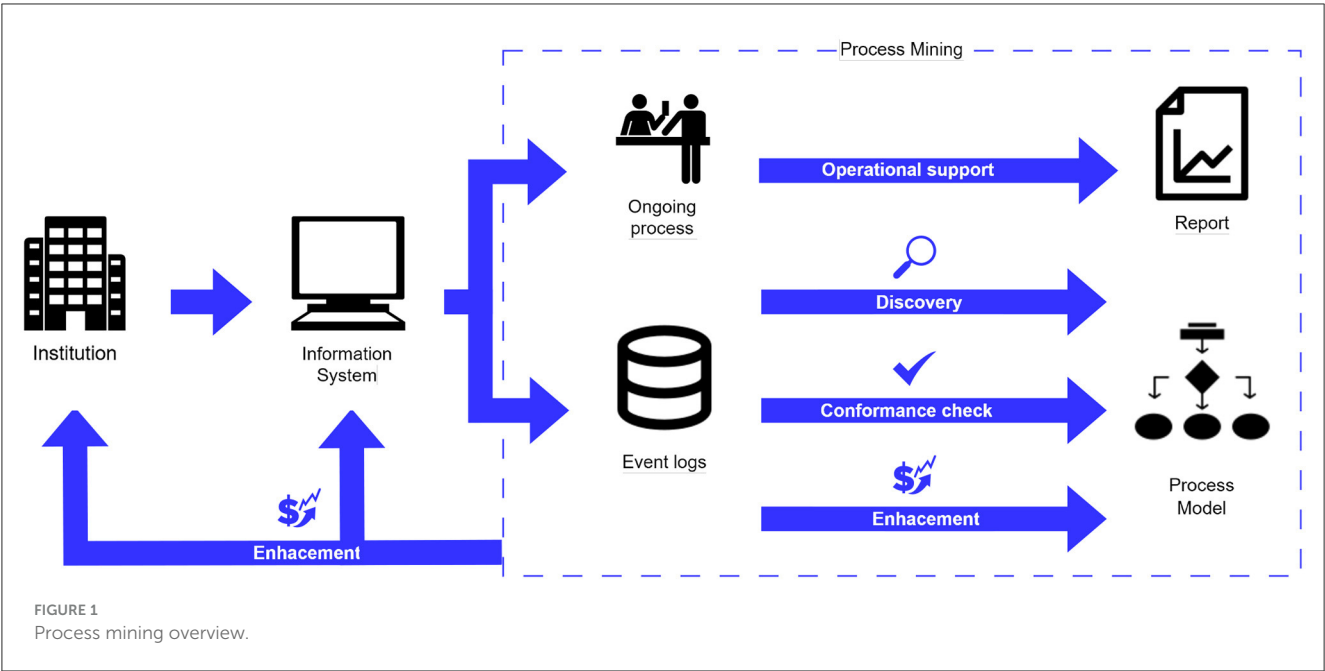
2.1 Educational process mining

EPM can be seen as a subarea of PM dedicated to understanding and improving educational processes. Models discovered through EPM have a wide range of applications that contribute to a deeper understanding of educational processes. By leveraging data from learning environments, it is possible to discover, analyze, and visually represent how these processes unfold (Bogarín et al., 2018a). Such models support the detection of learning styles (Wang et al., 2019), enabling early identification of learning difficulties, personalized recommendations, and targeted support for students with specific needs. They also provide valuable feedback to learners, educators, and researchers (Domínguez et al., 2021; Hachicha et al., 2021; Liu et al., 2022). Furthermore, EPM facilitates the analysis and categorization of student actions in relation to specific learning activities (Juhaňák et al., 2019), and supports conformance checking to determine whether observed behaviors align with predefined models (Salazar-Fernandez et al., 2021).

In summary, the use of support provided by EPM makes it possible to extract useful information and achieve a better understanding of learning, generating recommendations, and promoting continuous improvements in teaching (dos Santos Neto et al., 2022).

3 Literature review methodology

The methodology adopted in this study, so-called SLR, is described and validated by Petersen et al. (2015) and carried out with the essential characteristics, principles, and phases recommended. It consists of a literature search to identify a pool, i.e., a wide range of research articles that contain studies pertinent



to the outlined review question. This is achieved through an unbiased search strategy. The search approach was segmented as follows in [Figure 2](#).

3.1 Specify and structure questions, keywords, and search strings

The first step was to use the PICO (Population, Intervention, Comparison, and Outcomes) method suggested by [Kitchenham \(2012\)](#) and implemented in the adopted methodology for this study ([Petersen et al., 2015](#)) to identify keywords and form search strings based on the research questions. Population: For the area of EPM, population refers to a specific method, software category, or application area. For this study, the population refers to educational process mining. Intervention: In process mining, intervention

refers to a method, algorithm, tool, technology, or procedure. In the context of this study, techniques or algorithms applied to the educational domain are investigated. Comparison: In this study, different applications of process mining in the educational domain are compared by identifying different strategies and uses. Outcomes: For this study, the main results obtained in each article were observed, such as the process mined, its subsequent use in conformance analysis or prediction, as well as future notes for the area of EPM.

3.2 Questions to guide the review

The method starts by defining 5 questions whose answers could allow knowing what are authors and research groups of EPM studying in EPM (tasks, datasets, process workflows, methods,

techniques, algorithms, tools, and approaches), what are the results and their impact, and, mainly what are the future challenges in EPM.

- (i) What datasets and educational processes are explored? What PM tasks were applied?
- (ii) What methods, techniques, algorithms, tools, and approaches are used in each EPM step?
- (iii) What are the main results obtained? How have they been validated?
- (iv) What are the main directions and future research trends?

By collecting and analyzing the answers to the research questions, it is possible to map and categorize their predominant contributions, origins, publishing profiles, tools, techniques, and algorithms on EPM.

3.3 Inclusion and exclusion criteria

Inclusion and exclusion criteria serve to filter the papers that are really to be considered. For this paper, the criteria considered were:

- (i) studies on EPM that present techniques or algorithms applied to the educational domain; and
- (ii) studies that present a research method and results from the type of process mining analysis in the educational domain.

As a complement, the following exclusion criteria were considered:

- (i) non-peer-reviewed studies;
- (ii) studies not written in English; and
- (iii) books, chapters, editorials, short papers, collections, and technical reports.

In conjunction, inclusion and exclusion criteria allow for the composition of a robust, feasible, and yet complete set of material to be reviewed.

3.4 Conducting the review

The identified keywords were grouped to formulate the search string. Given that the keywords present in the research questions resemble the keywords identified in the PICO criteria, they were associated with corresponding sets. Each set of searches was performed on the IEEE Xplore, ACM, Springer, and Science Direct databases. These databases are used because, in engineering and computer science, these are the bases considered to have the best reputation and include the majority of the work on the topic when compared with other databases (Gusenbauer, 2022).

Considering that the search strings used for the IEEE Xplore, ACM, Springer, and Science Direct databases were the same, they were grouped in Table 1.

The EndNote Web reference management software was used to eliminate duplicates and manage a large number of references. The period considered in this study is from January 2018 (as the last SLR publication on EPM was in 2017) to April 2025. Table 2 presents the number of articles per database.

TABLE 1 Database search expressions.

Database	Search string
IEEE Xplore and ACM	("All Metadata";"educational process mining") OR
	("All Metadata";"educational workflow mining") OR
	("All Metadata";"educational mining workflow") OR
	("All Metadata";"educational data") OR
	("All Metadata";"educational mining") OR
	("All Metadata";"educational analysis") OR
	("All Metadata";"educational flow mining") OR
	("All Metadata";"educational mining flow")
Springer and Science Direct	"educational process mining" OR
	"educational workflow mining" OR
	"educational mining workflow" OR
	"educational data" OR "educational mining" OR
	"educational analysis" OR
	"educational flow mining" OR
	"educational mining flow"

TABLE 2 Database search results.

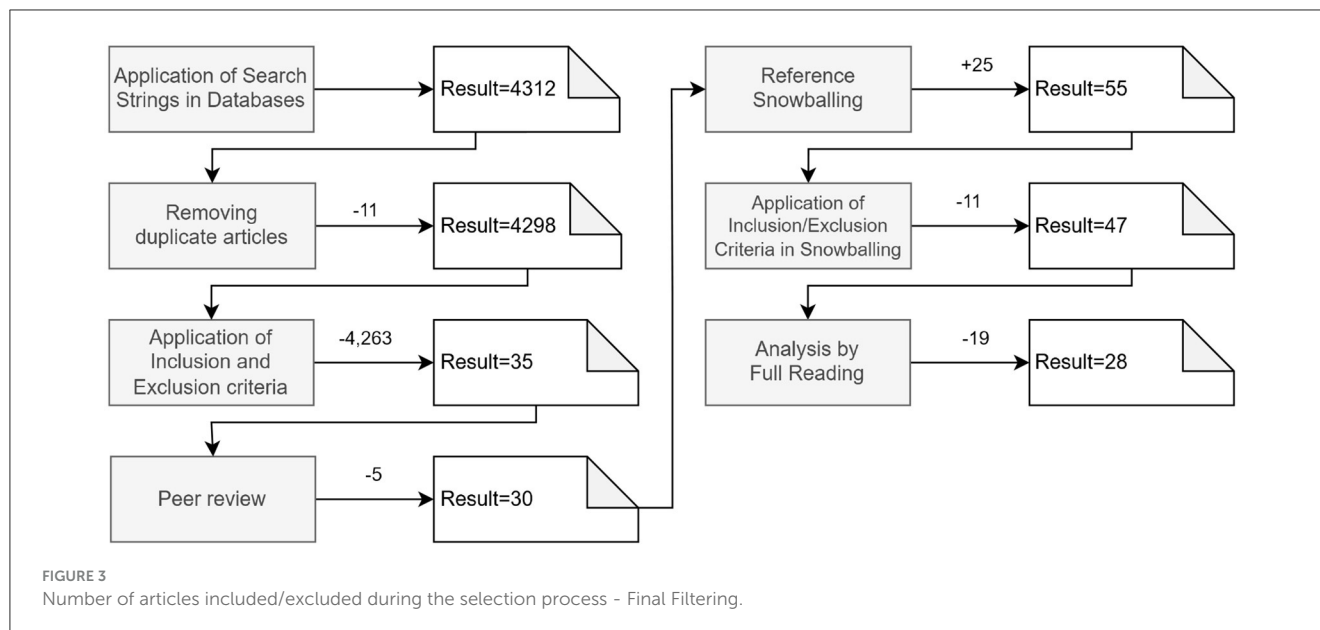
Database	Search results
IEEE Xplore	872
ACM	650
Springer	1,695
Science Direct	1,095
Total	4.312

Subsequently, in an initial screening, duplicated articles or articles that did not meet the criteria related to the defined search period were excluded from the bibliographic portfolio. As a second screening, each article's titles, abstracts, and keywords were reviewed by applying the inclusion and exclusion criteria described in Section 3.3.

To respect impartiality, this initial portfolio was subjected to peer review, characterized by another researcher who checked duplicates and short papers, in addition to reading the titles, abstracts, and keywords, checking for undue exclusions.

In the end, 28 articles made up the bibliographic portfolio and were integrally analyzed. These articles were carefully interpreted, observing the applications of process mining techniques and algorithms in the educational domain, with those that did not present such characteristics being excluded.

Following the guidelines of Wohlin (2014), the snowballing technique was applied to the bibliographic portfolio. Titles and abstracts from the references of the 28 selected articles were screened using predefined keywords, followed by the application of inclusion and exclusion criteria and peer review. After full-text analysis, 44 articles employing process mining in the educational context were identified. The analysis of the final portfolio ($n = 28$)



was guided by the research questions presented in Subsection 3.2 (Figure 3).

3.5 Data extractions

To extract the data, a table was created containing three columns: the variable names (item), the source of information extracted from the portfolio article (value), and its relevance to specific research questions (RQ). The items extracted with their respective values were:

- (i) Article ID: integer value identifying the article;
- (ii) Article Title: title description;
- (iii) Name of Author(s): listing the authors;
- (iv) Year of Publication: year;
- (v) Periodic: title of the periodic;
- (vi) Keywords: keywords of the article;
- (vii) EPM aspects explored: description of the main subject;
- (viii) Datasets: description of where the datasets came from;
- (ix) Methods, Techniques, Algorithms: description of methods, techniques, algorithms, and tools used;
- (x) Results: main results obtained;
- (xi) Validation: description of how the result was validated; and
- (xii) Future Research: description and indication for future research.

After extracting each features, they were systematically compared to reveal overarching patterns. First, articles were grouped according to the PM tasks addressed (Discovery, Conformance Checking, Enhancement, and Prediction), then cross-referenced by the type of dataset [Moodle, online platforms, educational information systems, integrated development environments (IDEs)] and the main educational workflow (learning patterns, instructional processes, problem-solving, etc.). This comparative approach uncovered shared trends—such as the strong reliance on discovery tasks and the prevalence of

Moodle logs—and highlighted where methodological gaps persist (e.g., limited validation or sparse use of predictive analytics). By mapping individual findings onto these broader categories, the review moves beyond a simple listing of studies, yielding a more integrative perspective on the current state of EPM.

The other authors oversaw both the extraction process and reviews to ensure consistency.

4 Results of the mapping

Figure 4 presents a heatmap summarizing the distribution of EPM publications by country and year from 2018 to April 2025. Peaks in research activity occurred in 2019 and 2024 (six publications each), while 2023 recorded the lowest output. These variations may reflect shifts in research priorities or external influences.

The most active countries include China, Spain, Brazil, Italy, Chile, Czech Republic, and Ecuador, with China showing consistent contributions across multiple years. Institutional highlights include Universidad de La Rioja (Spain) and Guizhou University (China), marking a novel granularity not addressed in prior reviews (Ghazal et al., 2017; Bogarín et al., 2018b; Sypsas and Kalles, 2022).

Most studies were published in journals, with only three appearing in conference proceedings (Real et al., 2020; Ardimento et al., 2019; Puttow Southier et al., 2024). Notably, *IEEE Transactions on Learning Technologies and Computers in Human Behavior* accounted for 24% of the portfolio, reinforcing their relevance for future research dissemination.

Unlike previous reviews, this study introduces a bibliometric perspective, offering original insights into publication trends by year, country, institution, and journal, thereby guiding future research navigation within the EPM field.

In comparison with existing literature reviews on EPM (Sypsas and Kalles, 2022; Bogarín et al., 2018b; Ghazal et al., 2017), none provided quantitative bibliometric data regarding the selected

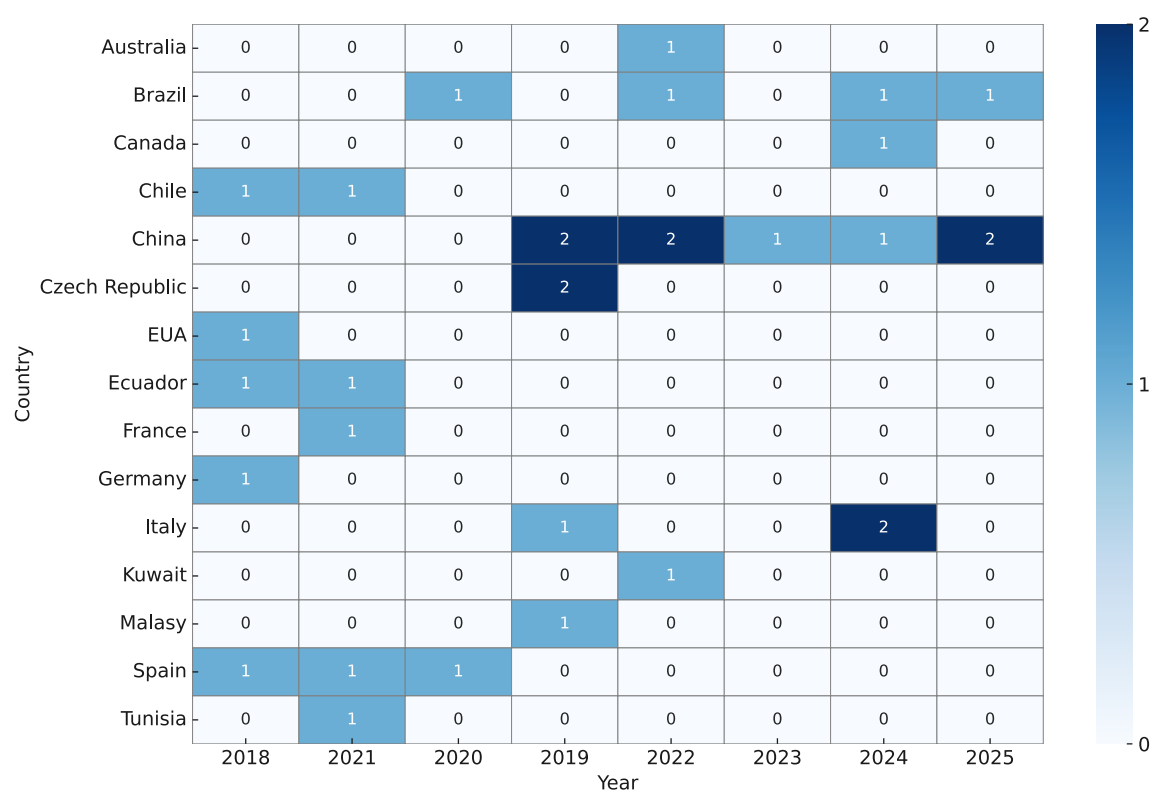


FIGURE 4
Heatmap of EPM publications by country and year (2018–2025).

articles. This study, therefore, introduces a novel contribution by incorporating this type of analysis.

5 Results: answers and findings

Based on the research questions in Section 3.2, the following subsections provide the corresponding answers and discuss the main findings. Additionally, Section 5.1 introduces the proposed literature classification, which is detailed in Section 5.1.1.

5.1 RQ2—What datasets and educational processes are explored? What PM tasks were applied?

Table 3 classifies the reviewed works according to PM tasks, the systems from which event logs were obtained, and the educational workflows analyzed.

Event log sources are primarily Moodle-based datasets (40%), with notable use of other online LMS platforms (28%), educational information systems (28%), and, less frequently, Integrated Development Environments (IDE) logs (4%). This distribution underscores the predominant focus on teaching and learning processes within Educational Process Mining (EPM), influenced by the accelerated adoption of e-learning solutions during the COVID-19 pandemic.

Despite Moodle datasets dominating the literature, featured in approximately 40% of studies, detailed discussions on their inherent structure, nature, and limitations remain scarce. Moodle logs generally contain timestamps, user identification, accessed resources, activity submissions, and click-stream data, providing rich but loosely structured event information. However, they often omit crucial metadata like user roles, assessment outcomes, and instructional design contexts, which restricts their application in conformance checking and enhancement tasks. The variability and granularity of data across platforms significantly impact the selection of mining methodologies, accuracy of derived models, and the generalizability of findings.

Thus, future research should incorporate comparative analyses of Learning Management Systems (LMS) datasets, emphasizing the effects of data quality, completeness, and semantic consistency on process model discovery and interpretation. Developing a comprehensive taxonomy of educational datasets—including their sources, event typologies, temporal resolutions, and privacy issues—would offer clearer guidance for researchers in dataset selection and preprocessing.

Analyzed educational workflows align closely with the utilized datasets, focusing predominantly on learning patterns (52%), followed by instructional processes (24%), learner behaviors (24%), and, to a lesser extent, problem-solving and learning objects (4% each).

Regarding mining tasks, process discovery is the most prevalent, employed in 80% of the studies. This task aims to

TABLE 3 Classification of works per type of EPM task.

PM Task	Log sources	Workflow	References
Conformance	IDE	Behavior	Ardimento et al., 2019
	Educational IS	Learning pattern	Wang et al., 2019 ; Porouhan and Premchaiswadi, 2017
		Instructional process	Diamantini et al., 2024 ; Puttow Southier et al., 2024
		Problem-solving	Liu et al., 2022
	LMS	Learning pattern	Van den beemt et al., 2018 ; Martínez-Carrascal et al., 2024 ; He et al., 2024 ; Zhang et al., 2024
Enhancement	LMS	Instructional process	Domínguez et al., 2021
Discovery	Educational IS	Learning pattern	Feng et al., 2022 ; AlQaheri and Panda, 2022
		Instructional process	Salazar-Fernandez et al., 2021 ; Diamantini et al., 2024
	LMS	Behavior	Juhaňák et al., 2019 ; Dolak, 2019 ; He et al., 2024 ; Puttow Southier et al., 2024 ; Xu et al., 2025 ; Real and Pimentel, 2025 ; Ma, 2025
		Learning pattern	Bogarín et al., 2018a ; Real et al., 2020 ; Bakar, 2019 ; Maldonado-Mahauad et al., 2018 ; Van den beemt et al., 2018 ; Zhang et al., 2024 ; Martínez-Carrascal et al., 2024 ; He et al., 2024
		Instructional process	Hachicha et al., 2021 ; Saint et al., 2021
		Learning objects	dos Santos Neto et al., 2022
		Problem-solving	Liu et al., 2022 ; Burhan Ogut and Yin, 2024
Prediction	Educational IS	Learning pattern	Feng et al., 2022
	LMS	Behavior	Nai et al., 2024

build accurate process models, helping stakeholders understand process execution and common trajectories. However, given the complexity inherent in educational environments, characterized by diverse technologies, fragmented systems, and data access barriers, research frequently prioritizes behavioral and learning path analysis over educational management concerns.

Conformance checking is employed by 36% of the reviewed articles, often in conjunction with discovery techniques. This task involves comparing process models against actual event logs to identify deviations and commonalities, which are crucial for auditing and model refinement. Examples include analyzing problem-solving abilities in programming education and evaluating student trajectories against predefined models.

Process enhancement, addressed by only 4% of studies, examines aspects such as temporal dynamics of student interactions. Prediction remains notably underexplored, appearing in merely two studies (8%), despite its potential for forecasting delays, recommending timely interventions, and anticipating learning outcomes.

In summary, the current EPM literature strongly emphasizes process discovery and model conformance tasks, highlighting an evident gap and significant opportunity in predictive approaches, which could notably advance educational management effectiveness and responsiveness.

5.1.1 Literature classification of EPM

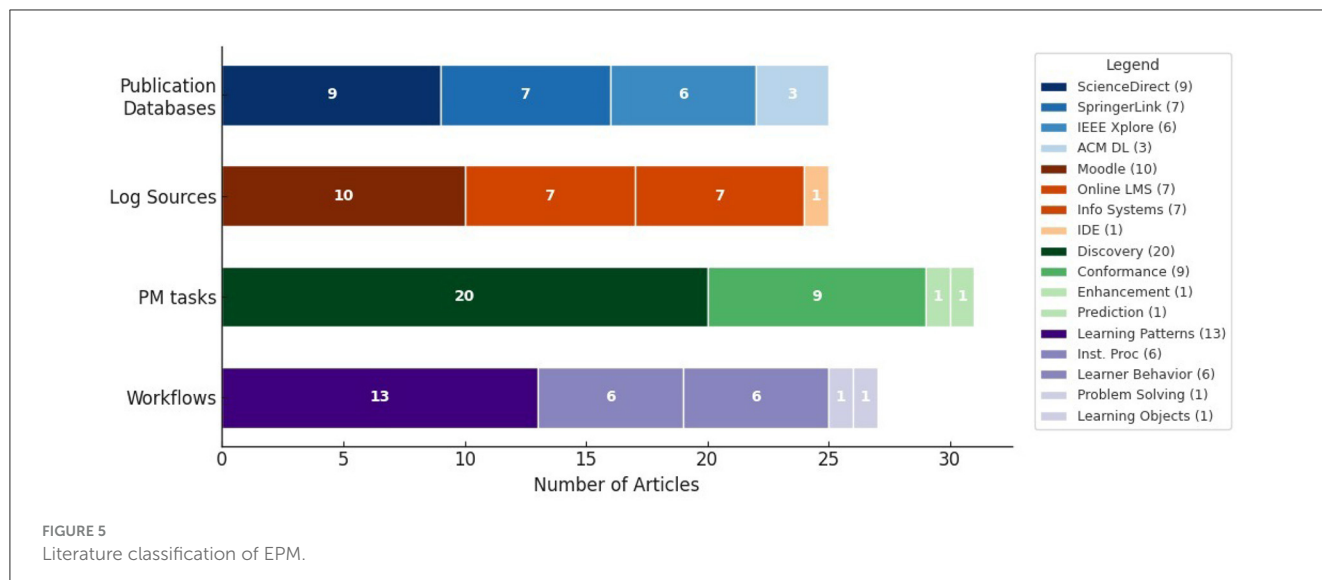
This study proposes a structured classification framework to organize the selected literature based on the core PM dimensions applied in educational contexts—namely discovery, conformance checking, enhancement, and prediction—alongside their associated

log sources and pedagogical workflows. A visual summary is presented in [Figure 5](#).

Although the final portfolio consists of 28 studies, [Figure 5](#) includes 33 classified entries, as several papers contribute to multiple PM perspectives. Discovery stands out as the dominant task, accounting for 62.5% of all occurrences (20 out of 32), and is frequently employed to extract behavioral insights ([Dolak, 2019](#); [Juhaňák et al., 2019](#); [He et al., 2024](#); [Ma, 2025](#); [Real and Pimentel, 2025](#); [Xu et al., 2025](#)), uncover learning patterns ([Maldonado-Mahauad et al., 2018](#); [Bogarín et al., 2018a](#); [Van den beemt et al., 2018](#); [Real et al., 2020](#); [Bakar, 2019](#); [Martínez-Carrascal et al., 2024](#); [He et al., 2024](#); [Zhang et al., 2024](#)), map instructional strategies ([Hachicha et al., 2021](#); [Wang et al., 2019](#); [Salazar-Fernandez et al., 2021](#); [Diamantini et al., 2024](#)), and model problem-solving ability ([Liu et al., 2022](#); [Martínez-Carrascal et al., 2024](#)).

Conformance checking is the second most cited task (28%), appearing in studies that focus on validating learning path fidelity ([Wang et al., 2019](#); [Dolak, 2019](#); [Van den beemt et al., 2018](#); [Martínez-Carrascal et al., 2024](#); [He et al., 2024](#); [Zhang et al., 2024](#)), identifying behavioral deviations ([Ardimento et al., 2019](#)), and evaluating instructional adherence ([Diamantini et al., 2024](#); [Puttow Southier et al., 2024](#)). Enhancement and prediction are notably underrepresented, with only one occurrence each (3%). Enhancement was explored in [Domínguez et al. \(2021\)](#), where process data was used to refine instructional strategies, while prediction appeared in [Feng et al. \(2022\)](#); [Nai et al. \(2024\)](#), aiming to anticipate student performance based on behavior logs.

Regarding log sources, *Learning Management Systems* (LMS) dominate as the primary source of event logs (22 out of 28), particularly Moodle and similar platforms. These systems provide comprehensive digital footprints, enabling robust discovery and



validation workflows. Educational Information Systems account for 8 entries, serving as sources for both behavioral and administrative process analysis (Feng et al., 2022; Wang et al., 2019; Dolak, 2019; Salazar-Fernandez et al., 2021; AlQaheri and Panda, 2022; Diamantini et al., 2024; Puttow Southier et al., 2024), while IDEs appear only once, applied in the context of conformance checking to model code writing behavior (Ardimento et al., 2019).

The categorization of pedagogical workflows reveals a strong focus on learning patterns (12 entries), followed by instructional processes (6), learner behavior (9), problem-solving ability (2), and learning objects (1). This distribution suggests that much of the current EPM research is oriented toward reconstructing and analyzing student trajectories, often leveraging process discovery and conformance validation. Studies such as Real et al. (2020); Bakar (2019); Zhang et al. (2024) utilize behavioral logs to map frequent learning routes, while others like (Hachicha et al., 2021; Wang et al., 2019) focus on instructional execution quality.

This classification not only highlights research concentration around discovery and LMS-based learning paths but also reveals opportunities for advancement. The sparsity of studies in enhancement and prediction, despite their strategic relevance for proactive feedback and adaptive learning, points to meaningful gaps. Furthermore, the limited exploration of alternative workflows, such as problem-solving and interaction with learning objects, suggests the need for more diverse analytical perspectives in EPM.

5.2 RQ3—What are the methods, techniques, algorithms, tools, and approaches used in each step of the EPM?

Based on the reviewed studies, the EPM techniques applied can be grouped into the following analysis categories:

- Case studies where the analysis relies on existing PM techniques available in current tools, without developing new

methods or algorithms. These may also include the use of well-established techniques from other research domains.

- Case studies where authors not only used existing PM techniques within the available tools but also incorporated techniques from other areas of investigation to improve their analysis (Feng et al., 2022; Wang et al., 2019; AlQaheri and Panda, 2022).

Most studies in the bibliographic portfolio used PM techniques and algorithms already implemented, functional, and available in tools like *ProM*; only two articles use BPMN available in the tool *Disco* (dos Santos Neto et al., 2022; Dolak, 2019). The most used algorithm was the Inductive Miner (22.22%), followed by the Heuristic Miner (17%) and Fuzzy Miner (11.11%).

Among the studies in the portfolio, only two (Domínguez et al., 2021; Maldonado-Mahauad et al., 2018) explicitly describe the methodology adopted for conducting EPM. In both cases, the chosen approach was PM². So far, three prominent PM methodologies have been identified: the Process Diagnostic Method (PDM) (Porouhan and Premchaiswadi, 2017), the L* life cycle model (Van Der Aalst, 2016), and the PM² Methodology (Poncin et al., 2011).

For articles that do not make it explicit, two methodologies were identified. The first is based on the PDM, which comprises five quick phases: log preparation, log inspection, control flow analysis, performance analysis, and function analysis. The studies that used this were Juhaňák et al. (2019); Dolak (2019); Bogarín et al. (2018a); Van den beemt et al. (2018); Salazar-Fernandez et al. (2021); Real et al. (2020); Ardimento et al. (2019); Bakar (2019).

Several studies have expanded the PDM approach by introducing a preprocessing and log clustering step before control flow analysis. Notably, Hachicha et al. (2021); Liu et al. (2022); Saint et al. (2021) have all adhered to this extended methodology.

The second methodology observed in the studies is characterized by *ad-hoc* methods, steps, or guidelines developed and employed by the authors to apply EPM (Feng et al., 2022; Wang et al., 2019; AlQaheri and Panda, 2022; Xu et al., 2023).

It is evident that there is currently no contemporary, domain-specific methodology adapted for EPM or that is self-adaptive. The development of such a domain-specific methodology is essential to establish EPM as a standardized and replicable process in new case studies.

5.3 RQ4—What are the main results obtained? How have they been validated?

The articles in the portfolio report a variety of outcomes, including proposed architectures and models, as well as findings related to discovery, conformance, and enhancement in EPM. Some articles focus on process discovery and the analysis of learning patterns, instructional approaches, or behavioral models, with results pointing to several possibilities, such as:

- (i) Identifying interaction sequences: students' interaction patterns when accessing course materials were identified, providing insights into their behavior and engagement (dos Santos Neto et al., 2022; Bakar, 2019).
- (ii) Differentiating interaction patterns: patterns of interaction sequences, involvement, and behavior were distinguished, contributing to a deeper understanding of student dynamics (Maldonado-Mahauad et al., 2018; Domínguez et al., 2021; Juhaňák et al., 2019; Dolak, 2019; Van den beemt et al., 2018; Xu et al., 2023).
- (iii) Profiling students: distinct student profiles were identified based on their interactions and behaviors (Maldonado-Mahauad et al., 2018; Juhaňák et al., 2019).
- (iv) Improving instructional design: the application of generalizable problem-solving models enabled dynamic adjustments to learning paths, enhancing students' skills throughout the course (Liu et al., 2022; Feng et al., 2022; Bogarín et al., 2018a; Salazar-Fernandez et al., 2021; Wang et al., 2019).
- (v) Reducing dropout risks: process models integrated into recommendation systems effectively reduced the risk of dropout, particularly among students with a higher likelihood of failure (Hachicha et al., 2021; AlQaheri and Panda, 2022).
- (vi) Supporting teacher analysis: although the actual behavior of students tend to be more complex than models can capture, the discovered process still offers valuable support for teachers in assessing the effectiveness of their instructional methods (dos Santos Neto et al., 2022; Saint et al., 2021; Real et al., 2020).

In Ardimento et al. (2019), the authors analyzed event logs generated during developers' interaction with an IDE to understand individual coding behaviors and common challenges. Through conformance analysis, they identified behavioral patterns and variations, revealing that IDE usage differs notably based on developers' skills and performance.

Regarding validation approaches, the portfolio suggests the following:

- (i) Lack of clarity (33.33%): a significant portion of the studies do not clearly describe whether or how their results were validated (Hachicha et al., 2021; Feng et al., 2022; Saint et al.,

2021; Juhaňák et al., 2019; Bogarín et al., 2018a; Bakar, 2019). In most of these cases, validation is either absent or mentioned only vaguely, without specifying reproducible procedures, metrics, or comparison baselines.

- (ii) Re-experiments (33.33%): some studies conducted new experiments using the same dataset to validate their findings (Liu et al., 2022; Van den beemt et al., 2018; Salazar-Fernandez et al., 2021; Ardimento et al., 2019; Real et al., 2020).
- (iii) Literature-based validation (27.78%): other articles relied on frameworks or prior studies to support their results (Wang et al., 2019; dos Santos Neto et al., 2022; Domínguez et al., 2021; Maldonado-Mahauad et al., 2018; AlQaheri and Panda, 2022).
- (iv) Descriptive statistics (5.56%): a smaller fraction used descriptive statistical methods for validation (Dolak, 2019).

5.3.1 Case-based synthesis of EPM applications

To deepen the understanding of the practical applications of EPM, a targeted synthesis of five representative case studies from the reviewed literature was conducted, as shown in Table 4. These cases illustrate the diversity of datasets, methods, and research objectives, and their comparative analysis reveals recurring methodological choices, common limitations, and potential directions for improvement.

A key observation is the predominance of exploratory and descriptive approaches. Most studies focus on process discovery using default configurations of tools like ProM or Disco. While this provides initial insights into educational behaviors, few efforts go beyond surface-level analysis or attempt to triangulate findings through statistical, experimental, or expert validation.

Another recurring pattern is the lack of methodological justification for tool or algorithm selection. In several cases, the choice of mining technique appears to be driven by tool availability rather than alignment with the structure or semantics of the dataset. Moreover, only a minority of studies incorporate conformance checking or enhancement phases, and predictive approaches remain rare.

These findings suggest that while EPM research has made meaningful progress in uncovering learning dynamics, it still requires greater methodological rigor, cross-study synthesis, and application-driven modeling. This reinforces the importance of transparent methodological reporting and tailored analytical strategies. Moving forward, the field would benefit from adopting mixed-methods approaches that combine process mining with domain knowledge, predictive modeling, and qualitative validation. Doing so would elevate EPM from a descriptive tool to a more powerful instrument for educational diagnosis, adaptation, and transformation.

5.4 RQ5—What are the main directions pointed as future research trends?

Suggestions for future research can be organized into thematic cores. The first core is related to PM itself and involves

TABLE 4 Representative case studies in educational process mining.

Study	Dataset source	Method/tool used	Focus area	Key results	Limitations
Liu et al. (2022)	Online Platform Logs	Inductive Miner/ProM	Problem-Solving Skills	Mapped task sequences; conformance with ideal model	No statistical validation; single-course scope
dos Santos Neto et al. (2022)	Moodle Logs	BPMN Models/Disco	Flipped Classroom Detection	Detected process alignment with class format	Event-level only; no deeper behavioral insights
Maldonado-Mahauad et al. (2018)	MOOC Platform Logs	Inductive Miner + SRL Framework	Self-Regulated Learning Strategies	Identified learner profiles based on SRL theory	Limited generalizability; interpretability challenges
Hachicha et al. (2021)	Moodle Logs	Heuristic Miner + Recommender	Resource Recommendation	Suggested learning resources based on process deviations	No predictive modeling; lacks impact evaluation
Zhang et al. (2024)	Online LMS Logs	Log Skeleton + Prediction Model	Personalized Learning Paths	Matched learner traces to optimal trajectories	Prediction module lacks interpretability

research directions focused on event logs, improvements to mining techniques, and the development of PM algorithms, as outlined below:

- (i) **Events Logs:** grouping of event logs from educational systems added to a resource that can extract and simulate student behavior ([Hachicha et al., 2021](#); [Wang et al., 2019](#); [Juhaňák et al., 2019](#); [Van den beemt et al., 2018](#); [Salazar-Fernandez et al., 2021](#));
- (ii) **PM Algorithms:** generalization of process mining algorithms for application in different contexts ([Bogarín et al., 2018a](#)); It is
- (iii) **PM miners:** improvements to heuristic miners with compliance checking ([AlQaheri and Panda, 2022](#))

The second core identifies new trends in assessment, teaching strategies, skills, and behaviors:

- (i) **Evaluation:** introduction of recommendation oriented to process evaluation ([Hachicha et al., 2021](#));
- (ii) **Ability:** an in-depth analysis of the influence of event records on process decisions ([Liu et al., 2022](#); [Ardimento et al., 2019](#); [Xu et al., 2023](#)); and the mining of paths that enhance the chances for students to obtain their academic level ([Feng et al., 2022](#));
- (iii) **Teaching:** development of more effective, interactive and modern teaching strategies ([Liu et al., 2022](#); [Feng et al., 2022](#); [Saint et al., 2021](#); [Real et al., 2020](#); [Bakar, 2019](#));
- (iv) **Behavior:** association of time-related characteristics with learning behavior ([Feng et al., 2022](#); [Saint et al., 2021](#));

From all articles, only 4 ([dos Santos Neto et al., 2022](#); [Domínguez et al., 2021](#); [Maldonado-Mahauad et al., 2018](#); [Dolak, 2019](#)) did not recommend future research.

Although some solutions have been implemented in practice, a key challenge remains: these investigations were mostly conducted in isolation and in such a way that they depend on the insights and expertise of the individual involved. In addition, efforts were focused on issues related to learning, improving grades, teaching plans, course programs, and/or instructional design.

Up to now, there is no standard methodology or framework for conducting new studies, making it challenging to establish EPM as a replicable and consistent technique. Furthermore, event

logs in the educational context can be obtained in the case of e-learning systems such as Moodle and, in some cases, in educational systems related to grades and attendance. However, data related to educational management, coupled with its heterogeneity and complexity, and the challenges in structuring event logs within the management context, have hindered the effective application of process mining by researchers in the fields of administration and management.

In terms of future research, the previous EPM reviews discussed the improvement of the model discovery process, detection of improvements in learning processes, dissemination of results from the application of EPM context in order to reach everyone involved and provision of EPM tools and data free of charge ([Sypsas and Kalles, 2022](#)); implementation of portable solutions focusing on the development of a Process-Aware Educational Information System ([Ghazal et al., 2017](#)); the use of predictions and recommendations in EPM, application of EPM in emerging educational domains such as games and making EPM datasets available in the public domain ([Bogarín et al., 2018b](#)). They are, therefore, aligned with the issues raised by this study. Therefore, future efforts are suggested to focus mainly on the following aspects:

- (i) **Academia:** use of EPM for students to recognize their academic levels and improve their performance; adjustments to teaching and course programs; use of the time variable for performance analysis in the process model and with dotted graph analysis techniques to visualize the propagation of events over time; early detection of groups of students based on their learning behavior and personal goals; and architectural projection for analyzing learning paths to improve planning or redesign subjects and courses.
- (ii) **Administrative Management of Educational environments:** process discovery, modeling, conformance analysis, enhancement, and prediction.

Furthermore, future researches are expected to prioritize the implementation of versatile and self-adaptive solutions that enable the construction of data engineering approaches for collecting refined event logs, as well as self-adaptive data science approaches that can be applied to education, enabling a more direct and precise application of PM techniques.

5.4.1 Emerging trends: artificial intelligence and machine learning in EPM

Despite the increasing complexity of educational environments and the growing volume of generated data, the integration of advanced technologies such as Artificial Intelligence (AI) and Machine Learning (ML) in EPM remains limited. Most of the reviewed studies rely on traditional PM algorithms (e.g., Inductive Miner, Heuristic Miner) and predefined toolsets, without incorporating adaptive, data-driven models capable of learning patterns or predicting outcomes.

AI and ML techniques—such as neural networks, decision trees, deep learning, and reinforcement learning—offer significant potential to enhance EPM by enabling predictive analytics, anomaly detection, and real-time feedback loops. For instance, deep learning models could be trained on historical learning trajectories to predict student dropout risks or recommend personalized learning paths. Reinforcement learning could support automated adjustments in instructional strategies based on process deviations.

Some initial efforts have already demonstrated the feasibility of such integrations. [Feng et al. \(2022\)](#), for example, applied predictive models to learning behavior logs, while [AlQaheri and Panda \(2022\)](#) incorporated pattern recognition to enhance student profiling. However, these approaches stand as individual initiatives, often limited to proof-of-concept experiments, and do not converge efforts toward a broad and real-life impact.

Future research, therefore, should focus on the systematic integration of AI and ML into the EPM lifecycle. This includes the development of hybrid frameworks that combine PM with predictive modeling, the design of interpretable models that maintain transparency for educators, and the implementation of scalable architectures capable of processing large, multi-source educational datasets. Additionally, benchmarking studies that compare traditional mining techniques with AI-driven approaches will be essential to validate their added value and applicability in real educational contexts.

6 Discussion

While this SLR provides a structured overview of the EPM landscape, a critical reflection reveals that the field is still in a formative stage, with several methodological and conceptual limitations. A large proportion of studies rely on established tools—primarily ProM and Disco—and default to using pre-implemented algorithms such as the Inductive Miner or Heuristic Miner. Although these tools are robust and widely adopted, their impact is limited by the available functionalities only, leaving more general analysis uncovered, suggesting a pattern of tool-driven rather than question-driven analysis.

Moreover, only a few studies apply methodological rigor in validation. Roughly one-third of the portfolio offers no clear strategy for validating process models or assessing their reliability. When validation is present, it is often limited to reapplying the same method to a single dataset without statistical evaluation or external benchmarks. This practice restricts the generalizability and scientific robustness of the findings.

Another relevant gap is the lack of comparative analysis across studies. Despite the diversity in contexts, datasets, and

tools, most articles present isolated case studies without cross-case synthesis or discussion of methodological trade-offs. There is little effort to compare, for instance, the impact of using different mining algorithms on similar educational workflows or to evaluate the strengths and weaknesses of datasets sourced from different LMS platforms.

Additionally, while some works introduce innovative techniques—such as predictive models or custom visualizations—these are rarely contrasted with conventional approaches to demonstrate their added value. As a result, the evolution of the field remains fragmented, and the opportunity to build cumulative knowledge is hindered.

To advance the maturity of EPM as a research domain, future studies must move beyond tool replication and embrace critical methodological reflection. This includes aligning mining techniques with specific research questions, providing transparent validation protocols, and fostering comparative and cross-institutional studies. Only through such rigor will EPM evolve into a reliable foundation for data-informed educational transformation.

6.1 Real-world impact of EPM

A noteworthy aspect of the reviewed literature is the evidence that EPM can deliver tangible benefits in real educational settings. Several studies suggest that EPM has the potential to influence practice by informing adaptive learning strategies, refining instructional design, and reducing dropout risks. For instance, research by [dos Santos Neto et al. \(2022\)](#) and [Hachicha et al. \(2021\)](#) show that integrating process models into LMS can help identify critical deviations in student behavior and trigger timely interventions. These studies suggest that when PM outputs are effectively integrated into decision-making frameworks, educators are better equipped to adjust course content and teaching methods to meet learners' evolving needs.

Furthermore, some case studies have reported improved student profiling and enhanced feedback mechanisms, which contribute to more personalized learning experiences. By mapping student interactions and learning trajectories, EPM not only uncovers hidden patterns in behavior but also supports targeted adjustments in instructional design. Such findings underscore the promise of EPM as a tool for enhancing student outcomes and for guiding educational planning at both the classroom and institutional levels.

Despite encouraging results, it is important to note that the majority of the studies reviewed remain exploratory, with most evidence stemming from short-term or single-institution investigations. There is a notable absence of longitudinal studies that confirm the sustained, large-scale impact of EPM on student performance. This gap highlights the need for future research to extend beyond pilot implementations and establish robust, replicable frameworks that can capture the long-term benefits of EPM in diverse contexts.

In summary, while early successes indicate that EPM can positively impact teaching and learning practices, further applied research is required to validate these effects on a broader scale. Strengthening the empirical evidence through multi-institutional

studies and long-term evaluations will be crucial for positioning EPM as a reliable foundation for transforming education.

6.2 Limitations

While this SLR provides a comprehensive overview of EPM, several limitations must be recognized.

First, the review was restricted to peer-reviewed articles published in English and sourced from selected bibliographic databases (IEEE Xplore, ACM, Springer, and ScienceDirect). Although these sources cover a broad range of scholarly publications, this approach excludes gray literature and studies in other languages that might offer additional insights.

Second, the review period was confined to studies published between 2018 and 2024. While this timeframe was chosen to capture recent advancements in EPM, it may omit earlier foundational works that could provide valuable historical context or inform the field's evolution.

Third, the application of strict inclusion and exclusion criteria—focusing solely on peer-reviewed articles and excluding short papers, editorials, and conference abstracts—was necessary to maintain methodological rigor. However, this may have resulted in the omission of innovative or preliminary studies that could contribute to a broader understanding of emerging trends in EPM.

Additionally, the synthesis of the selected articles relies heavily on the quality and transparency of the reporting in the original studies. Inconsistent reporting across studies poses challenges for cross-study comparisons and may lead to an underestimation of certain methodological shortcomings, thus affecting the robustness of the conclusions.

Finally, the heterogeneity of research contexts—including variations in LMS, educational levels, and institutional settings—makes it challenging to develop universally applicable recommendations. This diversity limits the generalizability of the findings, as the observed trends may not be directly transferable to all educational environments.

By recognizing these limitations, this study lays the groundwork for future research applying addressing these challenges. In particular, more inclusive, cross-institutional, and longitudinal studies are needed to refine EPM methodologies and to ensure that subsequent research yields more generalizable and actionable insights.

7 Conclusion

This article presented a comprehensive overview of EPM by conducting a systematic literature review of 28 peer-reviewed studies published between 2018 and 2024. The analysis strongly emphasized process discovery tasks, with fewer studies addressing conformance checking and process enhancement. A predominance of datasets from platforms such as Moodle was observed, reflecting the widespread use of learning management systems in educational research.

A central contribution of this study lies in the proposed classification framework, which organizes the literature according to the dimensions of PM (discovery, conformance, and enhancement), types of educational workflows, and

methodological approaches. In addition, the review identified the most active authors, institutions, tools, and algorithms in the field. These findings map the current landscape and support the development of more rigorous, adaptable, and domain-specific methodologies for EPM.

Beyond methodological contributions, the review highlights the importance of ethical considerations in EPM. Detailed educational event logs raise critical concerns regarding data privacy, security, and regulatory compliance, particularly under frameworks such as the General Data Protection Regulation (GDPR) and the Brazilian General Data Protection Law (LGPD). Ensuring proper anonymization, informed consent, and responsible data governance is essential for the ethical application of PM in educational settings. Future studies should incorporate these ethical dimensions from the outset of research design.

While most reviewed studies are exploratory and focused on identifying behavioral patterns or instructional inefficiencies, a subset of articles has suggested a tangible impact on real-world educational practices. For example, PM has been used to support adaptive learning strategies, refine instructional design, and reduce dropout risks through targeted interventions. These early applications suggest that EPM has the potential to enhance student outcomes when appropriately integrated into teaching and decision-making processes. However, evidence of large-scale, longitudinal impact remains scarce, underscoring the need for more applied research that evaluates the effectiveness of EPM.

Finally, the study identifies promising directions for future research, including the use of EPM for real-time academic performance monitoring, predictive analytics, and the development of self-adaptive systems to support educational planning and institutional management. By offering a structured synthesis of the field and concrete pathways for advancement, this review contributes to the consolidation and maturation of EPM as a research domain with practical, ethical, and scientific relevance.

Author contributions

RS: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. JS: Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. MW: Supervision, Writing – original draft, Writing – review & editing. LS: Writing – original draft, Writing – review & editing. DC: Resources, Writing – review & editing. MT: Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. We received financial support from CNPq (grant 305069/2023-3), CAPES (PROEX), FINEP, Araucária Foundation, Graduate Programa on Electrical Engineering and Industrial Informatics CPGEI-CT/UTFPR, and DIRPPG-PB/UTFPR through Call 3/2024 - Support for Scientific Publications.

Acknowledgments

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) through the Academic Excellence Program (PROEX) and the National Council for Scientific and Technological Development (CNPq).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- AlQaheri, H., and Panda, M. (2022). An education process mining framework: unveiling meaningful information for understanding students; learning behavior and improving teaching quality. *Information* 13:29. doi: 10.3390/info13010029
- Ardimento, P., Bernardi, M. L., Cimitile, M., and Maggi, F. M. (2019). "Evaluating coding behavior in software development processes: a process mining approach," in *International Conference on Software and System Processes (ICSSP)*, pages 84–93. doi: 10.1109/ICSSP.2019.00020
- Bakar, M. (2019). A process mining approach to understand self regulated-learning in moodle environment. *Int. J. Adv. Trends Comput. Sci. Eng.* 8, 74–80. doi: 10.30534/ijatcse/2019/1581.32019
- Bogarín, A., Cerezo, R., and Romero, C. (2018a). Discovering learning processes using inductive miner: a case study with learning management systems (LMSS). *Psicothema* 30, 322–329. doi: 10.7334/psicothema2018.116
- Bogarín, A., Cerezo, R., and Romero, C. (2018b). A survey on educational process mining. *WIREs Data Min. Knowl. Disc.* 8:e1230. doi: 10.1002/widm.1230
- Burhan, O., Blue Webb, J. H. R. C., and Yin, M. (2024). Exploring mathematical problem-solving through process mining: Insights from large-scale assessment log data. *Comput. Schools* 2024, 1–31. doi: 10.1080/07380569.2024.2416422
- Diamantini, C., Genga, L., Mircoli, A., Potena, D., and Zannone, N. (2024). Understanding the stumbling blocks of italian higher education system: a process mining approach. *Expert Syst. Appl.* 242:122747. doi: 10.1016/j.eswa.2023.122747
- Dolak, R. (2019). "Using process mining techniques to discover student's activities, navigation paths, and behavior in lms moodle," in *Innovative Technologies and Learning*, eds. L. Rønningsbakk, T.-T. Wu, F. E. Sandnes, and Y.-M. Huang (Cham: Springer International Publishing), 129–138. doi: 10.1007/978-3-030-35343-8_14
- Domínguez, C., García-Izquierdo, F. J., Jaime, A., Pérez, B., Rubion, L., and Zapata, M. A. (2021). Using process mining to analyze time distribution of self-assessment and formative assessment exercises on an online learning tool. *IEEE Trans. Learn. Technol.* 14, 709–722. doi: 10.1109/TLT.2021.3119224
- dos Santos Garcia, C., Meincheim, A., Faria Junior, E. R., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., et al. (2019). Process mining techniques and applications – a systematic mapping study. *Expert Syst. Appl.* 133, 260–295. doi: 10.1016/j.eswa.2019.05.003
- dos Santos Neto, J. F., Marques Peres, S., Correia, P., and Fantinato, M. (2022). Is my classroom flipped? using process mining to avoid subjective perception. *ELearn* 2021:12. doi: 10.1145/3508017.3495212
- Dutt, A., Ismail, M. A., and Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access* 5, 15991–16005. doi: 10.1109/ACCESS.2017.2654247
- Feng, G., Fan, M., and Ao, C. (2022). Exploration and visualization of learning behavior patterns from the perspective of educational process mining. *IEEE Access* 10, 65271–65283. doi: 10.1109/ACCESS.2022.3184111
- Ghazal, M. A., Ibrahim, O., and Salama, M. A. (2017). "Educational process mining: a systematic literature review," in *2017 European Conference on Electrical Engineering and Computer Science (EECS)*, 198–203. doi: 10.1109/EECS.2017.45
- Gusenbauer, M. (2022). Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics* 127, 2683–2745. doi: 10.1007/s11192-022-04289-7
- Hachicha, W., Ghorbel, L., Champagnat, R., Zayani, C. A., and Amous, I. (2021). Using process mining for learning resource recommendation: a moodle case study. *Procedia Comput. Sci.* 192, 853–862. doi: 10.1016/j.procs.2021.08.088
- He, S., Demmans Epp, C., Chen, F., and Cui, Y. (2024). Examining change in students' self-regulated learning patterns after a formative assessment using process mining techniques. *Comput. Human Behav.* 152:108061. doi: 10.1016/j.chb.2023.108061
- Juhaňák, L., Zounek, J., and Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Comput. Human Behav.* 92, 496–506. doi: 10.1016/j.chb.2017.12.015
- Kaymakci, O., Anik, V. G., and Ustoglu, I. (2010). "A local modular supervisory controller for a real railway station," in *International Conference on System Safety*, 1–6. doi: 10.1049/cp.2010.0844
- Kitchenham, B. A. (2012). "Systematic review in software engineering: where we are and where we should be going," in *Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies, EAST '12* (New York, NY, USA: Association for Computing Machinery), 1–2. doi: 10.1145/2372233.2372235
- Liu, F., Zhao, L., Zhao, J., Dai, Q., Fan, C., and Shen, J. (2022). Educational process mining for discovering students' problem-solving ability in computer programming education. *IEEE Trans. Learn. Technol.* 15, 709–719. doi: 10.1109/TLT.2022.3216276
- Ma, F. (2025). Learning behavior analysis and personalized recommendation system of online education platform based on machine learning. *Comput. Educ.* 8:100408. doi: 10.1016/j.caeai.2025.100408
- Maldonado-Mahaud, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., and Munoz-Gama, J. (2018). Mining theory-based patterns from big data: identifying self-regulated learning strategies in massive open online courses. *Comput. Human Behav.* 80, 179–196. doi: 10.1016/j.chb.2017.11.011
- Martínez-Carrascal, J. A., Muñoz-Gama, J., and Sancho-Vinuesa, T. (2024). Evaluation of recommended learning paths using process mining and log skeletons: conceptualization and insight into an online mathematics course. *IEEE Trans. Learn. Technol.* 17, 555–568. doi: 10.1109/TLT.2023.3298035
- Nai, R., Sulis, E., and Genga, L. (2024). Enhancing e-learning effectiveness: a process mining approach for short-term tutorials. *J. Intell. Inf. Syst.* 62, 1773–1794. doi: 10.1007/s10844-024-00874-9
- Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: an update. *Inf. Softw. Technol.* 64, 1–18. doi: 10.1016/j.infsof.2015.03.007
- Poncin, W., Serebrenik, A., and Brand, M. (2011). "Mining student capstone projects with FRASR and prom," in *Macromolecular Symposia - MACROMOL SYMPOSIA*, 87–96. doi: 10.1145/2048147.2048181
- Porouhan, P., and Premchaiswadi, W. (2017). Process mining and learners' behavior analytics in a collaborative and web-based multi-tabletop environment. *Int. J. Pedagog. Course Des.* 7, 29–53. doi: 10.4018/IJOPCD.2017070103
- Puttow Southier, L. F., Teixeira, M., Casanova, D., and Scalabrin, E. E. (2024). "Towards a labeling method for education process mining and a case study on higher education," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* (New York, NY, USA: Association for Computing Machinery), 654–660. doi: 10.1145/3605098.3635968
- Real, E., and Pimentel, E. (2025). *An educational process mining model on students' paths data from virtual learning environments*. TechKnowLearn. doi: 10.1007/s10758-025-09824-y

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Real, E. M., Pinheiro Pimentel, E., de Oliveira, L. V., Cristina Braga, J., and Stiubiener, I. (2020). "Educational process mining for verifying student learning paths in an introductory programming course," in *IEEE Frontiers in Education Conference (FIE)*, 1–9. doi: 10.1109/FIE44824.2020.9274125
- Saint, J., Fan, Y., Singh, S., Gasevic, D., and Pardo, A. (2021). "Using process mining to analyse self-regulated learning: a systematic analysis of four algorithms," in *International Learning Analytics and Knowledge Conference, LAK21* (New York, NY, USA: Association for Computing Machinery), 333–343. doi: 10.1145/3448139.3448171
- Salazar-Fernandez, J. P., Munoz-Gama, J., Maldonado-Mahauad, J., Bustamante, D., and Sepúlveda, M. (2021). Backpack process model (BPPM): a process mining approach for curricular analytics. *Appl. Sci.* 11:4265. doi: 10.3390/app11094265
- Sypsas, A., and Kalles, D. (2022). Reviewing process mining applications and techniques in education. *Int. J. Artif. Intell. Applic.* 13, 83–102. doi: 10.5121/ijaia.2022.13106
- Van den beemt, A., Buys, J., and Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): an approach based on process mining and clustering. *Int. Rev. Res. Open Distr. Learn.* 19:3748. doi: 10.19173/irrodl.v19i5.3748
- Van Der Aalst, W. (2016). *Process Mining*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-662-49851-4
- Wang, Y., Li, T., Geng, C., and Wang, Y. (2019). Recognizing patterns of student's modeling behaviour patterns via process mining. *Smart Learn. Environ.* 6:26. doi: 10.1186/s40561-019-0097-y
- Wohlin, C. (2014). "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *International Conference on Evaluation and Assessment in Software Engineering, EASE '14* (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/2601248.2601268
- Xu, W., Lou, Y.-F., Chen, H., and Shen, Z.-Y. (2023). Exploring the interaction of cognition and emotion in small group collaborative discourse by heuristic mining algorithm (HMA) and inductive miner algorithm (IMA). *Educ. Inf. Technol.* 28, 1–26. doi: 10.1007/s10639-023-11722-8
- Xu, X., Zhao, W., Li, Y., Qiao, L., Tao, J., Liu, F. (2025). The impact of visualizations with learning paths on college students' online self-regulated learning. *Educ. Inf. Technol.* 30, 2917–2940. doi: 10.1007/s10639-024-12933-3
- Zhang, F., Feng, X., and Wang, Y. (2024). Personalized process-type learning path recommendation based on process mining and deep knowledge tracing. *Knowl. Based Syst.* 303:112431. doi: 10.1016/j.knosys.2024.112431