Describing educational trajectories of engineering students in individual high-failure rate courses that lead to late dropout.

Juan Pablo Salazar-Fernandez[1, 2][0000-0001-7200-0614], Marcos Sepúlveda[2][0000-0002-9467-7666], Jorge Munoz-Gama[2][0000-0002-6908-3911]

[1] Universidad Austral de Chile, Valdivia 5111187, Chile
[2] Pontificia Universidad Católica de Chile, Macul 7820436, Chile
juansalazar@uach.cl

Abstract. Understanding the phenomenon of late dropout has been gaining importance in engineering education. The process that leads to the decision to drop an academic program after freshman year, usually covers several academic periods, and is influenced by different factors. Longitudinal analysis is a good approach to analyze previous events that lead to late dropout. In this case study, we use a process mining approach to answer how educational trajectories of engineering students may describe the process that finishes in late dropout. It was conducted at the Universidad Austral de Chile, using academic records of high-failure rate courses. Through the analysis of educational trajectories in each one of the high-failure rate courses, we found that trajectories of students that dropout early and those that graduate on-time are clearly distinguishable from each other, but the trajectories of students that dropout late or graduate late need to be analyzed in more detail. Late dropout is higher among students that fail in freshman year courses compared to those that fail in sophomore year courses, and passing a course after failing it several times can be a dropout risk factor, which is consistent with the Investment Model. Such findings may be useful for managers and policy makers, because these trajectories can be related to entrance conditions and permanence requirements.

Keywords: Late dropout, educational trajectories, high-failure rate courses, process mining.

1 Introduction

Although several researchers have addressed dropout in higher education [1], in the Western world the number of graduates in engineering programs has remained low and universities are being pressed by governments to increase it, in order to meet the society needs [2]. Most research about dropout in higher education has been concentrated on freshmen [1], but studies and improvements focused on early dropout have possibly contributed to delayed dropout to later years [3]. Currently, more than 40% of college students who do not finish, leave their programs after second year, and most students who finish late, remain at risk of dropping out until close to graduation [4]. Therefore, research on late dropout in engineering education has been recently gaining an increasing importance. In particular, courses with a high-failure rate are common in engineering programs and have been linked to dropout and late graduation [5]. Moreover, recent research suggests that the transition to more rigorous courses, as are found in sophomore year, may present a crucial barrier to completion [4].

We have identified two particular gaps that require more research. First, most of existing research linking high-failure rate courses and dropout in engineering has been developed using qualitative techniques [5] [6]. Second, not enough attention has been paid to dropout that occurs after freshman year [7], and only recently research on dropout as a process, using longitudinal techniques, has increased [8] [9].

The purpose of this case study is to contribute to the understanding of the dropout that occurs in engineering undergraduate programs, two or more years after enrollment. The research question is: How can educational trajectories in individual high-failure rate courses that are associated with late dropout be described?

To answer it, a process mining approach was used to analyze educational trajectories in each high-failure rate course, for engineering students at the Universidad Austral de Chile (UACh). Process mining is a relatively young research discipline that act as a bridge between data science and process science approaches, helping to extract knowledge from event logs that are obtained from information systems [10].

This study contributes to our understanding of late dropout in engineering education, by analyzing educational trajectories at the course level. The results show that educational trajectories of students that dropout early or graduate on-time, are easier to distinguish, but the trajectories of students that dropout late or graduate late are more similar, and need to be analyzed in more detail. Moreover, among students that remain enrolled for more than a two-year period, failure followed by approval in high-failure rate courses during the freshman year relates more to late dropout, while failure followed by approval for sophomore courses are more related to late graduation. These results could be used by managers and policy makers, focusing effort, either to reduce dropout or improve graduation times.

2 Related work

2.1 High-Failure Rate Courses and Late Dropout

Although studies that relate GPA (Grade Point Average) with dropout in engineering are not conclusive [11], there is more agreement among those that relate the performance in high-failure rate courses with dropout [2] [5]. According to the Suresh study [5], it is common in engineering education that faculty promotes the existence of difficult courses that act as an obstacle to success. Students know the high failure rates in those courses and, if they have a poor performance in them, they question their decision to follow engineering.

Efforts have been made to improve retention and reduce failure rates in freshman courses, but failure rates in subsequent courses have increased. It has contributed to increasing late dropout and now, about 40% of all university dropouts are post-freshman students [4]. Literature has linked late dropout with inadequate curricular advancement, possibly as a result of the increasing difficulty of the courses [4], and the reduction of motivation [12] and commitment among students. This particular phenomenon is known as 'sophomore slump', since the trend usually begins during the second year of study [13], although its characteristics may last in subsequent years [14].

2.2 Educational Process Mining

Process mining is a research discipline that helps us to discover process models, make conformance analysis and enhance process models, using data stored in information systems that support processes [10]. The applications of process mining in the educational domain are known as educational process mining [15]. In recent years, research has been conducted on educational data to understand the behavior of students on learning platforms [16], but its use as a support to understand and predict dropout, has been less frequent [16] [17].

Process mining may be useful for exploring data on educational trajectories, but the conclusions extracted will only be useful if the models of analysis are based on theoretical models, which are valid for the research community in the higher education domain [18]. Therefore, in the present case study, the analysis of educational trajectories is based on the Investment Model [19], that explains commitment based on satisfaction, quality of alternatives and investment size, and the Social Cognitive Theory [20], which have been used by the higher education research to explain how self-efficacy beliefs are developed by students and how they are related to student dropout [21], since persistence is only one manifestation of motivation [12].

3 Methods

This case study analyzes the educational trajectories of engineering students in each one of the high-failure rate courses, by means of the application of Process Mining PM^2 methodology [22]. This methodology includes the stages of data extraction, event log generation, model discovery, and model analysis. The data used correspond to anonymized enrollment and course-grade records for 6 cohorts of engineering students who were admitted to the Universidad Austral de Chile between 2004 and 2009. Some programs had a duration of 10 semesters whereas others had 12 semesters, according the Chilean classification between technology-based (first group) and scientific-based (second group) programs [23]. Course-grade records used were from 2004 to 2014. The discovery and analysis of models was conducted using bupaR [24], a process mining tool designed to be used on R [25].

The data extraction stage considered the application of filters and the definition of final status events, according to the credit progress and the number of periods with enrolled courses: ON-TIME-GRAD, for those who had graduated in a period of time equal to or shorter than the nominal duration of each program; LATE-GRAD, for those who had graduated in a longer period; LATE-DROPOUT, for those who had enrolled for more than 2 years and had abandoned their programs; EARLY-DROPOUT, for those who had enrolled in at least one course and had remained for a maximum period of two years. To achieve more representative information, only were considered those courses with a total of 50 or more students enrolled. These courses were ordered according to their average failure rate, and those in the top 5% were considered as high-failure rate. After that, the grades obtained by the students in these courses where filtered from their academic records, to build the event log.

To answer the research question, and to describe the educational trajectories in individual high-failure rate courses that are associated with late dropout, an event log was built and analyzed. The event log is a set of records where each one has the following attributes: CASE-ID, EVENT-NAME, INITIAL-TIMESTAMP, and FINAL-TIMESTAMP. In process mining, the CASE-ID is the label that identifies a sequence of related events [10]. In this case, the CASE-ID is composed by the student identifier, the academic program and the course code. There are three types of event types in this event log: enrollment, results and final-status. The first event in all educational trajectories is ENR (enrollment). After that, each educational trajectory is composed by one or more results-type events. For each student matriculated in a program who took a high-failure rate course, each one of the enrollments for that particular course, which, at the end of the semester, would have led to a pass or fail, were filtered. The EVENT-NAME for each result was tagged as COURSE-PASS or COURSE-FAIL-n, where n is the number of times that course had been failed. Finally, each educational trajectory was finished with the final status reached (EARLY-DROPOUT, LATE-DROPOUT, LATE-GRAD or ON-TIME-GRAD).

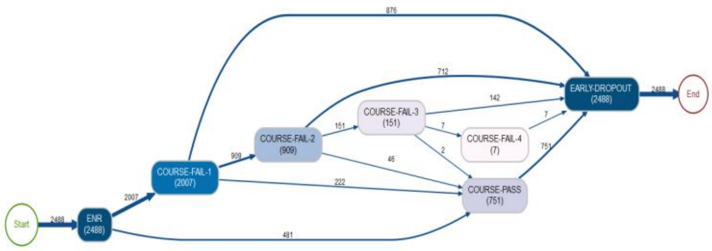
The trajectories were analyzed at a general level, taking into account the aggregate behavior of all the high-failure rate courses, and subsequently in accordance with the positioning of the course in the curriculum (distinguishing between freshman and sophomore years). Frequencies and times were compared, according to the final status event reached, using a proportions analysis [25] to evaluate the differences. Finally, the time elapsed between the passing of a course and the LATE-DROPOUT event was compared for different trajectories.

4 Results

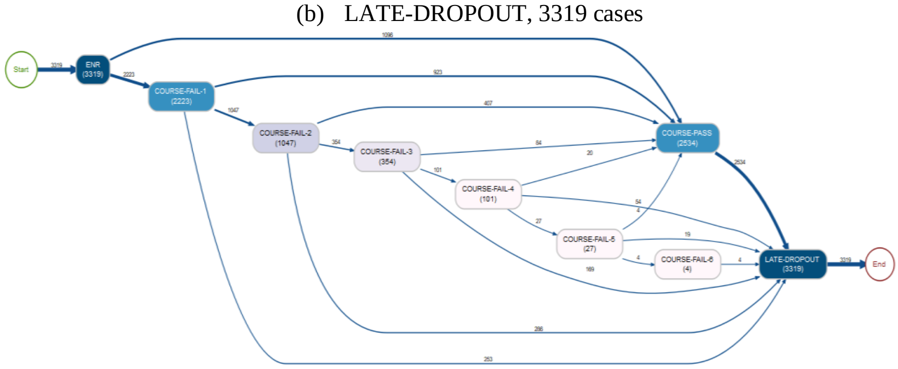
The number of students considered in this case study was 1,886. They enrolled in an engineering program at the UACh between 2004 and 2009, and in May, 2015, they had reached one of the four final states defined. Of these, 780 students ended in EARLY-DROPOUT, 488 in LATE-DROPOUT, 383 in LATE-GRAD and 235 in ON-TIME-GRAD.

Fig. 1 shows the educational trajectories followed by each student in each high-failure rate course from the event log. Clear differences can be identified between those that lead to ON-TIME-GRAD versus EARLY-DROPOUT. There is greater variability among the trajectories that lead to EARLY-DROPOUT, with 8 variants, compared to only 3 variants for ON-TIME-GRAD. The trajectories that lead to EARLY-DROPOUT, for the most part (69.82%), include the failure of high-failure rate courses, although not their subsequent passing. Conversely, in the vast majority (87.34%) of the educational trajectories that lead to ON-TIME-GRAD, the courses are passed at the first opportunity in which they are enrolled. Educational trajectories of students that dropout early and trajectories of students that graduate on-time show clear differences. In contrast, the differences between those of students that dropout late or graduate late, are less evident in the overall analysis and are therefore discussed in more detail below.

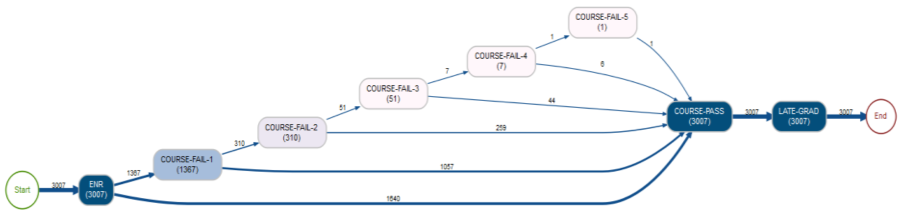
(a) EARLY-DROPOUT, 2488 cases



(b) LATE-DROPOUT, 3319 cases



(c) LATE-GRAD, 3007 cases



(d) ON-TIME-GRAD, 1817 cases

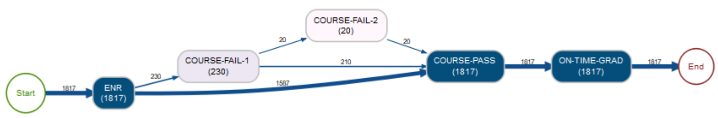


Fig. 1. Educational trajectories in individual high-failure rate courses, grouped for each one of the 4 final states. The darker color represents a higher frequency of occurrence of the event. The thickness of the arcs represents the frequency of the transitions, the exact value of which is displayed adjacent to each arc.

Table 1. Proportion of LATE-DROPOUT and LATE-GRAD, for trajectories that include the approval of a freshman high-failure rate course.

Trajectory	LATE-DROPOUT		LATE-GRAD		TOTAL
	%	n	%	n	n
PASS	45.72%	727	54.28%	863	1,590
FAIL-1 > PASS	51.45%	728	48.55%	687	1,415
FAIL-2 > PASS	70.51%	318	29.49%	133	451
FAIL-3 > PASS	79.75%	63	20.25%	16	79
FAIL-4 > PASS	83.33%	15	16.67%	3	18
FAIL-5 > PASS	100.00%	3	0.00%	0	3

Table 2. Proportion of LATE-DROPOUT and LATE-GRAD, for trajectories that include the approval of a sophomore high-failure rate course.

Trajectory	LATE-DROPOUT		LATE-GRAD		TOTAL
	%	n	%	n	n
PASS	32.79%	359	67.21%	736	1095
FAIL-1 > PASS	35.65%	185	64.35%	334	519
FAIL-2 > PASS	41.87%	85	58.13%	118	203
FAIL-3 > PASS	44.44%	20	55.56%	25	45
FAIL-4 > PASS	62.50%	5	37.50%	3	8
FAIL-5 > PASS	50.00%	1	50.00%	1	2

To determine the relationship between trajectories prior to passing a course and the decision to undertake late dropout, we calculated the proportion of LATE-DROPOUT and LATE-GRAD for those that had passed the high-failure rate courses, based on the number of times each course had previously been failed and the academic year in which each course is included in the study program. This information is presented in Table 1 for freshman high-failure rate courses and in Table 2 for sophomore high-failure rate courses. For freshman year courses, only 45.72% of trajectories that do not include failure of the high-failure rate course led to LATE-DROPOUT; a value that rises to 70% if the course had been failed on two or more occasions, prior to being passed. On the other hand, for sophomore year courses, 32.79% of trajectories that do not include failure of the high-failure rate course led to LATE-DROPOUT. This value rises slightly if the high-failure rate course has been failed once, twice or three times, prior to being passed. Therefore, it can be concluded that, for freshman year courses, the proportion of trajectories that lead to LATE-DROPOUT shows a greater increase as the number of course failures rises, prior to them being passed. Fig. 2 shows the proportion of LATE-DROPOUT presented in Table 2. It illustrates both the substantive increase in the proportion of LATE-DROPOUT that occurs when one or more freshman year courses are failed, as well as the difference in the proportion of LATE-DROPOUT in freshman and sophomore year courses for the same number of previous failures. The statistical significance of these differences is discussed below.

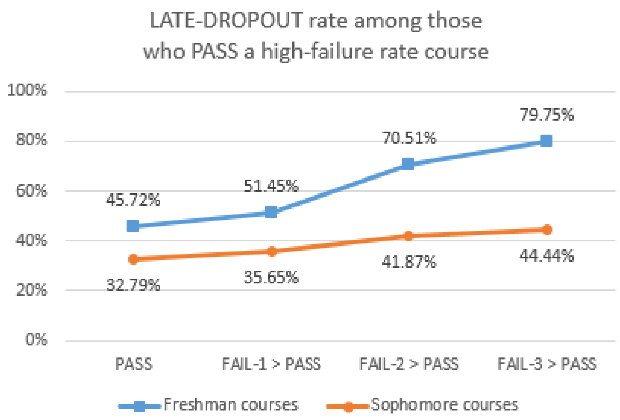


Fig. 2. Proportion of LATE-DROPOUT among those who pass a high-failure rate course from a particular academic year, after having failed 0, 1, 2 or 3 times.

Table 3. Proportion of LATE-DROPOUT among those students that passed a high-failure rate course and led to either LATE-DROPOUT or LATE-GRAD, for freshman and sophomore courses, grouped by when the students approved the course. p-value is obtained by comparing the proportion between those students that approved the course in the second time versus the first time, in the third time versus the second time, and so on.

Trajectory	Freshman Courses		Sophomore Courses	
	% LATE-DROPOUT	p-value	% LATE-DROPOUT	p-value
PASS	45.72%		32.79%	
		0.00195*		0.2806
FAIL-1 > PASS	51.45%		35.65%	
		1.81e-12*		0.1418
FAIL-2 > PASS	70.51%		41.87%	
		0.1214		0.8814
FAIL-3 > PASS	79.75%		44.44%	
		0.9865		0.5766
FAIL-4 > PASS	83.33%		62.50%	

* statistically significant differences, p-value < 0.05

Although the most frequent trajectories that lead to LATE-DROPOUT include the approval of the high-failure rate course, as shown in Fig. 1, late dropout is related to the trajectory the students follow before passing the course. Table 3 shows that the differences in the proportion of LATE-DROPOUT and LATE-GRAD are statistically significant among those who have failed a freshman high-failure rate course once compared to none, and those who have failed it twice compared to once. In the case of sophomore high-failure rate courses, the proportion of LATE-DROPOUT is lower than that of LATE-GRAD and, although it increases slightly as the number of failures rises, the differences are not statistically significant. Failing a freshman high-failure rate course is, therefore, strongly related to the probability of subsequent dropout, even if that course is passed at a later semester. On the other hand, this relationship is not as strong in the case of sophomore high-failure rate courses.

Table 4 shows that, for trajectories that lead to LATE-DROPOUT and include failing a course two or more times, the students who finally pass the course, remain on average 4.37 semesters more after passing the course. On the contrary, students who do not pass the course, remain on average only 1.25 semesters more, after COURSE-FAIL-2. Even if students pass a course, the fact they have failed it (and the number of times they have failed it) should be considered as a dropout risk factor.

Table 4. Average number of semesters that students who end in LATE-DROPOUT remain in the academic program, after failing a course twice.

Trajectory	Freshman Courses	Sophomore Courses	All Courses
After failing the course twice, the student finally pass the course, but ends in late dropout anyway	4.57	3.68	4.37
After failing the course twice (or more) and not being able to pass it, the student ends in late dropout	1.28	1.20	1.25

5 Discussion and Conclusions

This case study illustrates the application of process mining to the analysis of educational trajectories in individual high-failure rate courses, contributing to the comprehension about how they are related to late dropout.

The educational trajectories that ended in EARLY-DROPOUT and ON-TIME-GRAD showed clear differences, being satisfactory the vast majority of those that ended in ON-TIME-GRAD, while those that ended in EARLY-DROPOUT were mostly unsatisfactory. By the contrary, the differences between those that ended in LATE-DROPOUT and LATE-GRAD were less evident.

Analysis of trajectories that include passing and failing individual high-failure rate courses allowed us to confirm the importance of prior investment in time and resources regarding the decision to continue [8] [19], as well as to reinforce the hypothesis of academic efficacy [20]. In freshman high-failure rate courses, trajectory prior to passing a course was of significant importance, with a high proportion of students ending in LATE-DROPOUT, especially those who passed the high-failure rate course after having failed it on two or more occasions. Conversely, the proportion of LATE-GRAD was higher than that of LATE-DROPOUT among those who passed sophomore high-failure rate courses, regardless of the number of times the course had been previously failed. Passing a course after failing two or more times should be considered a dropout risk factor for several semesters after passing it.

These findings have implications for entrance conditions and program permanency requirements. First, the inadequate match between entrance requirements and academic achievement has been documented as a major cause of dropout in engineering [11], and bad results in freshman courses produce not only EARLY-DROPOUT, but also LATE-DROPOUT, even in students who have passed freshman high-failure rate courses after failed them. Second, if chances to finish engineering are very low between students who have failed high-failure rate courses two or more times, particularly freshman courses, requirements of permanency for students with unsatisfactory trajectories should be checked.

Main limitations of this research are associated to the specific context of analysis and the techniques used. First, although conclusions could be useful in different contexts, our findings are based on a specific case study, developed at a specific university. Second, the analysis was limited only to the curriculum dimension and did not include withdrawn courses, but only those in which the student obtained a final grade at the end of the academic period. Third, the conclusions drawn from process mining depend on the accuracy and completeness of the data used [26].

We believe that our analysis could be replicated in other context, and it could be useful to analyze educational trajectories in high-failure rate courses at different aggregation levels, as well as to incorporate different attributes that describe the personal and institutional context in which that academic results are obtained.

Acknowledgements

This case study was funded by CONICYT/DOCTORADO NACIONAL 2015 (Grant N° 21150985) and supported by FONDECYT Regular (Grant N° 1150923). Anonymized academic records for engineering students were provided by the Universidad Austral de Chile.

References

1. Tinto, V.: Research and Practice of Student Retention: What's Next? Journal of College Student Retention. 8(1),1-19 (2006).
2. Boles, W., Whelan, K.: Barriers to student success in engineering education. European Journal of Engineering Education. 42(4), 368-381 (2017).
3. Nora, A., Barlow, E., Crisp, G.: Student Persistence and Degree Attainment Beyond the First Year in College. The Need for Research. College Student Retention: Formula for Student Success. Rowman & Littlefield Publishers. (2005).
4. Mabel, Z., Britton, T.: Leaving late: Understanding the extent and predictors of college late departure. 69, 34-51 (2018).
5. Suresh, R. The relationship between barrier courses and persistence in engineering. Journal of College Student Retention. 8(2), 215-239 (2006).
6. Strenta, A.C.. Elliot, R., Adair, R., Matier, M., Scott, J.: Choosing and Leaving Science in Highly Selective Institutions. Research in Higher Education. 35(5), 513-547 (1994).

7. Willcoxson, L., Cotter, J., Joy, S.: Beyond the first-year experience: the impact on attrition of student experiences throughout undergraduate degree studies in six diverse universities. Studies in Higher Education. 36(3), 331-352 (2011).
8. Yue, H., Fu, X.: Rethinking Graduation and Time to Degree: A Fresh Perspective. Research in Higher Education. 58, 184-213 (2017).
9. Moncada, L., Arias, M., Negrete, J.: Dropout, lag in the completion of studies, and graduation. Analysis of the evolution of a cohort of students enrolled in the Ecuadorian distance study system. I Conferencia Latinoamericana de Analíticas de Aprendizaje. (2018).
10. Van der Aalst, W.M.P.: Process mining: Data science in action. Springer (2016).
11. Geisinger, B.N., Raman, D.R.: Why they Leave: Understanding Student Attrition from Engineering Majors. International Journal of Engineering Education. 29(4), 914-925 (2013).
12. Tinto, V.: Through the Eyes of Students. Journal of College Student Retention: Research, Theory & Practice. 19(3), 245-269 (2017).
13. Webb, O.J., Cotton, D.R.E.: Deciphering the sophomore slump: changes to student perceptions during the undergraduate journey. Higher Education. 2018.
14. Marra, R.M., Rodgers, K.A., Shen, D., Bogue, B.: Leaving engineering: A multi-year single institution study. Journal of Engineering Education. 101(1), 6-27 (2012).
15. Bogarín, A., Cerezo R., Romero, C.: A survey of educational process mining. WIREs: Data Mining Knowledge Discovery. 8(1), (2017).
16. Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R.F., Morales, N., Munoz-Gama, J.: Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. Computers in Human Behavior. 80, 179-196 (2018).
17. Deeva, G., De Smedt, J., De Konick, P., De Weerdt, J.: Dropout Prediction in MOOCs: A Comparison Between Process and Sequence Mining. Business Process Management Workshops, BPM2017. Vol. 308, Springer (2018).
18. Reinmann, P., Markauskaite, L., Bannert, M.: e-Research and learning theory: What do sequence and process mining methods contribute? British Journal of Educational Technology. 45(3), 528-540 (2014).
19. Rusbult, C., Martz, J., Agnew, C.: The Investment Model Scale: Measuring commitment level, satisfaction level, quality of alternatives and investment size. Personal Relationships. 5, 357-391 (1998).
20. Bandura, A.: Self-efficacy: The exercise of control. Freeman, New York (1997).
21. Hutchinson, M.A., Follman, D.K., Sumpter, M., Bodner, G.M.: Factors Influencing the Self-Efficacy Beliefs of First-Year Engineering Students. Journal of Engineering Education. 95(1), 39-47 (2006).
22. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P. (2015). PM2: a Process Mining Project Methodology. CAiSE 2015, LNCS 9097, pp. 297-313.
23. CNA-Chile. Criterios de Evaluación para Carreras de Ingeniería. Comisión Nacional de Acreditación. Ministerio de Educación, Chile (2007).
24. Janssenswillen, G., Depaire, B.: BupaR: business process analysis in R. Proceedings of the BPM Demo Track (2017).
25. R. Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2017).
26. Bose, R.J.C., Mans, R.S., van der Aalst, W.M.P.: Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously. Computational Intelligence and Data Mining (CIDM), pp. 127-134 (2013).