Moodle2EventLog: A Tool for Pedagogically-Driven Log Enrichment and Analysis

Noura Joudieh[1][a], Wil M. P. van der Aalst[2][b], Ronan Champagnat[1][c], Mourad Rabah[1][d] and Samuel Nowakowski[3][e]

[1]*L3i, La Rochelle University, La Rochelle, France*

[2]*PADS, RWTH Aachen University, Aachen, Germany*

[3]*LORIA, Lorraine University, Nancy, France*

{*noura.joudieh, ronan.champagnat,mourad.rabah*}*@univ-lr.fr, wvdaalst@pads.rwth-aachen.de,*

Keywords: Process Mining, Educational Process Mining, Moodle, Learning Analytics, Quality of Education.

Abstract: Learning Management Systems like Moodle generate detailed logs from student interactions, offering significant potential for learning analytics and educational process mining. However, raw logs capture interaction-based actions rather than actual learning processes, limiting their pedagogical relevance. To address this, we developed *Moodle2EventLog*, a tool that automates the cleaning, preprocessing, and semantic enrichment of Moodle logs. The tool operates in two modules: the first cleans and structures logs by generating event logs with key elements (case IDs, activities, timestamps), and the second enriches them by grouping low-level events into context-aware sub-processes and maps them to ”Semantic Activities” based on Bloom's Taxonomy. We tested *Moodle2EventLog* on logs from 65 Computer Science courses at Frederick University (471 students) from 2018–2022, and one course from La Rochelle University (36 students) in 2023, which serves as the use case in this paper. The enriched logs enabled deeper pedagogical analysis, such as identifying learning phase frequencies, studying specific activities and resource usage, and extracting semantically informed learner profiles linked to performance. Evaluation and instructor feedback validated the tool's effectiveness, demonstrating its ability to transform raw logs into pedagogically rich data, enabling the discovery of learning paths and providing insights unattainable with original Moodle logs.

1 INTRODUCTION

The widespread use of Learning Management Systems (LMSs) has transformed e-learning, generating huge amounts of data, including logs that capture every interaction between students, instructors, and course materials. This data offers significant potential for Learning Analytics (LA), Educational Process Mining (EPM) (Bogarín et al., 2017), and Data Mining with objectives such as understanding student profiles, analyzing learning strategies, correlating online behaviors with academic performance (Bey and Champagnat, 2022), providing personalized recommendations (Joudieh et al., 2023), and visualizing student learning paths (Álvarez et al., 2016). Understanding the learning process of students while covering a course is critical for both educators and learners. A learning process refers to the sequence of phases that a learner undergoes to acquire a course or skill, involving interactions with educational materials, engaging in discussions, practicing newly acquired knowledge, and applying it in new contexts. By uncovering these processes, educators can understand how students are studying their course, allowing them to improve course design and enhance the students' learning experience. However, the effectiveness of these analyses relies on data quality, leading to significant research on ensuring high-quality educational data (Umer et al., 2022).

Moodle, short for Modular Object-Oriented Dynamic Learning Environment, is a widely used open-source LMS that offers a flexible platform for creating, managing, and delivering online courses. Its adaptability and community-driven development

[a] https://orcid.org/0000-0003-3142-2962
[b] https://orcid.org/0000-0002-0955-6940
[c] https://orcid.org/0000-0001-5256-5706
[d] https://orcid.org/0000-0001-8136-5949
[e] https://orcid.org/0000-0001-7845-5425

Joudieh, N., van der Aalst, W. M. P., Champagnat, R., Rabah, M. and Nowakowski, S.
Moodle2EventLog: A Tool for Pedagogically-Driven Log Enrichment and Analysis.
DOI: 10.5220/0013327300003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 1*, pages 452-463
ISBN: 978-989-758-746-7; ISSN: 2184-5026

make it popular across educational institutions (Cole and Foster, 2007), supporting a wide range of teaching methodologies, including online and blended learning. Despite its rich feature set, Moodle logs primarily capture system interactions rather than the actual learning processes of students, thus lacking pedagogical relevance in its raw form.

Learning, however, has a **process-oriented nature** (Gašević et al., 2015), and Process Mining (PM) (Van der Aalst, 2016) has emerged as a valuable field for understanding these learning processes (Reimann et al., 2014). It bridges the gap between data analysis and process modeling, used to discover, monitor, and improve real processes by extracting knowledge from event logs available in information systems. EPM in particular focuses on analyzing sequences of student activities, such as interactions with e-learning platforms to identify patterns that influence educational outcomes (Bogarín et al., 2017). This specialized field helps teachers understand and impact the learning processes by visualizing how students engage with their learning environment and detecting bottlenecks or successful strategies.

To address the dual challenge of enhancing the pedagogical relevance of Moodle logs and preparing them for analysis, we propose a method for their semantic enrichment and we put it into action via *Moodle2EventLog*, a tool designated to automate the cleaning, preprocessing, and enrichment of Moodle logs. It operates in two modules: the first generates structured event logs by capturing essential elements such as case IDs, activities, and timestamps from raw Moodle logs, and the second enriches clean logs by grouping low-level events into context-aware sub-processes and maps them to **Semantic Activities**—higher-order learning processes derived from Bloom's Taxonomy. We define Semantic Activities, such as "Study," "Exercise," "Synthesize," and "Assess," to reflect the pedagogical intent behind student actions extracted from logs, offering a more meaningful analysis of student learning behavior. Through this transformation, the logs capture key stages of cognitive development, offering instructors a more insightful view of students' learning behaviors and processes. This enriched data serves as a valuable input source for **learning analytics**, **educational process and data mining**.

By using *Moodle2EventLog*, instructors and researchers can perform advanced pedagogical analyses, such as identifying learning phase frequencies, studying specific activities and resource usage, and extracting semantically informed learner profiles linked to performance. Moreover, the enriched logs are suitable for process mining, facilitating the discovery of process models that visualize students' learning paths. This dual benefit—enhancing learning analytics and supporting process mining—provides a more comprehensive understanding of student behavior in e-learning environments.

Our research is thus guided by the following questions:

- **RQ1.** How can Moodle logs be transformed and enriched to be pedagogically relevant for extracting the learning process of students in a course?
- **RQ2.** What impact does the categorization of low-level events into higher-order semantic activities have on the analysis of student learning processes?
- **RQ3.** How does *Moodle2EventLog* facilitate the use of enriched event logs for learning analytics and process mining, and what are the implications for improving instructional design?

The structure of this paper is as follows: Section 2 outlines key concepts and background knowledge. Section 3 reviews related work, emphasizing the importance of data quality and methods for processing educational logs, particularly those from Moodle. Section 4 describes the proposed tool, its architecture, functionalities, and how Semantic Activities are defined and mapped to Moodle logs. To demonstrate its effectiveness, Section 5 presents a case study, showcasing the tool's application, insights, analysis, and evaluation. Finally, Section 6 concludes the paper with a discussion of findings and future directions.

2 PRELIMINARIES

2.1 Process Mining

PM (Van der Aalst, 2016) analyzes sequences of events to reveal the execution of activities in business or educational settings. The main types of PM are:

- Process Discovery: Extracting a process model from event logs.
- Conformance Checking: Comparing the event log with a predefined model to identify deviations.
- Enhancement: Improving an existing process model using event logs to add new perspectives, like performance metrics.

2.2 Event Log

An event log is a structured collection of data that records specific activities or events as they occur within an information system. Formally, an **event log** $L = \{t_1, t_2, ..., t_k\}$ is a set of k **traces** where each trace t_i ($1 \leq i \leq k$) is a set of n_i consecutive events $t_i =< e_{i1}, e_{i2}...e_{in_i} >$ made by the same *case_id*. For process mining, an event log typically consists of the following main columns:

- Case ID: An identifier for each process instance.
- Activity: The specific action or task performed, such as submitting an assignment.
- Timestamp: The exact date and time when the activity occurred.

Event logs are commonly stored and shared in the XES (eXtensible Event Stream) format, an open standard that supports multiple attributes, hierarchical structures, and extensions, making it ideal for complex PM analyses.

2.3 Moodle Logs

Moodle logs typically include:

- Time: Timestamp of the logged event.
- User full name: The user performing the action.
- Affected User: The user targeted by the event.
- Component: Moodle module where the event occurred (e.g., Assignment, File).
- Event Context: The course or resource associated with the event.
- Event Name: The executed action.
- Description: Brief details, including user IDs and resources.
- Origin: Source of the event (e.g., web,...).
- IP Address: Originating IP address.

3 RELATED WORKS

EPM Using Moodle Logs

EPM has paved the way for a rich body of research that combines the strengths of data mining and process analysis to uncover insights from educational event data (Costa et al., 2020). Among the various sources of educational data, Moodle logs have been a focal point in numerous studies for revealing valuable insights into student learning behaviors (Wafda et al., 2022). For instance, in an effort to better understand students' learning processes, (Juhaňák et al., 2019) applied process mining to extract patterns from students' online quiz activities within an LMS. Similarly, (Cenka and Anggun, 2022) conducted weekly assessments of student activities over a semester, uncovering the most frequently accessed features, usage patterns, and the relationships between engagement and academic performance. Further, (Real et al., 2021) used both PM and Sequential Pattern Mining (SPM) to explore learning paths in an introductory programming course. By analyzing Moodle event logs, this study delved into specific activities, their sequence, and the actions performed by students, uncovering distinct behaviors and learning strategies.

Educational Data Quality

Despite the potential of Moodle logs for educational analysis, the quality of insights drawn depends heavily on the quality of the event logs. These logs often contain noise, making it necessary to filter and enrich the data to extract meaningful learning processes (Suriadi et al., 2017). Addressing this issue, (Umer et al., 2022) highlighted the importance and challenge of ensuring high-quality educational data for EPM and LA. They developed methods to extract standalone activities from Moodle's database and reformatted them to explicitly link learner data to process instances, thereby converting process-unaware logs into process-oriented event logs with a focus on quiz-taking activities. Similarly, (Aulia and Waspada, 2019) introduced an application for the preprocessing and exploratory data analysis of Moodle logs, using heuristic filtering and visualization techniques like flow control and dotted charts. However, despite these advancements, further improvements to the filtering techniques have yet to be explored.

The importance of proper data cleaning and preparation in Moodle logs cannot be overstated, as failure to do so can lead to overly complex or unstructured process models when applying PM algorithms (Etinger et al., 2018). One key challenge in process mining with Moodle data is creating models that accurately reflect general student behaviors without being too large or complex for teachers and students to interpret (Bogarin et al., 2014). In educational contexts, the comprehensibility of models is critical, as it ensures that both students and teachers can effectively monitor learning processes and use the feedback for improvements (Romero et al., 2016).

From Micro-Interactions to High-Level Learning Actions

In addition to Moodle-based research, several studies have focused on analyzing micro-interactions and low-level data, primarily in the context of Massive

Open Online Courses (MOOCs), which differ significantly from Moodle in terms of their data structure and log information. For example, (Yu et al., 2021) investigated how interactive navigational behaviors in connectivist MOOCs can be used to measure learning indicators such as engagement, progress, and achievement. Their analysis involved browser events like page loads, as well as mouse and keyboard interactions data. By applying sequence pattern mining and thematic analysis, they were able to transform these low-level interactions into higher-order behaviors.

These methodologies have also been applied to Self-Regulated Learning (SRL) strategies (Song et al., 2024). In such studies, trace data from various MOOCs and LMSs is first transformed into learning actions, which are then mapped to SRL processes using a pattern dictionary (Osakwe et al., 2024). A trace parser typically implements these mappings by comparing sequences of learning actions to a predefined pattern dictionary to identify SRL processes. For example, in (Maldonado et al., 2018), trace data from a MOOC was organized into six distinct learning actions, such as "Video-Lecture Begin" and "Assessment Pass," with process mining used to identify the most frequent sequences of these actions. By comparing these sequences to SRL theories, the researchers developed a pattern dictionary that mapped the actions to SRL processes like elaboration, evaluation, help-seeking, and task exploration. Similarly, (Li et al., 2024) employed a theoretically informed trace parser based on Bannert's SRL framework, categorizing SRL processes into cognition, metacognition, and emotion. Their trace parser consisted of an action library to convert raw trace data into learning actions and a process library to label these actions with specific SRL processes. Using Moodle logs, (Cerezo et al., 2020) assessed students' SRL skills in an online Spanish undergraduate course by analyzing 21,629 events. Upon preprocessing four key attributes were selected—time, anonymized student IDs, actions, and action details—by filtering out duplicates, irrelevant records, and nonessential actions like calendar checks. The authors refined 42 default Moodle actions into 16 SRL-relevant ones, grouped into five categories (Planning, Learning, Executing, Review, and Forum Peer Learning) aligned with Zimmerman's SRL model phases. This approach enabled analysis of student behaviors in relation to SRL theory, segmenting logs into pass/fail categories and course units. In contrast to the authors' approach of directly mapping individual Moodle events to high-level SRL categories, our method avoids this direct mapping as it isolates events from the broader context of the sub-processes they form. Instead, we first transform the log to extract the sub-processes that a specific resource undergoes with a particular student, which collectively contribute to the full process of the student's journey through the course. These sub-processes, enriched with their contextual information, are then mapped to higher-level activities, guided by Bloom's taxonomy, which we found more relevant given our focus on learning path recommendations within our framework (Joudieh et al., 2023).

Summary and Proposed Contributions

In summary, Moodle logs provide rich data for educational analysis, offering insights into student behavior, learning paths, and engagement to enhance the educational process. However, the effectiveness of such analyses depends greatly on the quality and richness of Moodle log data. While existing research has advanced data preparation and abstraction techniques, challenges specific to Moodle logs remain.

Our proposed tool addresses these by semantically enriching Moodle logs, transforming raw files into pedagogically meaningful event logs ready for process mining (XES format) without manual intervention or deep expertise in Moodle's structure. This enrichment enables process models to illustrate students' pedagogical learning paths rather than simple material interactions. Beyond EPM, the tool supports analyses such as studying learner behaviors, extracting study patterns, developing learner profiles, and examining learner performance and resource usage. Thus, our work contributes to improving the quality and value of educational data, specifically Moodle logs, enabling analyses that improve student learning experiences and support educators in refining teaching strategies and course design.

4 *Moodle2EventLog*

This section provides a detailed overview of the *Moodle2EventLog*, its core components, and their functionalities. The overall architecture is illustrated in Figure 1.

In a nutshell, the tool processes a log file downloaded from a Moodle course in Comma-Separated Value (CSV) format. A configuration file (referred to as "Config File" in Figure 1) accompanies the log and specifies metadata such as column names, file structure, output location, language (currently, only English is supported), time format, and enrollment method (Moodle or external). Using this information, the initial log file is processed in Module 1. The output of the former is a clean, filtered, student-centered

Moodle event log with the key columns necessary for process mining, as discussed in Section 2. While Module 1 generates a well-structured event log, Module 2 of the tool is dedicated to transforming and enriching this log with semantic information. The output of Module 2 can be one or more files provided in the XES standard, ready for use with any process mining discovery algorithm, as well as in CSV format for other types of analysis. The number of XES/CSV files generated is determined by the configuration file, as different files are created for different activities, resulting in distinct XES/CSV files for each. In the following sections, a detailed explanation of these two modules is provided.

4.1 Module 1: Cleaning and Preprocessing

This module relies on understanding Moodle logs - their structure and the information they contain to clean them and prepare them as event logs. In this module, two key pieces of information are extracted: the Moodle IDs of users in the log and their roles. As discussed in Section 2, a Moodle log includes all users interacting with a course. However, to focus on student behaviors and paths, the logs must be filtered to include only student actions.

First, the Moodle ID attribute is added by parsing the description field, which details user actions. If the action is administrative (e.g., ”Item Created with id...”), a dummy value is assigned to the Moodle ID. For actions involving students (e.g., ”The user with id 1...”), the student's Moodle ID is extracted. While Moodle logs can include multiple IDs for various resources, this tool focuses exclusively on student activities. To determine user roles, two methods are used depending on the enrollment system. If enrollment is managed externally, the log lacks role information, so filtering is done using the ”User full name” field, excluding non-students. If enrollment occurs within Moodle, the ”Role assigned” event is parsed (e.g., ”The user with id '1' assigned the role with id '5' to the user with id '2'”) to identify student roles. The Moodle IDs are matched with their corresponding role IDs, filtering for id '5', which typically denotes students.

After filtering, the log includes only student interactions, with the Moodle ID as the *case_id*, uniquely identifying each student as a process instance.

4.2 Module 2: Transformation and Enrichment

Structurally, the output of Module 1 is ready for use as input for process mining algorithms and other analyses. However, to move beyond low-level interaction records, it is necessary to elevate and enrich these logs to support higher-level analysis of the learning process. To accomplish this, Module 2 maps Moodle interaction events to the pedagogical actions a student takes while learning. This approach focuses on capturing how a student learns a course rather than merely how they navigate it through Moodle, as indicated by the *Event Name* in the original logs.

Moodle is composed of various components where each can be viewed as a process instance with its own sequence of events (Rotelli and Monreale, 2023). Typically, one would examine the process of a specific component to understand its sequence of interactions. However, in this case, the student is treated as the process instance. We observe the student's learning process as a sequence of sub-processes, each corresponding to interactions with different contexts within Moodle. With this in mind, the enrichment step is preceded by a transformation step for context-aware grouping of events that represents these sub-processes. As illustrated in Figure 2, events are grouped by *case_id* and *event context*, maintaining the sequence across different contexts. For instance, *case_id* 1 interacts with *event context* c1 three times at times t1, t2 and t6, interrupted by an interaction with context c2 at time t3. Thus, the interaction of *case_id* 1 with c1 is divided into two different groups and remain separate due to intervening interactions in other contexts. The time is then adjusted to reflect the start time of the first event in each group.

In the enrichment step, each group of events is mapped to one of the Semantic Activities outlined in Table 1 by a rule-based approach explained in what follows. An optional column, ”Peda Activity”, can be added and returned as a new XES/CSV file, combining the semantic activity with the original event context to provide a finer level of granularity, such as indicating that ”a student is studying **lecture 1**”. While the enrichment considers the student as the case_id the user can further select different IDs for the case_id to perform the analysis. Currently, the tool in Module 1 extracts only students IDs.

4.2.1 The up Bringing of ”Semantic Activities”

Benjamin Bloom's Taxonomy of Learning Domains, developed in 1956, categorizes learning into three domains: cognitive, affective, and psychomotor. Bloom views these domains as progressive, with learners

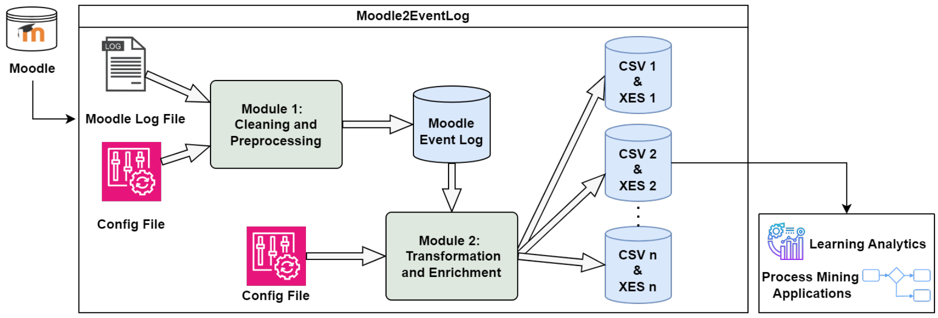

Figure 1: *Moodle2EventLog* Architecture.

case id	time	event context	event name
1	t1	c1	e1
1	t2	c1	e2
1	t3	c2	e1
2	t4	c1	e2
2	t5	c1	e1
1	t6	c1	e3
2	t7	c2	e3
2	t8	c2	e3
2	t9	c2	e4

Moodle Event Log

case id	time	event context	event list
1	t1	c1	[e1, e2]
1	t3	c2	[e1]
2	t4	c1	[e2, e1]
1	t6	c1	[e3]
2	t7	c2	[e3, e3, e4]

Transformation Step

Semantic Activity	Peda Activity
s1	s1: c1
s2	s2: c2
s1	s1: c1
s3	s3: c1
s4	s4: c2

Enrichment Step

Figure 2: Event log transformation and enrichment with Semantic and Pedagogical Activities.

advancing through six stages in each domain as their knowledge, attitudes, and skills evolve (Bloom, 1956). While the affective and psychomotor domains are important, the cognitive domain—focused on intellectual capabilities like knowledge and 'thinking'—is most measurable through digital interactions in LMSs. For this reason, Bloom's Taxonomy, a framework organizing cognitive objectives into six hierarchical levels, is crucial for analyzing student data from Moodle logs. Bloom's Taxonomy (Anderson and Krathwohl, 2001) classifies cognitive processes into Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating, guiding educators in defining learning outcomes (Fastiggi, 2019; Shabatura, 2022). By mapping Moodle logs to these cognitive levels, teachers can see how students' interactions correspond to deeper cognitive processes, validating engagement with course material in alignment with these objectives.

Using Bloom's Revised Taxonomy, we mapped Moodle activities to cognitive levels, as shown in Table 1. Activities like viewing course resources correspond to Remembering and Understanding (semantic activity **Study**), while exercises align with Applying and Analyzing (semantic activity **Exercise**). Higher-order cognitive tasks such as "Evaluating" and "Creating" are translated into the semantic activities **Assess** and **Synthesize**, respectively, often associated with project work or knowledge reflection and evaluation. Additional activities like **View, Feedback,** and **Interact** occur during the learning process but are not directly tied to Bloom's levels, instead supporting the overall learning progression. For activities like Study, Exercise and Assess we differentiate between passive and active modes (denoted by _P and _A, respectively). For example, downloading lecture materials is considered a *passive* action, as it does not guarantee completion, while submitting an assignment is an *active* action, signifying completion. By integrating Bloom's framework into Moodle logs, we transform the system-based interaction data into a rich pedagogical tool. This allows instructors to not just see what students are doing, but how their actions align with different stages of learning, providing deeper insight into the cognitive progression of each student and validating learning outcomes at both individual and course levels.

4.2.2 Rule-Based Semantic Activity Extraction

After the transformation step, events with the same context and case_id are grouped into an eventList, as previously explained. This eventList, along with the context and component information recorded in the Moodle logs, is used in the subsequent rule-based approach (Table 2), which applies a set of rules to deter-

Table 1: Semantic Activities to Bloom's Taxonomy.

Semantic Activity	Bloom's Level	Meaning: Explanation
Study (Passive/Active)	Remember, Understand	Course review: Acquiring and comprehending knowledge.
Exercise (Passive/Active)	Apply, Analyze	Practical work: Solving problems and applying knowledge.
Assess (Passive/Active)	Evaluate	Evaluation: Testing understanding and self-reflection.
Synthesize	Create	Project/Practical: Applying knowledge in new contexts.
View	N/A	Exploration: Exploring course materials.
Feedback	N/A	Feedback: Receiving grades or comments.
Interact	N/A	Interaction: Participating in chats or forums.

mine the appropriate Semantic Activity for each contextualized subprocess (eventList + component + context).

To develop the mapping algorithm, we gained a base on Moodle components and processes from its official website, various studies (Rotelli and Monreale, 2023; Costa et al., 2020; Nammakhunt et al., 2023), and a collection of Moodle logs that we collected. The algorithm consists of a series of rules structured in an if-else format. Essentially, each group of events associated with a case_id, together with the context and component, is translated into a semantic activity. This translation is achieved by identifying key events, parsing event contexts for relevant keywords, and considering specific components.

To facilitate this process, we defined a set of keyword dictionaries relevant to certain semantic activities, such as "exercise," "assess," "interact," "feedback," and "outline." For instance, the "exercise" dictionary includes terms like *"lab", "exercise", "worksheet", "homework", ...*, while the "outline" dictionary identifies terms like *"outline", "syllabus", "agenda", "description", ...* to differentiate between viewing course descriptions and actual studying.

Some components consistently lead to the same semantic activity; for example, both `Forum` and `Chat` are always mapped to the "Interact" activity. However, certain event names, such as "Course Module Viewed," can indicate different semantic activities depending on the resource type, identified through keyword searches. For instance, viewing an example test file suggests a passive assessment activity, while viewing a lab exercise sheet indicates a passive exercise activity. Joining a Zoom session, on the other hand, is categorized as an active study activity. Events that do not signify any pedagogical value are labeled as "Others" and can be filtered out later by instructors.

These semantic activities assist teachers in interpreting student behavior, identifying areas needing support, and adjusting teaching strategies to better meet student needs. For example, if students engage in passive exercises ("Exercise_P") without making active submissions ("Exercise_A"), this may indicate several potential issues: the course design might lack sufficient opportunities for active engagement and submission, students could be struggling to complete the assignments, or additional study materials may be necessary to help them grasp the concepts.

5 *Moodle2EventLog* IN ACTION

To illustrate the proposed semantic enrichment for Moodle logs using the *Moodle2EventLog* tool, we analyzed logs from two universities: 471 students across 65 computer science courses at Frederick University, Cyrpus (2018–2022), and four log files from a 2022–2023 course at La Rochelle University, France. Detailed results from Frederick University are presented in our previous work (Joudieh et al., 2024), where we applied trace clustering on semantically enriched traces to demonstrate how semantic activities can establish learner profiles from discovered process models, focusing on a novel Trace Clustering Algorithm. In this paper, we focus on a 2023 Process Mining course for Master 1 students at La Rochelle University, which lasted three months and involved 36 students. The course included three assignments, a practical test and a course test, both conducted as Multiple Choice Questions (MCQ) exams, and a final project.

This section presents insights from the semantic enrichment applied to this course, followed by advanced analyses, such as trace clustering and the extraction of learner profiles using semantic activities. These profiles are linked to students' grades in each cluster. Finally, we evaluate Moodle2EventLog by comparing input and output logs and sharing instructor feedback on their experiences with the tool.

5.1 Insights Brought by the Semantic Enrichment

This section explores the analyses possible with the enriched Moodle logs produced by *Moodle2EventLog*, focusing on model discovery, statistical analysis of semantic activities, and an examination of study behavior using dotted charts.

Figure 3 presents the directly follows graph discovered using the original Moodle Event names from the cleaned event log generated from Module 1 of the tool vs Figure 4 with semantic activities. This comparison highlights the primary advantage of our tool:

Table 2: Rule-based Mapping Algorithm for Extracting Semantic Activities.

Input: eventList, context, component, **Output:** semantic activity (SA) **Define:** exercise, assess, view, interact, feedback, outline, project	
Rule	**Condition and Semantic Activity Mapping**
Rule 1	If `'Course activity completion updated'` in eventList: If context contains exercise → SA = 'Exercise_A' Else if context contains assess → SA = 'Assess_A' If context contains interact → SA = 'Interact' Else → SA = 'Study_A'
Rule 2	If component in {`Assignment, File submissions`}: If any of {`A submission has been submitted, File uploaded, Submission updated`} in eventList: If context contains project → SA = 'Synthesize' Else → SA = 'Exercise_A' Else if `'Course module viewed'` in eventList → SA = 'Exercise_P' Else → SA = 'others_submission_viewing'
Rule 3	If component = 'Quiz': If any of {`Quiz attempt started/submitted or updated`} in eventList: If context contains project → SA = 'Synthesize' Else → SA = 'Assess_A' Else if `'Course module viewed'` in eventList → SA = 'Assess_P' Else → SA = 'others_quiz_attempt_viewing'
Rule 4	If context = 'other' → SA = 'others'
Rule 5	If component in {`Forum, Chat, Choice`} → SA = 'Interact'
Rule 6	If component in {`H5P Package, H5P`} → SA = 'Study_P'
Rule 7	If component in {`Feedback, Overview report, User report`} → SA = 'Feedback'
Rule 8	If component in {`Scheduler, User tours`} → SA = 'View'
Rule 9	If component = 'Lesson': If `'Question answered'` in eventList → SA = 'Exercise_A' Else if any of {`Lesson started, Lesson resumed`} in eventList → SA = 'Study_A' Else if any of {`Message viewed, Group message sent`} in eventList → SA = 'Interact' Else if `'Question viewed'` in eventList → SA = 'Exercise_P' Else if context contains: exercise → SA = 'Exercise_A', assess → SA = 'Assess_A'
Rule 10	If any of {`Course module viewed, Zip archive of folder downloaded`} in eventList: If context contains: exercise → 'Exercise_P', assess → 'Assess_P', outline → 'View', feedback → 'Feedback', interact → 'Interact', otherwise → 'Study_P'
Rule 11	If component = 'System': If any of {`Badge listing viewed, User graded`} in eventList → 'Feedback' Else → 'View'
Rule 12	If component = `Zoom meeting`: If `'Clicked join meeting button'` in eventList → SA = 'Study_A' Else → SA = 'Study_P'
Rule 13	Else → SA = 'others'

it simplifies process models while providing a pedagogically relevant illustration of the underlying learning processes in the course. The analysis of the directly follows graph reveals several insights into student learning behaviors. The most common learning paths begin with Study_P, followed by Exercise_P and Synthesize, indicating a typical sequence of learning, practice, and assessment. Feedback loops between Exercise_P and Synthesize, as well as Assess_A and Assess_P, suggest iterative cycles of practice and assessment. Also, alternative paths in the graph imply that students may take different routes based on individual needs or preferences. Figure 5 illustrates the frequency of each semantic activity, showing that the two most prevalent activities are passive studying and exercising, with exercising being the more frequent. This is justified by the course structure, which includes two exercise sessions for each lecture. Figure 6 provides another perspective, allowing teachers to focus on specific activities and visualize student paths through a dotted chart. Each line in the chart represents a student's learning path filtered for the Study_P activity, with colors indicating different resources or event contexts. Notably, the chart shows random access behavior on December 4th and 5th, which teachers confirmed coincided with the quiz period when students review all course materials in preparation.

Through these analyses, this tool can be integrated with Moodle via a dashboard for teachers, while offering researchers an effective means to apply process mining and educational analysis to Moodle logs.

5.2 Extracting Semantic Learner Profiles

In this part, we present the results of an advanced analysis to extract learner profiles from the enriched logs. We apply trace clustering using the Improved FSS encoding, as described in (Joudieh et al., 2024), and Hierarchical Agglomerative Clustering (HAC)

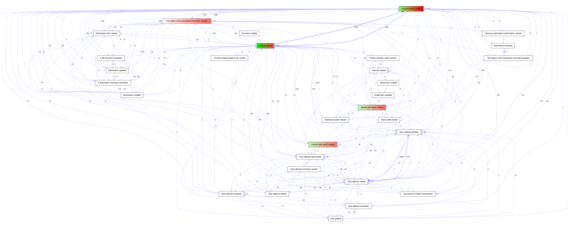

Figure 3: Process Model Before Enrichment using Moodle Event Name.

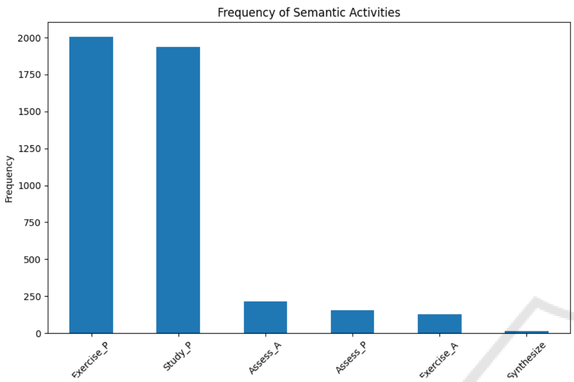

Figure 5: Frequency of Semantic Activities.

with Ward linkage, a bottom-up approach that merges similar clusters. The dendrogram in Figure 7 suggests a cut, indicated by the red line, which results in four distinct clusters.

Table 3 provides a summary of the resulting clusters and their profiles. For each cluster, it details the number of traces, trace lengths, the relative frequency of studying and exercising, as well as the average theoretical (course grade) and practical (lab grade) scores. The relative frequency of a semantic activity is calculated as the proportion of its occurrences relative to the total activities within the cluster. Studying and exercising were specifically chosen for analysis because other activities, such as assessment and synthesis tasks (e.g., the group project), are typically mandatory or collaborative, making them less reflective of individual differences.

Table 3: Resulting Cluster Analysis and Profiles.

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Number of Students	12	5	8	11
Mean Trace Length [Min-Max]	[58-123]	[141-212]	[56-86]	[108-160]
Mean Lab Grade	12.4	11.6	9.4	10.9
Mean Course Grade	12	13.8	10	13.4
Relative Frequency of Exercise (in %)	46	42.1	51.4	39.3
Relative Frequency of Study (in %)	47.6	52.7	42.5	56.3

Analyzing the profiles of the clusters reveals distinct learning patterns and their impact on performance. Cluster 0, the largest group, demonstrates a balanced approach to studying and exercising, resulting in relatively consistent grades across both the course and lab exams. In contrast, Clusters 1 and 3

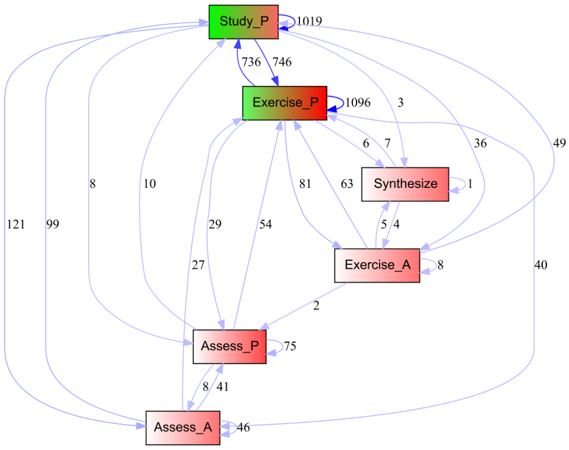

Figure 4: Process Model Using Semantic Activities.

Figure 6: Dotted Chart for Study_P.

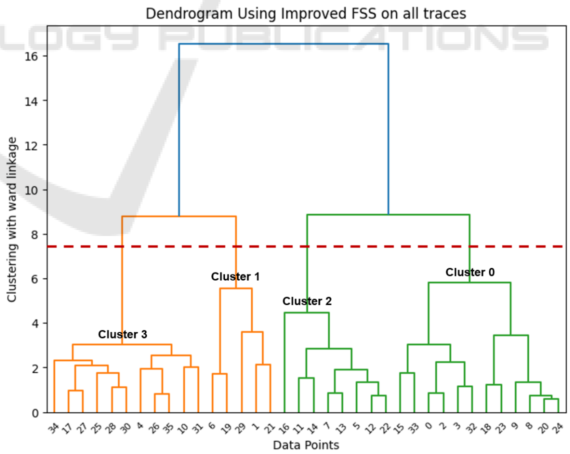

Figure 7: Dendrogram from HAC with ward linkage.

emphasize studying more than exercising, which correlates with higher performance in the course MCQ. However, this focus comes at the expense of their lab grades, with Cluster 3 performing particularly lower than Cluster 1. This difference may be attributed to Cluster 1's slightly higher engagement in exercises, which likely supports better practical understanding. Cluster 2 however, showcases a different learning be-

Table 4: A Comparison between the input and output of *Moodle2EventLog*.

	Input Moodle Log	Output Moodle Event Log
Number of Cases	45	36
Number of Event Classes	66	6
Number of Events	19422	7188
Trace Length [Min-Max]	[7-1706]	[56-212]

havior characterized by a lower emphasis on studying relative to exercising. This suggests that students in this cluster may tend to dive into exercises without adequate preparation or understanding of the course materials. Consequently, their grades in both MCQ exams are relatively low, indicating potential gaps in understanding and knowledge retention. Particularly concerning is the notably failing grade in the lab exam, suggesting a lack of foundational understanding or practical application skills. This analysis highlights the importance of a balanced approach to studying and exercising. A focus on thorough preparation before engaging in exercises appears to be crucial for achieving consistent performance in both course assessments and lab exams. The use of semantic activities in this analysis provided a deeper understanding of students' performance by linking their learning actions to outcomes, which the raw Moodle event names could not have achieved.

5.3 Evaluation of *Moodle2EventLog*

For evaluation, the tool was used on the course logs, while selecting the ”Semantic Activity” as the ”activity” for the final output files.

5.3.1 Raw vs Enriched Event Logs

To evaluate *Moodle2EventLog*, we compared the input and output files, as shown in Table 4. This comparison reveals that the number of cases decreased from 45 to 36, effectively filtering out non-student users, as confirmed by the course instructors. Additionally, the reduction in event classes from 66 to 6 –from *Event name* in the input to *Semantic Activity* in the output–demonstrates the tool's success in abstracting event-level details. This abstraction is evident in the reduced log length and the changes in minimum and maximum trace lengths. The time taken to process the input CSV file and generate the XES output file for this course log was 17 seconds.

5.3.2 Instructors' Feedback

To evaluate the effectiveness of *Moodle2EventLog* and its alignment with our research questions, feedback was collected via a questionnaire from five instructors of the Process Mining course who used the tool to analyze their Moodle course log data. The evaluation consisted of seven questions answered using a 5 point Likert scale, where instructors rated various aspects of the tool from 1 (strongly disagree) to 5 (strongly agree). Additionally, three open-ended questions allowed instructors to reflect on their experiences with the tool (Questionnaire can be accessed via this link). The results of the Likert scale questions, summarized in Table 5, provide insights into the tool's strengths and highlight areas for further improvement.

The tool's ability to help instructors identify new patterns in student behavior received an average score of 4.2 (SD = 0.83), indicating that the categorization of low-level events into semantic activities revealed insights that were not previously visible to instructors. This is a key finding relevant to **RQ2**, as it highlights the significant impact of semantic enrichment on the analysis of student interactions with course content, making the learning process more understandable.

Regarding the interpretability of the enriched event logs, the tool was rated highly, with a score of 4.4 (SD = 0.54). This further emphasizes how the semantic categorization simplifies the logs, making them easier for instructors to use in their analysis, thereby supporting **RQ2** and **RQ3**. However, when asked about the tool's ability to influence teaching strategies, instructors provided a lower rating of 3.2 (SD = 0.83). While the tool provided valuable insights, further refinement is needed to ensure that these insights are actionable enough to directly inform instructional design, as explored in **RQ3**.

In terms of accuracy, the tool received a score of 3.6 (SD = 0.54) for capturing how students interacted with course materials, suggesting some room for improvement. Nonetheless, the process models generated from the semantic activities were rated as pedagogically relevant (4.2, SD = 0.83) and simplified (4.4, SD = 0.89), indicating that the semantic activities facilitated a clearer and more useful understanding of student behavior. This reinforces the importance of log enrichment for generating meaningful pedagogical models, as discussed in **RQ2** and **RQ3**.

In addition to the structured feedback, instructors provided valuable insights through open-ended questions regarding the tool's functionality. When asked for suggestions on enhancing the tool's ability to provide more meaningful insights, instructors emphasized the need for more detailed statistical data on user learning sequences, such as the duration between events and time of day. They also highlighted the importance of refining semantic activities to make them more precise and meaningful. For instance, distinguishing between different types of ”study” activities—such as passive study related to reading lectures versus exploring external resources—could signifi-

Table 5: Instructors Feedback Likert Scale Results.

Evaluation	Mean	Standard Deviation (SD)
Reflecting Learning Process	4	0.70
Identifying New Patterns	4.2	0.83
Interpretability of Event Logs	4.4	0.54
Influencing Teaching Strategies	3.2	0.83
Accuracy of Captured Data	3.6	0.54
Process Models Pedagogical Relevance	4.2	0.83
Process Models Simplicity	4.4	0.89

cantly improve clarity. Furthermore, instructors noted the potential for enriched logs to provide insights into students' engagement with course materials from other courses, which could influence their current assignments. This feedback suggests that while the current semantic activities enrich the models, there is a need for further refinement to better capture indicators of student engagement. Regarding expectations for additional analyses from enriched Moodle logs, instructors expressed interest in tracking the time spent working on courses outside of class and linking completed work to learner progress through assessments.

In summary, the instructor feedback shows that *Moodle2EventLog* provides a valuable tool for enriching Moodle logs, making them more pedagogically relevant and improving their interpretability for learning analytics and process mining. While the tool has demonstrated strong results in helping instructors uncover new insights into student behavior and generating simplified models of learning processes, additional work is required to ensure that these insights can be integrated into instructional design and teaching strategies.

6 CONCLUSION

Moodle generates extensive data invaluable for analyzing student behavior, learning profiles, and engagement. This data helps educators refine course design and supports students in improving their learning experiences. While previous studies have examined aspects like quiz behavior and resource usage, challenges with data quality, cleaning, and preparation persist. The granularity of Moodle data further complicates extracting meaningful insights. To address these issues, we developed Moodle2EventLog, a tool that converts raw Moodle logs into student-centered event logs optimized for process mining. Validated across multiple courses, the tool cleans and transforms logs by filtering non-student activities and aggregating detailed events into semantic activities, enabling clearer analysis and insightful visualizations of learning patterns.

Despite its benefits, *Moodle2EventLog* has limitations. The tool could be enhanced by expanding language support, integrating natural language processing for advanced semantic extraction, and adding features for analyzing additional Moodle resources. While our research focused on computer science courses, adaptations may be needed for other domains. Improving the user interface would also increase accessibility for educators and researchers. Additionally, Moodle's logging limitations, such as inadequate recording of H5P packages and lessons, may lead to misinterpretations of student engagement, and certain user interactions are only captured in site logs, not course logs.

Feedback from instructors indicates that while the semantic enrichment of Moodle logs yields meaningful pedagogical insights and simplifies process model interpretation, there is a demand for more granular insights—particularly regarding student transitions between learning activities. Future iterations of *Moodle2EventLog* will prioritize refining these aspects to enhance clarity and educational relevance. In conclusion, our work demonstrates that by transforming and enriching Moodle logs, we can significantly improve their pedagogical relevance, ultimately supporting more effective learning analytics and process mining applications.

REFERENCES

Anderson, L. W. and Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.

Aulia, D. and Waspada, I. (2019). The design of exploratory application and preprocessing of event log data in lms moodle-based online learning activities for process mining. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, 5:124–133.

Bey, A. and Champagnat, R. (2022). Analyzing Student Programming Paths using Clustering and Process Mining:. In *Proceedings of the 14th Int. Conf. on Computer Supported Edu.*, pages 76–84.

Bloom, B. (1956). A taxonomy of cognitive objectives. *New York: McKay*.

Bogarin, A., Romero, C., Cerezo, R., and Sanchez-Santillan, M. (2014). Clustering for improving educational process mining. In *Proceedings of the 4th Int. Conf. on Learning Analytics And Knowledge*, page 11–15.

Bogarín, A., Cerezo, R., and Romero, C. (2017). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8.

Cenka, N. and Anggun, B. (2022). Analysing student behaviour in a learning management system using a process mining approach. *Knowledge Management & E-Learning: An Int. Journal*, 14:62–80.

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 Edu.*, 32(1):74–88.

Cole, J. and Foster, H. (2007). *Using Moodle: Teaching with the popular open source course management system*. "O'Reilly Media, Inc.".

Costa, J., Azevedo, A., and Rodrigues, L. (2020). Educational process mining based on moodle courses: a review of literature. In *20th Conf. of the Portuguese Association for Information Systems (CAPSI)*.

Etinger, D., Orehovacki, T., and Babic, S. (2018). Applying process mining techniques to learning management systems for educational process model discovery and analysis. In *Intelligent Human Systems Integration (IHSI 2018)*, pages 420–425. Springer.

Fastiggi, W. (2019). Applying bloom's taxonomy to the classroom. *Technology for Learners*.

Gašević, D., Dawson, S., and Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59:64–71.

Joudieh, N., Eteokleous, N., Champagnat, R., Rabah, M., and Nowakowski, S. (2023). Employing a process mining approach to recommend personalized adaptive learning paths in blended-learning environments. In *12th Int. Conf. in Open and Distance Learning, Athens, Greece*, volume 12.

Joudieh, N., Trabelsi, M., Champagnat, R., Rabah, M., and Eteokleous, N. (2024). Using trace clustering to group learning scenarios: An adaptation of fss-encoding to moodle logs use case. *In Proceedings of the 16th Int. Conf. on Computer Supported Edu.*, pages 247–254.

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*, 92:496–506.

Li, T., Fan, Y., Srivastava, N., Zeng, Z., Li, X., Khosravi, H., Tsai, Y.-S., Swiecki, Z., and Gašević, D. (2024). Analytics of planning behaviours in self-regulated learning: Links with strategy use and prior knowledge. In *Proceedings of the 14th Learning Analytics and Knowledge Conf.*, page 438–449.

Maldonado, 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. *Computers in Human Behavior*, 80:179–196.

Nammakhunt, A., Porouhan, P., and Premchaiswadi, W. (2023). Creating and collecting e-learning event logs to analyze learning behavior of students through process mining. *Int. Journal of Information and Edu. Technology*, 13(2):211–222.

Osakwe, I., Chen, G., Fan, Y., Rakovic, M., Singh, S., Molenaar, I., and Gašević, D. (2024). Measurement of self-regulated learning: Strategies for mapping trace data to learning processes and downstream analysis implications. In *Proceedings of the 14th Learning Analytics and Knowledge Conf.*, page 563–575.

Real, E. M., Pimentel, E. P., and Braga, J. C. (2021). Analysis of learning behavior in a programming course using process mining and sequential pattern mining. In *2021 IEEE Frontiers in Edu. Conf. (FIE)*, pages 1–9.

Reimann, P., Markauskaite, L., and Bannert, M. (2014). e research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, 45(3):528–540.

Romero, C., Cerezo, R., Bogarín, A., and Sánchez-Santillán, M. (2016). *Educational Process Mining: A Tutorial and Case Study Using Moodle Data Sets*, chapter 1, pages 1–28. John Wiley and Sons, Ltd.

Rotelli, D. and Monreale, A. (2023). Processing and understanding moodle log data and their temporal dimension. *Journal of Learning Analytics*, 10:126–141.

Shabatura, J. (2022). Using bloom's taxonomy to write effective learning outcomes. *University of Arkansas*.

Song, Y., Oliveira, E., Kirley, M., and Thompson, P. (2024). A case study on university student online learning patterns across multidisciplinary subjects. In *Proceedings of the 14th Learning Analytics and Knowledge Conf.*, page 936–942.

Suriadi, S., Andrews, R., ter Hofstede, A., and Wynn, M. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems*, 64:132–150.

Umer, R., Susnjak, T., Mathrani, A., and Suriadi, S. (2022). Data quality challenges in educational process mining: Building process-oriented event logs from process-unaware online learning systems. *Int. Journal of Business Information Systems*, 39:569 – 592.

Van der Aalst, W. (2016). *Process mining: data science in action*. Springer.

Wafda, F., Usagawa, T., and Mahendrawathi, E. (2022). Systematic literature review on process mining in learning management system. *IEEE Int. Conf. on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pages 160–166.

Yu, H., Harper, S., and Vigo, M. (2021). Modeling micro-interactions in self-regulated learning: A data-driven methodology. *Int. Journal of Human-Computer Studies*, 151:102625.

Álvarez, P., Fabra, J., Hernández, S., and Ezpeleta, J. (2016). Alignment of teacher's plan and students' use of lms resources: Analysis of moodle logs. In *15th Int. Conf. on Information Technology Based Higher Education and Training (ITHET)*, pages 1–8.