Improving PhD Student Journeys with Process Mining: Insights from a Higher Education Institution

Kanika Goel[1][0000−0002−6250−2589], Sander J. J. Leemans[1][0000−0002−5201−7125], Moe T. Wynn[1][0000−0002−7205−8821], Arthur H. M. ter Hofstede[1][0000−0002−2730−0201], and Janne Barnes[2]

[1] School of Information Systems, Queensland University of Technology, Brisbane, Australia
{k.goel, s.leemans, m.wynn, a.terhofstede}@qut.edu.au
[2] Business Change Manager, Queensland University of Technology, Brisbane, Australia
janne.barnes@live.com.au

Abstract. The socioeconomic consequences of not successfully completing PhD studies have motivated universities to expend dedicated efforts on improving student journeys. These journeys leave traces in a variety of university IT systems and this trace data can be exploited to derive insights through the application of process mining. Process mining is a form of data-driven process analytics, where process data, collated from different IT systems, is analysed to uncover the real behaviour and performance of processes. Despite its potential application, process mining hitherto has not been applied to visualise, analyse, and improve PhD student journeys, to the best of our knowledge. This paper reports on the findings of a process mining case study conducted at an Australian University that had espoused a digital transformation initiative to improve PhD student journeys. The case study utilised interactive and comparative process mining techniques and focused on clarifying the way a PhD student journey eventuates, visualising the differences between the real (actual) and prescribed (recommended) processes, comparing the performance of different cohorts, identifying root causes for adverse outcomes, and providing evidence-based recommendations for the digital transformation initiative. The findings from this study resulted in restructuring of HDR services and the introduction of a new research management system.

Keywords: Process Mining · Digital Transformation · Higher Education · Process Improvement.

1 Introduction

Technology is transformative. Throughout the world, higher education is undergoing digital transformation [24], which is a result of increasing competition among universities and affordance of digital technologies [13]. Queensland University of Technology (QUT), situated in Australia, undertook one such digital transformation project that aimed to restructure the university's Higher Degree Research (HDR) services and replace core research management systems. To do so, an objective was to understand and improve PhD (a HDR degree) student journeys, i.e., the various steps PhD students take throughout their degrees as reflected in their interactions with university systems.

Analysing PhD journeys and associated processes of QUT was expected to contribute to better design of the new research management system and timely completion of PhD.

Completion rates of PhD programs at universities have been a long-standing concern for national governments [5,22,8] and a pressing issue for higher education. While there are many facets to the complex problem of timely PhD completions the focus in this paper is PhD student *journeys*. A student journey can be seen as a *process* as it involves a number of well-defined steps with inter-dependencies and documents which are needed as input or are produced by these steps.

Taking this process lens allows one to leverage methods and techniques from the discipline of process mining. Process mining is a specialised form of data-driven process analytics, which enables the extraction of detailed insights regarding process behaviour, process performance, conformance of processes to existing process models, and process improvement opportunities from event logs [1]. Process mining can thus be used to understand complex unstructured PhD journeys and assist in identifying unnecessary variations, reasons for delay, and complexity drivers. A process-oriented view also helps to transcend the isolated view of data collections that dominates traditional data mining techniques [4] prevalent in higher education.

Some initial applications of process mining in higher education can be found in literature [4,6]. While these applications show the potential of process mining, the full breadth of process mining techniques, such as, the interactive discovery and conformance checking of process models, comparative process mining, and the identification of root causes has not been explored in depth.

In this paper, we presents the insights from a process mining case study conducted at QUT over a period of six months[3]. The case study shows (1) *how a range of process mining techniques can be applied to better understand PhD student journeys* and (2) *how this understanding can be translated into concrete recommendations for process improvement and changes to systems at a higher education institution* .

2   Related work

Universities today rely heavily on IT systems to support their processes. These systems record huge amounts of data which can be studied to reveal valuable information to improve these processes. To this end, various data-driven analysis approaches have been applied. These can be subdivided into data mining and process mining techniques.

Several data-mining techniques have been used to address questions similar to those addressed in this paper. Regression has been used to study the influence of variables on one another [11], for instance, to study the relation between the amount of leave taken on time taken to complete a degree. Such techniques are able to identify high-level correlations between context factors and outcome variables, however, they are not able to provide views of the process/journey that led to these outcomes. Sequential pattern mining studies the occurrence of events in a sequence in order to find statistically relevant patterns and to predict student outcomes [15,14]. While such techniques can analyse sequences of events in a student journey, they do not provide complete end-to-end process views and abstract away choices and parallelism. Consequently, such

---

[3] Janne Barnes was the business change manager at QUT during this period.

techniques are unable to identify deviations, loops, and infrequent behaviour which is relevant to answer some process-related questions [1].

Process mining has been applied in the context of online learning environments such as Massive Open Online Courses (MOOCs). For example, [2] analysed log data of student interactions with online course materials, clustered cases to derive accurate process models, and compare different cohorts. [20] used process mining to derive student learning workflows from virtual learning environment logs with decision trees capturing the rules that control students' adaptive learning. [19] proposed a process mining approach that can predict student dropouts in MOOCs at an early stage.

Furthermore, process mining has been employed to mine curricula providing insights for university administrators and course coordinators about the different units a student may choose throughout their course ([21], [18]). [7] developed a foundational learning analytics model for higher education, which they claim can be used to provide personalised learning and support services to students but to the best of our knowledge has not been applied to improve student journeys.

3   Introduction to the Case Study

In Australia, timely completion is reinforced through PhD scholarship rules and the Australian Quality Framework (AQF) that explicitly state the term of PhD studies as 3-4 years. The PhD program at QUT is expected to be completed within a three-year period. During these three years, students need to complete three major milestones at which the scope, plan and progress of the student is scrutinised: Stage 2, the Confirmation seminar, and the Final seminar. A student may withdraw from the degree at any point which terminates the PhD journey. Independent of the three milestones, the student may take periods of leave, apply for extensions, complete annual progress reports, which are all done using online e-forms. These e-forms assist with monitoring the progress of a PhD student thus contributing to timely completion of student journeys. The varied number of activities involved, inherent uncertainties of research, and the multi-year duration make the PhD journey a complex process. As a result, the PhD journey is a much less unstructured process although the milestones are clearly defined.

The digital transformation program aimed to improve student supervisor experience by restructuring HDR services around these experiences. Another change imperative was the centralisation and standardisation of services for HDR management to create economies of scale. The project was primarily a process improvement initiative at the enterprise level, with the overarching goal of more efficient research management and enriched PhD journey experiences.

4   Actions Taken

Our process mining study consisted of four phases: identification of analysis questions, data extraction and pre-processing, process mining analysis, and interpretation of results and provision of recommendations.

Identification of analysis questions. The following six questions were identified:

AQ1 Assess the quality of time/event data for undertaking process mining analysis of a PhD student journey;

AQ2 Discover the different behaviours exhibited by students to better understand their journey;

AQ3 Compare the actual and prescribed process of the PhD student journey to identify deviations;

AQ4 Conduct a detailed performance analysis of the PhD student journey, particularly in regards to completion time;

AQ5 Identify root causes and patterns in the process models which could be used as risk predictors for withdrawal from PhD programs;

AQ6 Identify ways to improve the processing of e-forms to facilitate the student journey.

Data extraction and pre-processing. A student journey dataset which covered all the PhD students enrolled in the past 16 years (2002 to 2018) was analysed. In addition, seven data sets each relating to one of the *seven e-forms* covering a period of 2016 to 2018 were analysed. Both datasets were complemented by a data set with demographic factors such as: faculty (eight faculties in total), type of study (full-time, part-time), mode of study (domestic, international), gender (male, female), and degree (various HDR degrees). Datasets had to be merged to obtain logs appropriate for analysis. Following this, a significant amount of effort was spent on data cleaning and pre-processing (AQ1). The logs were cleaned to ensure consistency of activity labels across datasets. Events without complete timestamps were filtered as they were marked as active cases and were not desired to be included in the analysis. In discussion with stakeholders, logs were filtered for full-time students who completed their PhD (1139 cases with 13056 events) and who withdrew (498 cases with 7558 events). Furthermore, sublogs for each milestone (Stage 2, Confirmation, and Final Seminar) were created to conduct an in-depth analysis.

Process mining analysis. Once the data sets were retrieved, cleaned and pre-processed, they were subjected to analysis. The main tools we used for analysis were: multiple ProM plug-ins (in particular the Inductive visual Miner (IvM) [12] and ProcessProfiler3D [23]), DISCO [4], and SQL Server. IvM was used to automatically discover and manually build process models. ProcessProfiler3D was used to visualise and compare the performance of different cohorts. DISCO was used for its elaborate filtering features, and SQL server was used for cleaning data as well as aggregate calculations.

Interpretation of results and provision of recommendations. During our analysis, we frequently sought feedback on intermediate results from the stakeholders, rather than sharing the results at the end. In addition to regular informal meetings, we presented three times to the stakeholders. This way we gained a deeper understanding of the intricacies of the domain and were able to adjust our next steps to increase the accuracy of our analysis. This was essential given the complexity of the PhD journeys and the data quality issues we encountered. Close collaboration with stakeholders also assisted with live conformance checking. For instance, we manually built the models of some processes with the assistance of the IvM in collaboration with the stakeholders to visualise deviations from prescribed models, rules, and regulations.

---

[4] `https://fluxicon.com/disco/`

5   Results Achieved

5.1   Discovery of the PhD Student Journey (AQ2)

Given the complexity of the PhD journey, the stakeholders were interested in *visualising* the key activities undertaken in the PhD journey. The stakeholders were also interested to know about the distribution of leave in the PhD journey as leave has been identified as a potential risk factor for delays in PhD completion [17].

To discover process models, an automated process discovery technique, IvM, was used. However, automatic process discovery algorithms tend to overgeneralise the presence of recurring activities, i.e., show excessive behaviour that does not reveal much about the actual behaviour of processes. For example, the automatically discovered model showed leave (a recurring activity) in parallel to a milestone with a frequency of 250. The frequency indicates that the total number of leaves taken are 250, however, it gives little information about when the leaves were taken in the journey, which the stakeholders were interested in. Therefore, we manually adapted the automatically derived models to better fit the needs of the stakeholders. E.g., we adapted the process model to display activities for the different types of leave taken just before and after major milestones. Figure 1 shows this adaption for the final seminar where different types of leave such as maternity, sick and other approved leave are taken prior to the final seminar. The adapted models showed key events, such as students being placed under review, students resubmitting documents, and students taking leave, and how often these occurred. The models also displayed *when* in the journey these events happened as well as their occurrence frequency. The resulting models also revealed both, frequency and duration of various types of leave.

5.2   Deviations From Expected Student Journey (AQ3)

The stakeholders wished to find out whether Annual Progress Reports (APR) were only submitted after Confirmation. To this end, we constructed two process models to see if there were students completing an APR before Stage 2 or between Stage 2 and Confirmation. Interestingly, we found that 308 students submitted APRs before Stage 2, which surprised the stakeholders and provided an example of unnecessary utilisation of resources. *This discovery resulted in revision of the APR workflow, with the first APR request initiated six months after confirmation.*

In addition to constructing models, we also visualised deviations from frequent paths observed in the log using Inductive visual Miner. Figure 1 shows an example of such deviations (in red). It was interesting to discover that eight students skipped Final Seminar yet completed their PhD. This visualisation also helped to identify infrequent behaviour, which assisted with understanding the different trajectories PhD students have taken in the past.

5.3   Performance analysis of the PhD student journey (AQ4)

PhD completion times are a concern for both individuals and universities due to significant economic and psychological costs. At present, QUT expects a PhD journey to be
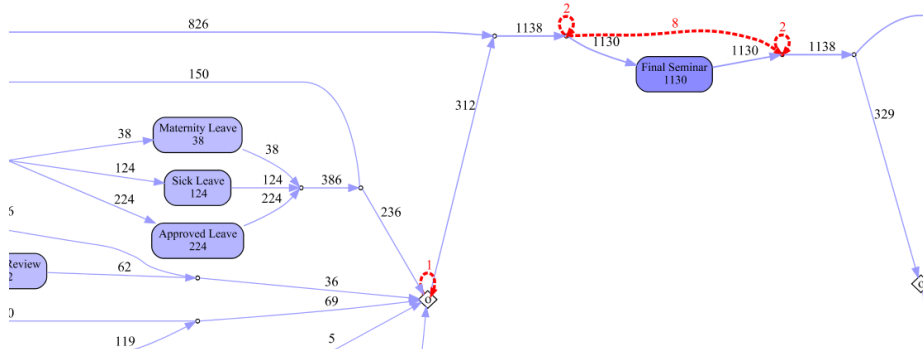
Fig. 1: Process model showing deviations in the PhD journey of completed students.

completed in three years. However, during their PhD journey a student may take periods of leave which extend the duration of the program.

We found that only 4.65% of the students complete their PhD within three years. 55% of the students completed within four years and 82% complete within four and a half years. These results surprised the stakeholders as they did not expect such a small percentage of the students to complete within three years. Longer completion times cause additional costs to the university due to higher resource utilisation. Further discussions with the stakeholders revealed that they considered durations from *enrolment* to *lodgement of thesis*, while in fact the actual duration of a PhD journey is better viewed as starting at *enrolment* and ending at *submission of thesis after addressing external examiners comments* as students still use university resources such as printer, desk, laptop etc. after thesis lodgement. *This insight was met with a positive response from the stakeholders and resulted in a new definition of the duration of a PhD journey as well as calculation of the associated costs*.

We also explored variations in performance of the students across demographic factors present in the log, such as gender, faculty, and more. We used ProcessProfiler3D for comparative process mining, to showcase multiple cohorts for each factor (e.g., different faculties) in one view. Figure 2 shows a high-level process model in ProcessProfiler3D showing performance measures of student cohorts grouped by faculty. We found no variations in performance for PhD students across the demographic factors, an interesting finding in itself for the stakeholders, as it implies that *no special attention needs to be paid to a particular cohort and no cohort-specific reforms are required*.

5.4   The Journey of Students that withdrew from the degree (AQ5)

Another question of interest was when students withdraw during the journey and why, i.e., what the factors are that influence the students' withdrawal from the course. To answer this question, we discovered a process model for each phase of the PhD journey (each corresponding to one of the three milestones).

We found that 35 out of 280 students withdrew just after starting the course. Another 31 students took some type of leave and then withdrew before Stage 2. Out of 280

Fig. 2: Performance of activities for completed PhD students by faculty. From left to right, the transitions are "Stage 2", "Confirmation" and "Final Seminar"; all sub-processes. Bar widths indicate relative number of occurrence, bar heights indicate mean duration of each activity and bar colour indicates faculty.

students, only 214 submitted their Stage 2 proposal. This finding was interesting to the stakeholders and raised the question why some students decided to withdraw so soon after enrolment without even attempting to achieve their first milestone. Though the actual reason cannot be ascertained from the model, one can think of multiple causes for this attrition, e.g. student-supervisor relationship, financial constraints, health [10].

Our observations also indicate that students who withdraw take more time to finalise their confirmation milestone than students who successfully complete the degree. Furthermore, we found that 50% of the students who withdraw after completing their confirmation withdraw within the next year. This insight highlights the need to ensure ample support and supervision in the first year of a student's PhD [3]. For the other milestones we did not observe major differences between the two student cohorts.

The stakeholders were also interested in a more in-depth analysis of duration and frequency of periods of leave before and after the three major milestones. Six process models were built, one for each milestone-cohort combination. Analysis of all six models shows that students who eventually withdraw from the program take leave more often and for longer periods of time during the journeys to Stage 2 and also from Stage 2 to Confirmation. Figure 3 demonstrates the difference in the frequency and duration of different leave types between completed and withdrawn students during the journey from Stage 2 until Confirmation.

This pattern observed for the withdrawn students was considered to be an early risk indicator of potential withdrawal. *Acknowledging this as an early risk indicator, QUT decided to initiate bi-annual health-check emails to ensure that PhD students have an opportunity to report on any health issues, which can be brought to the attention of the university as well as to the supervisors.*

5.5   Improvement in the processing of e-forms (AQ6)

E-forms are the main mechanism for students to interact with the university processes and are, therefore, an integral component of the PhD journey. Another question raised by the stakeholders was how the processing of e-forms could be improved (AQ6). To address this question, we checked for bottlenecks in the processing of e-forms, again using process models enhanced with performance information. We found that among all performed activities, the activity 'RSC approval' and the activities residing with the Research Student Centre (RSC) take the largest amount of time as is shown by the process model associated with the 'Student Leave' e-form. On sharing the median and average duration of processing e-forms with the stakeholders, it was confirmed that the

(a) Average duration
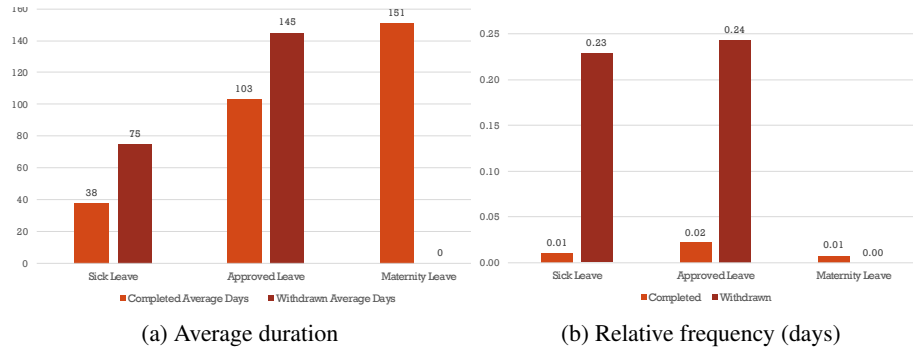

(b) Relative frequency (days)

Fig. 3: Relative frequency and average duration of leave types for different student co-horts during the journey from Stage 2 to Confirmation. The relative frequency is cal-culated as the percentage of students of a cohort that have requested that type of leave. The average duration for a cohort of students is calculated as the sum of periods of leave of all these students divided by the number of students.

time taken is much more than what they expected (note that this was not a given as certain activities may by their very nature take longer). We repeated the process for the remaining e-forms (seven e-forms in total) and found the same bottlenecks for all of them. In addition to this, we found that the students take considerable time filling in the e-forms. *These findings resulted in changes in the e-form workflows (e.g. pre-filling them automatically as much as possible) and a restructuring of the way the RSC processes these forms.*

In terms of the overall time taken to process e-forms, i.e., the case duration, it was found that some e-forms completed within an expected time frame, while others took exceptionally long to complete. Consequently, the stakeholders were interested in iden-tifying the reasons for these long delays. To answer this question, we started by discov-ering process models for each e-form. We found that all process models contain loops, which indicates the presence of rework in the processing of the e-forms.

To investigate these loops further, we abstracted out the rework loops in the process into sub-processes using hierarchical process models. The top level hierarchy showed the key activities of the loop and the bottom level was a simplistic version of the origi-nal process model. This enabled us to focus on those parts of the model that concerned rework and hence retrieve performance measurements for these loops. To get further information, we created and added an extra attribute to our event log, which indicated whether a particular e-form instance was long running or not. An e-form instance was considered to be long-running if it lasted at least 2.5 times the standard deviation longer than the mean duration of processing of e-forms. The resulting log and hierarchical pro-cess model were used as input for ProcessProfiler3D. Unsurprisingly, the visual repre-sentation brings out the striking difference between cohort of long-running cases versus the cohort of cases with an expected duration (see Figure 4). The performance visuali-sation shows that normal cases take much less time for certain activities than the long running cases.
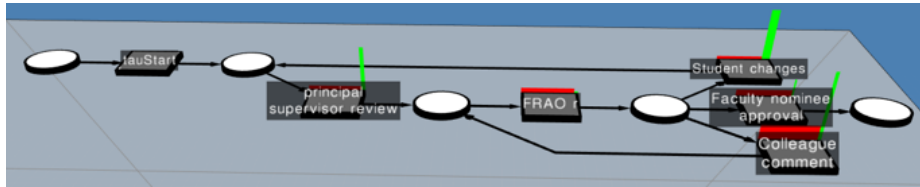
Fig. 4: Thesis submission sub-process. The bars on the activities show duration, split into normal cases (red) and long-running cases (green). From left to right the labels read "tauStart", "principal supervisor review", "FRAO", "Student changes", "Faculty nominee approval", "Colleague comment".

Table 1: Processing times by faculty for student leave e-forms.

Activity	Faculty 1	Faculty 2	Faculty 3	Faculty 4	Faculty 5	Faculty 6
Frequency	87	453	106	121	224	848
Create	0d 01:19	0d 02:40	0d 01:16	0d 00:05	0d 06:01	0d 00:04
Principal supervisor approval	0d 22:46	2d 01:41	2d 05:55	1d 21:36	1d 17:19	2d 01:48
Faculty feedback	n/a	2d 03:00	8d 02:44	2d 04:42	0d 08:52	2d 01:22
Faculty approval	1d 11:11	0d 22:33	1d 21:54	1d 11:43	1d 22:8	1d 8:15

To further understand why in some instances the processing of e-forms takes an exceptionally long time, we used the 'trace visualisation' feature of ProcessProfiler3D. This feature enables the user to visualise the trajectory of process instances through the process model. This time we divided the cases into five cohorts (where four cohorts correspond to the first, second, third, and fourth quartile respectively, and the fifth cohort correspond to the long-running outlier cases) by adding another attribute to the event log. The resulting visualisation, conveyed that long-running cases involve multiple loops resulting in substantially more rework than other cases.

Once activity duration and multiple loops were identified as underlying reasons of delay, we were also interested in investigating if any of the demographic factors were associated with long running cases. We used relevant data mining techniques (contrast set learning [9] and decision tree mining [16]) to uncover the potential influence of the attributes on the duration of cases. We found that none of the demographic factors (faculty, gender, scholarship holder, type of study) had an impact on the processing time of e-forms. This was insightful for stakeholders.

To identify more potential improvements in the processing of e-forms, the stakeholders also wanted an analysis of variations in the processing of e-forms across six faculties of interest. To address this, we filtered the event log by faculty (using DISCO) and collated data for each faculty, as shown in Figure 1. It is evident that Faculty 3 takes more time for 'faculty feedback' than other faculties. This surprised stakeholders as they assumed Faculty 3 to have the fastest processing times. Based on this and other similar findings, *a case for standardisation of processing of e-forms across faculties was proposed.*

6 Benefits and Lessons Learned

Application of process mining techniques brought forth numerous benefits for the digital transformation program at QUT. Specifically, the following benefits were achieved:

1. *Better visibility of the end-to-end PhD journey and identification of task dependencies in this journey*: Visualising the end-to-end journey of completed and withdrawn PhD students was considered insightful by the stakeholders. This analysis gave them a better understanding of the task dependencies as well as the different behaviours exhibited by students in the past.
2. *Enhanced visualisation of deviations from the expected student journey*: We observed deviations from typical student journeys and these observations resulted in introduction of changes to the PhD journey. New decision points were added in the workflow of APR and Stage 2 milestone e-form.
3. *Clear identification of patterns in the journey of withdrawn students as early risk indicators in the PhD journey*: Our analysis showed that most students withdrew in the first year of their PhD program. Additionally, the data points towards a pattern of increased frequency and duration of periods of leave for students that withdrew. To reduce attrition, QUT has decided to implement automated health check forms every 6 months as a means to monitor the well-being of students at an early stage.
4. *Data-driven insights for process improvement for e-forms*: We found that the time taken by the RSC to process e-forms and the duration of completion of these e-forms by students are performance bottlenecks. The data also revealed re-work taking a considerable time in the processing of e-forms. This finding resulted in a reformulation of task handover rules in the RSC. Furthermore, to reduce form completion times by students, pre-population of forms using information already available in the database was introduced.
5. *Better evidence for standardisation of processing of e-forms across faculties*: We found variations in e-form processing performance across faculties. This insight assisted in making a case for standardisation e-forms processing across faculties, which was approved by the relevant authority.

Insights from this case study had supported the introduction of a new research management system at QUT. According to the manager of the project, "*this project allowed for the review of policies to support the student journey and has underpinned innovative thinking in how to redesign processes and forms for students*." Furthermore, the approach presented here can be replicated by other universities enabling them to use process mining techniques in addition to other methods to improve PhD student journeys. Here, we summarise the lessons learned from this study:

1. *The need for interactive process mining*: The study brings forth the gravity of continual interaction with stakeholders to uncover relevant process models. This is also useful in scenarios where a standard or normative process is not documented. It prevents ruling out behaviours without any underlying reasons. Similarly, 'live' conformance checking can enrich the analysis with domain knowledge and assist in obtaining accurate insights. Additionally, the questions asked during such interactions can also assist stakeholders in assessing the correctness of existing policies and also point to them, if not documented.

2. *Significance of comparative process mining*: Universities usually have a certain degree of autonomy compared with other organisations, which is why variants of processes may be observed. Hence, analysing cohorts of interest in order to further standardise the process can contribute to better performance. Once logs of cohorts are obtained, process models can be discovered, performance measures calculated, and then compared. Comparative analysis enables the identification of variants and root causes of performance differences among cohorts. These findings can in turn be used by decision makers to provide targeted support.

3. *Significance of data-driven evidence to validate hypothesis*: The data-driven evidence provided by process mining analysis allowed stakeholders to validate, or disprove hypothesis about student journey. As mentioned by a stakeholder, "[Process mining] *approach allowed us to have a conversation about the student journey which were not based in beliefs, but data, often in process improvement initiatives the project doesn't have any authority on the subject, as the business areas have a better understanding of the process. Having data allowed the project to challenge conventional beliefs*."

7   Conclusion

The findings of this paper demonstrate the significance of application of process mining techniques in higher education. They also convey how process mining techniques and insights can be used to support a digital transformation initiative, in this case an overhaul of traditional research management systems and associated processes. Some of our findings are also reflected in existing literature, reinforcing the validity of the application of process mining in addressing the analysis questions presented in this study and providing further empirical evidence to the field of higher education studies. Furthermore, the approach presented here can be replicated by other universities enabling them to use process mining techniques to improve PhD student journeys. The work presented in this paper is limited to descriptive analysis, where future work can incorporate more advanced capabilities of predictive analysis. Techniques to systematically translate findings to improvements are also recommended. The work presented in this paper brings forth opportunities for future research, notably the conduct of process mining case studies in other higher education universities in Australia as well as internationally to improve PhD student journeys.

References

1. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, Heidelberg (2016)
2. van der Aalst, W.M., Guo, S., Gorissen, P.: Comparative process mining in education: An approach based on process cubes. In: International Symposium on Data-Driven Process Discovery and Analysis. pp. 110–134. Springer (2013)
3. Agné, H., Mörkenstam, U.: Should first-year doctoral students be supervised collectively or individually? effects on thesis completion and time to completion. Higher Education Research & Development **37**(4), 669–682 (2018)
4. Bogarín, A., Cerezo, R., Romero, C.: A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(1), e1230 (2018)

5. Caparrós-Ruiz, A.: Time to the doctorate and research career: some evidence from spain. Research in Higher Education **60**(1), 111–133 (2019)
6. Cerezo, R., Bogarín, A., Esteban, M., Romero, C.: Process mining for self-regulated learning assessment in e-learning. Journal of Computing in Higher Education **32**(1), 74–88 (2020)
7. De Freitas, S., et al.: Foundations of dynamic learning analytics: Using university student data to increase retention. British Journal of Educational Technology **46**(6), 1175–1188 (2015)
8. Geven, K., Skopek, J., Triventi, M.: How to increase phd completion rates? an impact evaluation of two reforms in a selective graduate school, 1976–2012. Research in higher education **59**(5), 529–552 (2018)
9. Hu, Y.: Treatment learning: Implementation and application. Ph.D. thesis, University of British Columbia (2003)
10. Hunter, K.H., Devine, K.: Doctoral students' emotional exhaustion and intentions to leave academia. International Journal of Doctoral Studies **11**(2), 35–61 (2016)
11. Koenker, R., Bassett Jr, G.: Regression quantiles. Econometrica: journal of the Econometric Society pp. 33–50 (1978)
12. Leemans, S.J., Fahland, D., van der Aalst, W.M.: Exploring processes and deviations. In: International Conference on Business Process Management. pp. 304–316. Springer (2014)
13. Littlejohn, A., Hood, N.: Reconceptualising learning in the digital age: The [un] democratising potential of MOOCs. Springer (2018)
14. Perera, D., Kay, J., Koprinska, I., Yacef, K., Zaïane, O.R.: Clustering and sequential pattern mining of online collaborative learning data. IEEE Transactions on Knowledge and Data Engineering **21**(6), 759–772 (2008)
15. Poon, L.K., Kong, S.C., Wong, M.Y., Yau, T.S.: Mining sequential patterns of students' access on learning management system. In: International conference on data mining and big data. pp. 191–198. Springer (2017)
16. Rokach, L., Maimon, O.Z.: Data mining with decision trees: theory and applications, vol. 69. World scientific (2008)
17. van de Schoot, R., Yerkes, M.A., Mouw, J.M., Sonneveld, H.: What took them so long? explaining phd delays among doctoral candidates. PloS one **8**(7), e68839 (2013)
18. 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: International Learning Analytics & Knowledge Conference. pp. 538–539. ACM (2017)
19. Umer, R., Susnjak, T., Mathrani, A., Suriadi, S.: On predicting academic performance with process mining in learning analytics. Journal of Research in Innovative Teaching & Learning **10**(2), 160–176 (2017)
20. Vidal, J.C., Vázquez-Barreiros, B., Lama, M., Mucientes, M.: Recompiling learning processes from event logs. Knowledge-Based Systems **100**, 160–174 (2016)
21. Wang, R., Zaïane, O.R.: Discovering process in curriculum data to provide recommendation. In: EDM. pp. 580–581 (2015)
22. Winchester-Seeto, T., Homewood, J., Thogersen, J., Jacenyik-Trawoger, C., Manathunga, C., Reid, A., Holbrook, A.: Doctoral supervision in a cross-cultural context: Issues affecting supervisors and candidates. Higher Education Research & Development **33**(3), 610–626 (2014)
23. Wynn, M.T., Poppe, E., Xu, J., ter Hofstede, A.H., Brown, R., Pini, A., van der Aalst, W.: ProcessProfiler3D: A visualisation framework for log-based process performance comparison. Decision Support Systems **100**, 93–108 (2017)
24. Xiao, J.: Digital transformation in higher education: critiquing the five-year development plans (2016-2020) of 75 chinese universities. Distance Education **40**(4), 515–533 (2019)