13th International Conference on Applied Informatics
Eszterházy Károly Catholic University
Eger, Hungary, February 19–21, 2026

Analyzing Student Pathways Through Process Mining

Erika Baksáné Varga[a], **Attila Baksa**[b]

[a]Institute of Informatics, University of Miskolc
erika.b.varga@uni-miskolc.hu

[b]Institute of Applied Mechanics, University of Miskolc
attila.baksa@uni-miskolc.hu

Abstract

Higher education institutions increasingly rely on data-driven methods to understand student progress, identify bottlenecks, and support evidence-based curriculum development. Process mining [1] offers a powerful framework for analyzing student pathways by transforming administrative records into event logs and comparing actual study trajectories with the intended curriculum structure. This extended abstract outlines a methodological framework for applying conformance checking techniques [2], [3] to evaluate how closely students follow the recommended study plan and to identify systematic deviations that may indicate curricular misalignment or structural challenges.

Introduction

Universities typically define a recommended curriculum that specifies the ideal sequence and timing of courses. In practice, however, students often diverge from this plan due to prerequisite issues, scheduling conflicts, academic difficulties, or personal circumstances. Understanding these deviations is essential for improving program design, supporting at-risk students, and enhancing institutional decision-making.

Process mining provides tools for modeling real student behavior and comparing it with the normative process. While previous studies have applied process mining in educational contexts [6], [4], there is still a need for systematic, reproducible methods that integrate high-quality event log construction with alignment-based

conformance checking. This research aims to address this gap by developing a transparent and interpretable framework tailored to higher-education data.

Research Objectives

The study focuses on the following objectives:

- to construct a clean, `PM4Py`-compatible event log representing student lifecycle data,
- to formalize the official curriculum as a reference process model suitable for conformance checking,
- to apply alignment-based techniques to compare actual student trajectories with the normative model,
- to identify and categorize typical deviation patterns (e.g., delays, course reordering, repeated attempts),
- to explore how these deviations can support curriculum refinement and student success initiatives.

Methodology

The research begins with the development of a reproducible event log pipeline that transforms raw academic records into structured traces. Each trace represents a student's progression through courses, enriched with timestamps, course codes, and status information. Special attention is given to preserving temporal order, handling repeated enrollments, and encoding grade or completion status using a consistent semantic hierarchy.

The official curriculum is modeled as a process structure [5] that captures recommended semester-level sequencing and prerequisite relationships. Using `PM4Py`, alignment-based conformance checking is applied to quantify the fitness of each student's trajectory and to detect deviations at the activity level. Deviation types are categorized to reveal recurring behavioral patterns, such as early or late course completion, prerequisite violations, or atypical course combinations.

The analysis is complemented by descriptive statistics and visualization techniques to highlight program-level trends and potential structural issues in the curriculum.

Expected Outcomes

The expected outcome of this research is a methodological contribution that demonstrates how educational process mining can be used to evaluate the alignment between actual student behavior and the intended curriculum. The study aims to:

- provide a transparent and reproducible framework for constructing educational event logs,
- identify common deviation patterns that may signal curricular bottlenecks or structural inconsistencies,
- support data-informed curriculum development and academic advising,
- contribute to the broader adoption of process mining techniques in higher-education analytics.

References

[1] W. M. P. van der Aalst: *Process Mining: Data Science in Action*, Springer, 2016.

[2] W. M. P. van der Aalst, A. Adriansyah, B. F. van Dongen: *Replaying history on process models for conformance checking and performance analysis*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2.2 (2012), pp. 182–192, doi: 10.1002/widm.1045.

[3] A. Adriansyah: *Aligning observed and modeled behavior*, PhD thesis, Technische Universiteit Eindhoven, 2014, doi: 10.6100/IR770080.

[4] A. Bogarín, R. Cerezo, C. Romero: *A survey on educational process mining. WIREs Data Mining Knowl Discov*, tech. rep., 8: e1230, WIREs Data Mining Knowl Discov, 2018, doi: 10.1002/widm.1230.

[5] S. J. J. Leemans, D. Fahland, W. M. P. van der Aalst: *Discovering Block-Structured Process Models from Event Logs - A Constructive Approach*, in: Application and Theory of Petri Nets and Concurrency, ed. by J. Colom, J. Desel, vol. 7927, PETRI NETS 2013. Lecture Notes in Computer Science, 2013, doi: 10.1007/978-3-642-38697-8_17.

[6] N. Trcka, M. Pechenizkiy, W. M. P. van der Aalst: *Process Mining from Educational Data. Ch. 9. in Handbook of Educational Data Mining*, Chapman & Hall, CRC Press, 2010, doi: 10.1201/b10274.