



## Process Mining on Curriculum-Based Study Data – a Case Study at a German University

---

Richard Hobeck, Luise Pufahl and Ingo Weber

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

November 7, 2022

# Process Mining on Curriculum-based Study Data: A Case Study at a German University

Richard Hobeck<sup>1</sup>, Luise Pufahl<sup>1</sup>, and Ingo Weber<sup>1</sup>

Chair of Software and Business Engineering,  
Technische Universität Berlin, Germany, `(firstname).(lastname)@tu-berlin.de`

**Abstract.** On their trajectory through educational university systems, students leave a trace of event data. The analysis of that event data with a process lens poses a set of domain-specific challenges that is addressed in the field of Educational Process Mining (EPM). Despite the vast potential for understanding the progress of students and improving the quality of study programs through process mining, a case study based on an established process mining methodology is still missing. In this paper, we address this gap by applying the state-of-the-art process mining project methodology (PM<sup>2</sup>) in an EPM case study with a focus on student trajectory analysis at a German university. We found that process mining can create actionable items to improve the quality of university education. We also point out domain-specific challenges, like handling reoccurring exams (retaken after failing) for future research in EPM. Finally, we observe insights of some value in our case.

**Keywords:** Educational Process Mining · Curriculum Mining · Case Study.

## 1 Introduction

Students are the lifeblood of universities and their main reason for existence. Universities' role in society is unquestionably large, with more than 40% of the population in Europe<sup>1</sup> and 24% in the U.S.<sup>2</sup> attaining at least a Bachelor degree. We aim to recruit young academic talents from the best graduating students to sustain excellent work across sectors. But talent is not spread equally: the international competition among universities for top students is largely driven by university rankings, in which the quality of teaching is a pivotal factor. On the other end of the spectrum, unfortunately large shares of students leave colleges and universities without a degree – for the U.S., this was the case for at least 20

<sup>1</sup> <https://www.statista.com/statistics/1093466/eu-28-adults-with-tertiary-education-attainment/>, 2020 data, accessed 2022-08-22.

<sup>2</sup> Estimated population: 328.24M; Bachelor, Master, and Doctoral degrees: 80.33M; both for 2019. Sources: <https://www.statista.com/statistics/183489/population-of-the-us-by-ethnicity-since-2000/> and <https://www2.census.gov/programs-surveys/demo/tables/educational-attainment/2019/cps-detailed-tables/table-1-1.xlsx>, accessed 2022-08-22.

million inhabitants in 2019<sup>3</sup>. In summary, universities can be expected to have a high level of motivation to increase student success.

Once students enroll in a program, they complete courses to receive their degrees – with varying levels of freedom, depending on the respective university system. On their trajectory through the university system, those students leave a trace of data. Some universities analyze that data with basic statistics, others employ more elaborate data mining techniques [4]. Insights from such analyses inform refinement of curricula and course offerings, to improve the quality of teaching and (global) competitiveness, and to lower drop-out rates [13].

As a set of process-focused data science techniques, process mining is widely used in various business environments [1]. A core value proposition of process mining is an increased understanding of the actual execution of processes and leveraging that knowledge for process improvements. Those benefits are not limited to an industrial or business context and can also be applied in other domains, such as education [16]. The process perspective on education interprets learning and teaching as an ordered number of activities (such as taking an exam) that are being executed over time (like each semester) [5]. The term *Educational Process Mining (EPM)* [9,15] covers the analysis of execution data with process mining techniques in the education domain. As the limited number of case studies and publications on process mining on student trajectory data [5] shows, the potential of process mining for improving university teaching has rarely been unlocked.

In this paper, we add a case study to the EPM field that has been conducted at a German university. In contrast to former case studies, our initiative is guided by the Process Mining Project Management methodology [7] (PM<sup>2</sup>). Our main contributions are two-fold: (i) we document our lessons learned and motivate potential refinements to cater for the application of PM<sup>2</sup> to student data; and (ii) through the case study, we obtain insights which, in some instances and for some contexts, indicate a rather clear value of educational process mining.

In the remainder, we elaborate on preliminary work in the field (Sect. 2), present our research methodology (Sect. 3) and execute the methodology based on a case study at a German university (Sect. 4). Sect. 5 sums up our lessons learned, before Sect. 6 concludes.

## 2 Related Work

Data analysis in education is widely applied and has drawn research attention in the past decades under the umbrella term Educational Data Mining (EDM) [2,13]. The process perspective on courses taken by students is covered in the subfield Educational Process Mining (EPM) [16,9,15], which by itself spans a spectrum of application areas. In our paper, however, the focus lies on curriculum mining – a term covering course-based analysis of student trajectories in

<sup>3</sup> See source in footnote 2, “Some college, no degree”, and <https://www.statista.com/statistics/235406/undergraduate-enrollment-in-us-universities/> for current enrollment numbers; accessed 2022-08-22.

educational institutions [5]. In its original meaning, curriculum mining described the pattern-based discovery of a curriculum from observed study data and its subsequent usage for process mining [16,11]. From this origin, authors picked up on the term curriculum mining and extended it, adding comparisons between individual curricula of successful and less successful students, while “success” was measured in course or program grading [18]. Due to the extension of the term curriculum mining, earlier case studies may be subsumed under the term in retrospect such as [17]. [10, Ch.8] explored student’s paths with a focus on the impact of failure on their trajectories. [18] furthermore proposed a process-based recommender system for students to provide guidance when choosing courses, which they published in 2018 [19]. Other initiatives aimed at creating specific techniques for curriculum mining. [3] contributed an approach for conformance checking multiple events with the same timestamps (e.g., courses taken in the same semester). However, they excluded student’s option to retake courses (e.g., after failing them) in their method, leaving room for improvement. Notably, some recent attempts to apply EPM on student curricula lagged behind expectations due to domain-specific challenges [14] that in part remained unsolved as yet. One of the related case studies [6] (in German) presents a method for a curriculum mining initiative, although that method focuses on gaining insights from three specific techniques: bubble-chart-analysis, fuzzy mining and inductive mining; disregarding established process mining methodologies such as PM<sup>2</sup> [7].

### 3 Research Method and Case

We employ a case study-based method to explore how student trajectories can be analyzed in a process mining project. A case study, as a qualitative research method, analyzes a phenomenon in its natural setting to gain insights into emerging topics [12, Ch.5]. With this work, we address the research question: *Does curriculum-based EPM with PM<sup>2</sup> provide benefits for the study program analysis and improvement?* In the process, we explore the arising challenges.

The case we selected deals with the study progress of Bachelor of Science students in Information Systems Management (B.Sc. ISM) at Technische Universität Berlin (TUB). B.Sc. ISM is the third-largest study program of the department for Electrical Engineering and Computer Science comprising 985 students (winter term 2021/22). The department provides a study plan which recommends taking certain courses in certain semesters. We chose this case study, due to our high influence on the study process (one of the authors of this paper serves as program director<sup>4</sup>) and our domain expertise as we teach in the program. In that respect, our team encompassed all roles for a process mining project. Next, the case and study design are described.

*Case.* TUB has around 34.000 enrolled students (winter term 2021/22) and 335 professors. The program B.Sc. ISM has existed since 2013. Currently, 717 B.Sc.

<sup>4</sup> [https://www.eecs.tu-berlin.de/menue/studium\\_und\\_lehre/studiengaenge/wirtschaftsinformatik\\_information\\_systems\\_management/beratung\\_und\\_service/studiengangsbeauftragter/](https://www.eecs.tu-berlin.de/menue/studium_und_lehre/studiengaenge/wirtschaftsinformatik_information_systems_management/beratung_und_service/studiengangsbeauftragter/), accessed 22-08-2022

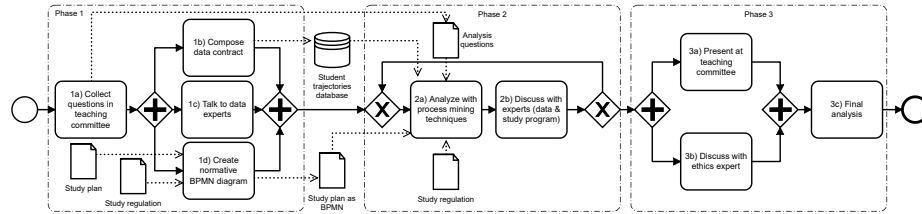


Fig. 1: BPMN diagram representing the case study design.

ISM students are studying based on the study regulation from 2015, and 268 B.Sc. ISM students are studying based on the study regulation from 2021. To increase the data quality for the case study, we focus on trajectories of the 696 students that started their program with or after the effective date of the 2015 study regulation (01-10-2017). (Bi-)Yearly study program meetings are a central element of the teaching quality management at TUB. In such meetings, the study program director, lecturers, and students discuss development and improvement options for the program based on experiences, teaching evaluations, and a high-level analysis of student trajectory data (e.g., cohort analysis of students in the program, drop-out rates, etc.). The controlling department of TUB collects student trajectory data of all study programs. Process mining for a more detailed analysis of student trajectory data was applied for the first time at TUB in this case study.

*Case Study Design.* The BPMN diagram in Fig. 1 provides an overview of our main case study that consisted of (1) a *preparation phase*, (2) a *process mining analysis*, and (3) a *finalization phase*. We prepared the case study in two teaching committee meetings, in which we gathered questions (1a) for a (process-centered) analysis of student trajectory data. Next, we composed a data contract (1b) between the controlling department and our research group to define the appropriation for the data usage, required data attributes, data storage means with respect to data security, and privacy protection mechanisms. This step was supported by the privacy protection department of the TUB. As an output, we received a defined subset of the student trajectory database for doing the process mining analysis. In parallel, we talked to the system experts (1c) from the controlling department regularly to understand the different data tables and attributes of the database and their quality issues. Finally, we also reviewed study program regulations as well as the study plan, and (1d) created a normative BPMN process diagram based on the study plan. After these preparation steps, we started with the analysis (2a). The first step was the creation of an event log from the database tables that suited a process representation of a student curriculum and qualified for answering the analysis questions. After mining and analysis of the event log, the results were discussed with the data and study program experts (2b). In the final phase, the resulting analysis was presented at the meeting of the teaching committee (3a). Feedback and further questions were collected. In parallel, the results and further ideas for analysis were presented to

an ethics expert (3b) to discuss how and in which way the results should be used. Based on these insights, we finalized our analysis of the student trajectories (3c), which is presented in the following section.

## 4 Applying PM<sup>2</sup> for Educational Process Mining

The PM<sup>2</sup> [7] is the defacto standard process mining methodology and consists of six main phases: (1) project planning, (2) extraction and (3) data processing of the event log, (4) mining and analysis, (5) evaluation and (6) process improvement & support. In the following, we describe them for our EPM case.

### 4.1 Planning

The process under scrutiny was the trajectory of students through the university course system. The goal of the project was 1) to learn about student’s paths through the university system, and 2) to find deviations from recommended path (the study plan) through the university system provided by the department.

*Selecting business processes.* As described in Sect. 3, the program B.Sc. ISM was well known to the members of the project team. Also, the project had lively support of the program director, which provided a lever on changeability of the process. Additionally, availability of the necessary data in the university’s information systems was given, following the signing of the data contract.

*Identifying research questions.* To avoid confusion with the research question in this paper, we refer to research questions of the project as “project question (PQ)”. The mentioned study plan recommends a certain order of courses to Bachelor students. The teaching committee did not yet have information about:

**PQ1** Do B.Sc. ISM students follow the study plan?

**PQ2** Are students that follow the study plan successful in their studies?

**PQ3** Which behavioral patterns can be observed in the data?

*Composing the project team.* The project team comprised four people with the following roles: the study program manager (process owner), three lecturers in the study program (business experts), a member of the controlling team with experience of data warehousing at TUB (system expert) and three process analysts. Additionally, we frequently sought feedback from TUB’s data protection unit about the use of the data. We also incorporated ethical advice for result exploitation.

### 4.2 Extraction

For data extraction, we had different sources of input. On the one hand, the university operates partially self-deployed information systems building on off-the-shelf student lifecycle management systems holding all available data about students’ study progress. On the other hand, several documents officially issued

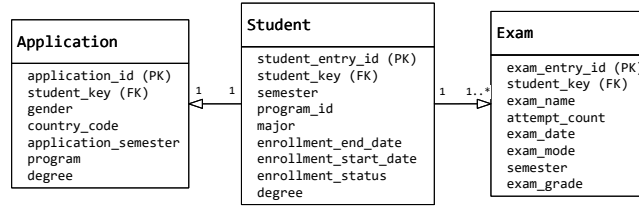


Fig. 2: Data scheme showing a subset of available data attributes. Note that the tables were not normalized (PK ... Primary key; FK ... Foreign key).

by the university contained descriptions of the process students go through at TUB.

*Determining the scope.* With respect to the project questions, two levels of granularity were required: 1) exams taken by the students on specific days during a semester, and 2) the semester the students were enrolled in, including the status of the semester (actively studying, sabbatical, exmatriculated, etc.). The relevant time-frame for B.Sc. ISM students of the most recent long-running version of the program was from September 2017 to January 2021. We were provided with a data schema covering student-related data attributes. An abbreviated version is depicted in Figure 2. Note that student’s study data are considered *personal data* according to the European Union’s General Data Protection Regulation and thus have to be handled with additional care (GDPR Art. 4). To sustain the ability to assign courses to individual students, we created pseudonyms for student ID numbers to which courses were assigned. Additionally, the data was delivered using an encrypted and secure data transfer mechanism.

*Extracting event data.* The data was delivered as a multi-table SQL dump. The tables, as we received them, were not normalized. As notion of an instance, we selected the student with the pseudonym of the matriculation number (`student_key`) and selected some case-specific attributes from the student table. The main two extracted event types are the (re-)enrollment of a student for a semester and their exams taken for which we also consider event-specific attributes (e.g., timestamp, name of the exam, (non-)compulsory exam).

Fundamentally, we joined the tables `Student` and `Exam` based on the attribute `student_key` to generate the event log. Note that we did not make use of the data in the table `Application`. Including the following additional preparatory steps to transform the data, the SQL commands for the event log generation amounted to 150 lines of code: 1) filtering for time frame the program B.Sc. ISM; 2) unifying event labels; 3) flagging compulsory and elective courses; 4) flagging events by their type; 5) initial filters, e.g. for the study program and its version; 6) creating and re-labeling semester re-enrollment events; and 7) adjust the timestamps for discovery of parallelism

*Transferring process knowledge.* We transferred knowledge from written document-bound information issued by TUB that described the study process such as the *General Study and Examination Regulations of TUB* (university-wide). For the

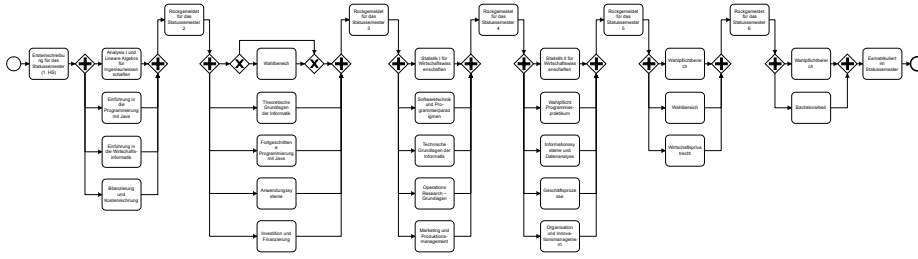


Fig. 3: Normative BPMN diagram capturing the study plan for B.Sc. ISM.

project, we created a normative BPMN diagram shown in Fig. 3 representing the study plan. We had to take a few aspects into consideration during modeling. First, how can we model the structure of semesters? To this end, the first activity in the process corresponds to initial enrollment (i.e., the registration for the first semester). This is followed by the activities for the first four modules in parallel (between a parallel split and a parallel join). Subsequently comes the activity to re-enroll for the second semester. In the parallel construct that follows, the top branch of the process has an optional activity, labeled “Wahlbereich”, i.e., free electives, while the other modules are compulsory. The rest of the process follows the same logic.

### 4.3 Data Processing

*Creating views.* To create meaningful views on the event data for the purposes of our project (analyzing course sequences taken by students) we chose student ID (pseudonyms) as a case notion. That way, for each student their exams as well as their status for a semester were aligned as a series of events.

*Aggregating events.* One major challenge of the event log are the two levels of granularity in the events: exams taken and semesters studied, while multiple exams can be taken in one semester. Since the study plans suggested an order of taking exams, there was a de facto part-of relation between the exams and the semesters. We thus aggregated the exams to the semester level for parts of the analysis.

*Enriching logs.* For this project, answering the project questions did not require enriching the log with additional information.

*Filtering logs.* The applied filters concerned various attributes. Most frequently, we filtered for particular cohorts of students (e.g., only students that enrolled in a specific semester) or particular courses (e.g., to learn about the order in which two exams were taken). Lastly, in April 2021, TUB was the target of an IT hack that led to multiple services being interrupted, including the exam database. Hence, most exams taken since could not be tracked as complete entries and were available to us as “unknown”, which caused a data quality issue for this case study. The “unknown” events were filtered out in almost all analyses, except for investigating the count of exams over the observed time period.



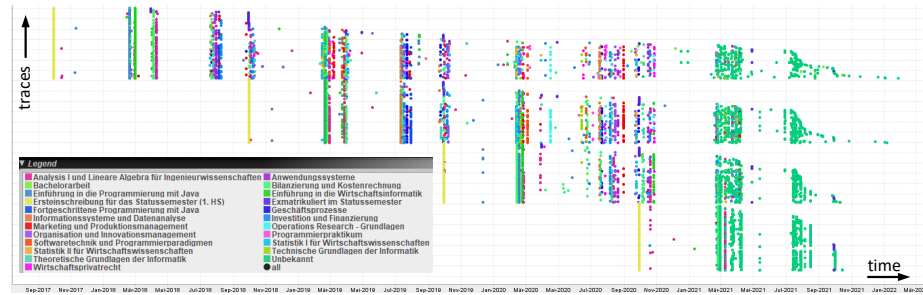


Fig. 4: High-level view on the event data in a dotted chart.

#### 4.4 Mining and Analysis

Mining and Analysis started with applying process discovery techniques to answer PQ2 and PQ3. PQ1 was approached with conformance checking.

**Process discovery** To gain an overview of the data and its time distribution, we generated a dotted chart<sup>5</sup> from the filtered data, which is shown in Fig. 4. The different cohorts of students starting in the five years appear clearly separated, with the yellow dots for matriculation forming basically solid lines in October of the respective years. Most students attend classes for the first semester before sitting their first exam, but interestingly some students do not; these might have transferred from a different university or program, or might have prior knowledge that allows them to sit an exam. The dots corresponding to some exams form almost solid lines, like the ones shown in blue, green, and pink after the first semester – e.g., in February–April 2018 – indicating that the corresponding cohorts take these exams in unison.

Very noticeable is also the decline of participation in exams over the years (e.g., following the first cohort at the top of Fig. 4): while almost all students sit some exam after the first and second semester, the number of exam events is subject to a sharp decline from the second to the third semester and slowly drops further. To understand this decline, it is worth noting that German universities often have a low bar for matriculation into a study program – first-semester students are then admitted without a test to measure their qualification for a particular program. Other students might have started a program without a solid understanding about its content, and decide that they prefer a different one. For these or similar reasons, it is rather common that about one third to half of the students drop out of a study program without a degree. One concrete question in our case was, whether students attempted (or passed) the advanced programming course (“Fortgeschrittene Programmierung mit Java”) before the introductory programming course (“Einführung in die Programmierung mit Java”). To exam-

<sup>5</sup> A number of events highlighted in green appears in the observed traces as of spring 2021 across cohorts. These are the “unknown” events that resulted from the TUB’s IT being compromised. Exams that are “unknown” were filtered out in all analysis (as described above), but included in the dotted chart for illustration purposes.

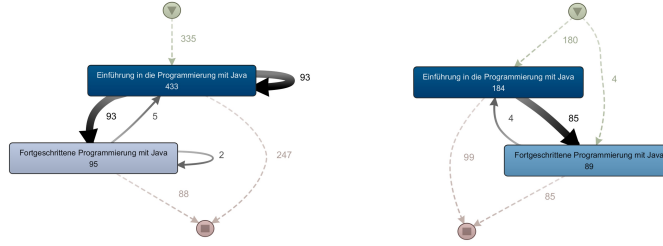


Fig. 5: Order of programming courses (left: all attempts; right: passed attempts).

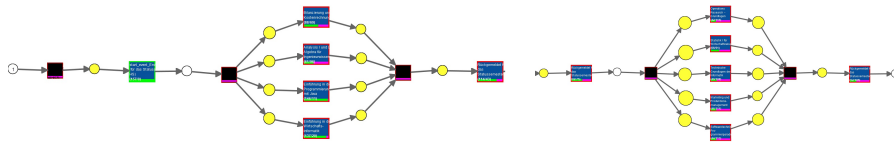


Fig. 6: Conformance in semester 1

Fig. 7: Conformance in semester 3

ine this, we filtered the DFG of individual exam attempts to only show these two exams – depicted in Fig. 5 in terms of all attempts (left) and passed attempts (right). Clearly the answer is “no”: 93 (85) times the introductory course is attempted (passed) before the advanced one, and only 5 (4) times the opposite is true. This is one of the more concrete questions that we have collected for the B.Sc. ISM program.

**Conformance checking** For evaluating the question: “To which degree do students follow the study plans of their programs?” conformance checking can be applied. As the variance in non-compulsory modules is high, we projected the process model (see Fig. 3) and the log to only re-enrollments and compulsory exams. We focused only on the first cohort of students that matriculated in September 2017 and had a chance to finish their studies (participate in all courses) to avoid result distortions from non-finalized cases of early semester students. With the help of ProM<sup>6</sup> 6.9, we converted the normative BPMN diagram into a Petri net and then, used it for the alignment-based conformance checking. The resulting Petri net’s layout is very long, hence we only show results for the first and the third semester, in Figures 6 and 7, respectively. Clearly observable is that the percentage of adherence (shown by the green bar at the bottom of a transition in the net) to recommended modules (i) varies somewhat within a semester, and (ii) decreases considerably from the first to the third semester. This trend continues further after the third semester, though at varying speeds as observed from the dotted chart above. In addition, we obtain statistics for the entire log: The average trace fitness was approx. 29% – for the compulsory modules. This indicates that students make ample use of the freedom provided by the study regulation at TUB.

<sup>6</sup> <https://www.promtools.org/doku.php>, accessed 2022-08-22

#### 4.5 Evaluation

For the diagnosis as well as the verification, we conferred with the teaching committee yet again. Most of the results of the study confirmed speculations, in particular the decreasing conformance of student trajectories with the study plan. Given the novelty of the study results for the committee, the meeting was largely used to define further research questions, e.g.: How can the findings be incorporated in study plans? Can process mining techniques be used to predict drop-outs? What are ethical ways to communicate potential prediction findings to students?

#### 4.6 Process improvement and support

The implementation of this phase is still in progress. For process improvements, we identified two areas to take action: 1) the study plan that could be adapted following the findings, 2) communicating results of the project to students and providing them an outside perspective on how they advanced in their studies. In both scenarios, ethical considerations ought to play a crucial role: e.g., nudging students away from using their freedom of choice in courses is disputable. Supporting operations, i.e., including process mining in the stable set of analysis tools to inform the university directors and teaching committees, was at the heart of the case study and is ongoing in collaboration with the controlling department.

### 5 Lessons Learned

In this section, we report on our lessons learned from the case study in the different phases of the PM<sup>2</sup> methodology.

**Planning.** Our *research questions* such as “Do B.Sc. ISM students follow the study plan?” were defined based on the related work (in particular [11] and [10, Ch.8]) and the discussion in the study program meeting. More specific questions can be added for each study program. Besides, data on students’ studies are considered as *personal data* that need to be handled with special care according to the GDPR. We worked with the TUB data privacy department for consultancy in this regard and recommend allocating time for these matters. *Defining student success* over the course of the whole degree is a non-obvious question. We could distinguish levels of success by the GPA, like [19], or overall grade, but might run the risk of overly simplifying.

**Extraction.** We showed a first approach to define a *normative process* diagram in BPMN representing the study plan for a program and the challenges thereby.

**Data processing.** Due to experiences with *existing regulations and new requirements*, study regulations are regularly changing (also see [11,14]). This leads to different student cohorts that should be analyzed independently (also see [17]). It is challenging to define consistent student cohorts for different event log views. In this study, we filtered for time-frame in which the regulation was valid. Additionally, timestamps of events of student trajectories are usually on different

levels of granularity, discussed as one of the *timestamp imperfections* of event logs by [8]. Whereas the information about the re-registration for a semester and its status is on a semester level, exams are captured finer-grained on a day-level. Also, German universities offer a very high degree of flexibility in (compulsory) *electives*. Based on our observations, for most analyses it makes sense to aggregate such courses.

**Mining and analysis.** In this phase we encountered several challenges. To many of them the obvious solution would have been to apply strong event filters to reduce noise in the log and simplify the analysis, coming at the cost of reducing the expressiveness of the log. These challenges include handling *exam repetitions* after failing (also see [3,14]), *semesters abroad* students spend trying to recognize modules as equivalent to offerings at their home university, *lateral program entry* by students switching programs at various points (also see [14]), and *handling non-finalized cases* which poses a particular challenge in conformance checking.

**Evaluation.** Students of TUB have access to a wide variety of (compulsory) electives in their B.Sc. studies. This results in many *snow flakes*, i.e., unique traces. As above, aggregation might be necessary to be able to obtain insights from this data, since the raw data might be too fragmented.

**Process improvement and support.** *Communicating* individual data analysis results back to students may cause unease, or be stressful to students. Additionally, purely data-based results do not include alternative cause of action yet. Opt-in options or coupling the communication with consultation offers should be considered. Lastly, the results of the case study analysis may nudge students to follow very specific succession of courses, which is opposing the *Humboldtian ideal*<sup>7</sup> of freedom of study for students – one of the basic principles of the German university system. Hence, there is an area of conflict between educational ideals and efficiency goals of EPM projects that need to be debated.

## 6 Conclusion

In this paper, we study how to apply process mining to curriculum-based study data. Our case encompasses a major Bachelor program at TU Berlin, a leading university in Germany. Sect. 4 describes our approach using the PM<sup>2</sup> methodology. We were able to answer concrete questions about student behavior and adherence to the study plan; as such, with regards to our research question we observe an indication that indeed curriculum-based EPM can provide insights of some value – e.g., checking if contents of succeeding lectures are coordinated in a way that lets students advance, or potential to translate student trajectories into curriculum recommendations which we expect to reduce study time and eventually (personal and social) costs. In Sect. 5 we reflected on specific challenges encountered during our study. If similar challenges appear in coming curricular EPM case studies, we suspect future domain-specific alterations of PM<sup>2</sup> might be justified considering its objective to “overcome common challenges” [7]. Given

<sup>7</sup> [https://en.wikipedia.org/wiki/Humboldtian\\_model\\_of\\_higher\\_education](https://en.wikipedia.org/wiki/Humboldtian_model_of_higher_education), accessed 2022-08-22

the research method, case study, we inherit typical threats to validity of such research [12, p.125], specifically threats of limited replicability and generalizability. The analysis team was also responsible for data preparation, which might have introduced bias.

**Acknowledgments.** We are very thankful for the input of Patrick Russ, data analyst at TUB's controlling department.

## References

1. Van der Aalst, W.M.: Process mining: data science in action. Springer (2016)
2. Baker, R.S., Inventado, P.S.: Educational data mining and learning analytics. In: Learning analytics, pp. 61–75. Springer (2014)
3. Bendatu, L.Y., Yahya, B.N.: Sequence matching analysis for curriculum development. *Jurnal Teknik Industri* **17**(1), 47–52 (2015)
4. Berland, M., Baker, R.S., Blikstein, P.: Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning* **19**(1), 205–220 (2014)
5. Bogarín, A., Cerezo, R., Romero, C.: A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery* **8**(1), e1230 (2018)
6. Buck-Emden, R., Dahmann, F.D.: Zur Auswertung von Studienverläufen mit Process-Mining-Techniken. Technical Report, Hochschule Bonn-Rhein-Sieg (2017)
7. Eck, M.L.v., Lu, X., Leemans, S.J., Van Der Aalst, W.M.: PM<sup>2</sup>: A process mining project methodology. In: CAISE. pp. 297–313. Springer (2015)
8. Fischer, D.A., Goel, K., Andrews, R., van Dun, C.G.J., Wynn, M.T., Röglinger, M.: Enhancing event log quality: Detecting and quantifying timestamp imperfections. In: BPM. pp. 309–326. Springer (2020)
9. Ghazal, M.A., Ibrahim, O., Salama, M.A.: Educational process mining: A systematic literature review. In: EECS Conference. pp. 198–203 (2017)
10. Janssenswillen, G.: Unearthing the Real Process Behind the Event Data. Springer (2021)
11. Pechenizkiy, M., Trcka, N., De Bra, P., Toledo, P.: Currim : Curriculum mining. In: Intl. Conference on Educational Data Mining, EDM. pp. 216–217 (2012)
12. Recker, J.: Scientific research in information systems: a beginner's guide. Springer, 2nd edn. (2021)
13. Romero, C., Ventura, S.: Educational data mining and learning analytics: An updated survey. *Wiley Interd. Rev.: Data Min. and Knowl. Disc.* **10**(3), e1355 (2020)
14. Schulte, J., Fernandez de Mendonca, P., Martinez-Maldonado, R., Buckingham Shum, S.: Large scale predictive process mining and analytics of university degree course data. In: Intl. Learning Analytics & Knowl. Conf. pp. 538–539 (2017)
15. Sypsas, A., Kalles, D.: Reviewing process mining applications and techniques in education. *Intl. J. Artificial Intelligence & Applications* **13**, 83–102 (2022)
16. Trcka, N., Pechenizkiy, M.: From local patterns to global models: Towards domain driven educational process mining. In: Intl. Conf. Intelligent Systems Design and Applications. pp. 1114–1119. IEEE (2009)
17. Trcka, N., Pechenizkiy, M., van der Aalst, W.: Process mining from educational data. *Handbook of educational data mining* pp. 123–142 (2010)
18. Wang, R., Zaïane, O.R.: Discovering process in curriculum data to provide recommendation. In: EDM. pp. 580–581 (2015)
19. Wang, R., Zaïane, O.R.: Sequence-based approaches to course recommender systems. In: Intl. Conf. Database and Expert Systems Applications. pp. 35–50 (2018)