Identifying Factors for Master Thesis Completion and Non-completion Through Learning Analytics and Machine Learning

Jalal Nouri¹✉, Ken Larsson¹, and Mohammed Saqr²

¹ Stockholm University, Stockholm, Sweden

{jalal, kenlars}@dsv.su.se

² University of Eastern Finland, Joensuu, Finland

mohammed.saqr@uef.fi

Abstract. The master thesis is the last formal step in most universities around the world. However, all students do not finish their master thesis. Thus, it is reasonable to assume that the non-completion of the master thesis should be viewed as a substantial problem that requires serious attention and proactive planning. This learning analytics study aims to understand better factors that influence completion and non-completion of master thesis projects. More specifically, we ask: which student and supervisor factors influence completion and non-completion of master thesis? Can we predict completion and non-completion of master thesis using such variables in order to optimise the matching of supervisors and students? To answer the research questions, we extracted data about supervisors and students from two thesis management systems which record large amounts of data related to the thesis process. The sample used was 755 master thesis projects supervised by 109 teachers. By applying traditional statistical methods (descriptive statistics, correlation tests and independent sample t-tests), as well as machine learning algorithms, we identify five central factors that can accurately predict master thesis completion and non-completion. Besides the identified predictors that explain master thesis completion and non-completion, this study contributes to demonstrating how educational data and learning analytics can produce actionable data-driven insights. In this case, insights that can be utilised to inform and optimise how supervisors and students are matched and to stimulate targeted training and capacity building of supervisors.

Keywords: Thesis · Master · Learning analytics · Completion · Dropout · Retention · Machine learning

1 Introduction

The master thesis is the last formal step in most universities around the world. However, all students do not finish their master thesis. A considerable number of students struggle with the thesis process, resulting in delays, disruptions, and non-completion of their degrees [1–3]. Such outcomes are devastating for individual students and academic institutions that waste time, money and energy, and for societies that are not

strengthened with high-skilled workers [2, 4, 5]. Therefore, ensuring that students enrolled in graduate programs obtain their degrees in a timely fashion is in the best interest of students, higher education institutions and societies [4, 6].

However, the thesis is a challenging endeavour that requires skills, aptitude, and determination for successful, timely completion [5, 7–10]. Thus, it is reasonable to assume that non-completion of higher education degrees should be viewed as a substantial problem that requires serious attention and proactive planning [2, 4, 5, 11–13].

Previous research related to thesis projects has identified some variables that influence the performance of students undertaking thesis projects; variables that, in particular, point out the relation between the student candidate and the supervisor [2, 5, 14]. The specific student variables that have been indicated as influencing thesis completion are students' attitudes and motivation [10], the students' average entry grade [12], and the students' communication and language skills [13]. Among the supervisor variables, it has been shown that the supervisor's experience, research output and workload constitute factors of thesis success [13, 15]. However, the review of the literature leads to the conclusion that there are few studies explicitly focusing on master thesis projects. Studies on completion of thesis projects mostly concern the doctorate thesis [16, 17] while studies on master thesis completion tend to focus on the whole program, not the thesis specifically [18–20]. Furthermore, most studies have used a qualitative approach to investigate factors for thesis completion; single factors have been looked at in an isolated way with a primary focus on student variables and on completion factors (and not on non-completion and supervisor variables) [21–23]. Furthermore, there are few contemporary studies that look at factors for success and failure related to thesis work.

Today, the introduction of thesis management systems, such as SciPro from Stockholm University [24] and Thesis Writer (TW) from Zurich University of Applied Sciences [25], generate a lot of data concerning many aspects of the thesis process. This paves the ground for using learning analytics techniques in order to gain data-driven insights about thesis management and the factors that affect thesis retention [26]. Learning analytics have been used successfully to early map the indicators of successful course completion, inform course design, provide insights and feedback to teachers and students, as well as improve education outcome [27].

This study takes as a departure point to better understand factors that influence completion – and in particular – non-completion of master thesis projects. More specifically, we ask: which student and supervisor variables influence completion and non-completion of master thesis? Can we predict completion and non-completion of master thesis using such variables in order to optimize the matching of supervisors and students?

To answer these research questions, we extracted data about supervisors and students from two thesis management systems, Daisy and SciPro from the Department of Computer and Systems Sciences, Stockholm University, which record large amounts of data related to the thesis process. The sample used was 755 master thesis projects supervised by 109 teachers. By applying traditional statistical methods (descriptive statistics, correlation tests and independent sample t-tests), as well as machine learning algorithms, we identify five central factors that can accurately predict master thesis completion and non-completion. Besides the identified factors and predictors that

explain master thesis completion and non-completion, this study contributes to demonstrating how educational data and learning analytics can produce actionable data-driven insights. In this case, insights that can be utilised to, on the one hand, inform and optimise how supervisors and students are matched, and on the other hand, stimulate targeted training and capacity building of supervisors.

2 Identified Factors in the Literature Explaining Thesis Completion and Non-completion

Our literature review has led to the identification of two groups of factors that influence thesis outcomes: related to the student candidate and the supervisor. Below we give an account of what is known about these two groups of factors.

Rennie and Brewer [10] using a grounded theory approach to investigate the problem of thesis delay proposed the term ‘thesis-blocking’. They propose that thesis blocking factors are more numerous than factors leading to completing it in a timely fashion. Successful thesis completion is dependent on the candidate’s conformity and acceptance of the process. Failure of the supervisor to handle a candidate’s negative feelings is the reason why many candidates to be stuck in the middle of the path [10]. House and Johnson’s findings point to the applicants’ average entry grade as a decisive predictive factor of successful, timely completion [12], a finding that was corroborated by Jiranek [13] and Wright and Cochrane [28].

On the other hand, studies have shown that entry grade is not a significant predictor of completion [28, 29]. In a study by Pascarella and Terenzini [30], it was shown that the background characteristics, including entry grades, only explain a small part of retention, while academic and social integration explain more.

Other student factors affecting the completion or non-completion include communication skills and language proficiency skills [13], self-reliance and independence [31]. However, a right balance and proactive planning along with institutional support could mitigate the impact and assist the candidates [2, 5, 13, 15, 31, 32]. Contrary to what is a common belief, part-time older candidates appear to be better than their counterparts in their approach to research, other duties and being independent [28].

It has also been shown that supervisors behaviours are crucial in every stage of the thesis work, in supporting the thesis writing process, rectifying errors, suggesting directions and being responsible for arranging the defence [33]. Rennie and Brewer compare the supervisor’s role in these cases to the writer’s block phenomenon [10]. They suggested that both share essential features, the main problem being the writer’s internalisation of the critical feedback by the supervisor and poor management of duties and time constraints.

A healthy relationship between student and supervisor is helpful for the success of the thesis. The thesis is an embedded social exercise more than most of the other educational projects, therefore collaborating with the supervisor, regular productive meetings and the ability to reach a shared understanding are central to the success of the project [2, 5, 14, 34, 35]. A relationship where the supervisor exerts a moderate control of the process and more significant affiliation was found to influence the successful

outcome in terms of time to completion and completion rates [36]. Supervisor experience and research production is a factor that might affect positively [13].

In general, the supervisors support through all the stages of the thesis process is an indispensable factor [5, 13, 32, 34, 37]. On the contrary, supervisors that are overwhelmed by research work, teaching or multiple students have less time for students who have negative results on the thesis work [2, 15]. Furthermore, students report that its central that supervisors provide constructive, on-time feedback, as well as encouragement [38].

3 Method

3.1 Sample and Context

The sample for this study consisted of master students' thesis projects ($n = 755$) during the period between 2010 and 2017 at the Department of Computer and Systems Sciences, Stockholm University, Sweden. Since it takes approximately 350 days for students to complete a thesis project (from course registration to grade registration), data from the year, 2018 were excluded as they contained many projects likely to be completed after the data extraction. The dropout rate for the thesis project at the department is approximately 43% for the period studied. We have included all master thesis projects that adhere to the present curriculum for thesis projects.

3.2 Data Collection

A challenge in data collection for learning analytics is to avoid amplifying errors from different standards in data sources, especially if some sources are external and out of control. In this study, to minimise this risk for all data sources, we used data that are under the control of the university.

Data collection was performed in several iterative steps. Using SQL (structured query language) queries, we extracted data from two different data systems used by the department to record data about the thesis projects. From these systems, we collected thesis project data concerning both students and supervisors. Informed by factors identified by previous research [12, 13], and taking into account additional variables that were available in the systems that record thesis data. We focused in general on three groups of factors that influence the academic thesis process, namely: (1) student's previous performance in the master program; (2) supervisor's thesis project performance and experience; and (3) supervisor's research output.

More specifically, we extracted the following variables:

- Thesis project: start and completion date. From this, the number of days to completion was calculated.
- The students ($n = 755$): the grade of the thesis, the average grade in the study before the master thesis, and the number of course credits received within the educational program.

- The supervisors ($n = 105$): number of scientific publications, the average number of scientific publications per year, number of complete/incomplete thesis projects, the average grade of thesis projects, number of started thesis projects, and average days of supervisors to complete thesis projects were calculated from the projects.

All data was anonymised by converting personal identifiers to fictive IDs. The researchers who did the analysis did not know the identity of the subjects. The data was subsequently prepared for statistical and predictive analytics by removal of extreme- and null values and through the computation of relevant variables.

Ethical approval for this study was obtained through the Regional Board of Ethical Vetting in Stockholm. Consent for participating in this research was also obtained from the selected supervisors in the sample. Six supervisors and their associated thesis projects were excluded due to no consent for using their data were received.

3.3 Data Analysis

The analysis was performed using SPSS, and R. Spearman correlation test was conducted to investigate the correlation between incomplete thesis projects (dropouts) and student and supervisor variables. Multiple independent sample t-tests were performed in order to explore differences between completers and non-completers with regards to student and supervisor variables. The Shapiro–Wilk test of normality was employed and confirmed that the assumptions for the t-tests were satisfied.

For the predictive analytics, seven supervised machine learning classifiers were applied: Naive Bayes, Logistic Regression, Gradient Boosted Trees, Neural Network, Deep Learning, Decision Tree, and Random Forest in order to predict completers and non-completers of thesis projects. These classifiers were chosen because they are frequently used for predicting dropout, and each has demonstrated good and comparable performance in predicting at-risk students and dropout [39, 40]. The data set was split into a training and testing set. The training set consisted of 70% of the total data set, and the testing set the remaining 30%. After the implementation of the predictive models, features were ranked using the information gain ratio. To prevent overfitting and increase robustness, 10-fold cross-validation was performed, where performances were measured from multiple iterations of cross-validation and averaged over iterations. To measure the prediction performance of the different models, the area under the receiver operating characteristic curve (AUC) was obtained, along with measures for precision and recall.

4 Results

In Table 1 the full descriptive statistics are presented. Among the 755 thesis projects studied, 57% were completed, and 43 remained non-completed (see Fig. 1).

After performing the descriptive analysis presented in Table 1, a correlations tests (Spearman's) was performed in order to study the correlation between completion of thesis projects and all student and supervisor variables presented in the above table. This analysis revealed that completion is significantly correlated with the ratio of

Fig. 1. Histogram over completed and non-completed thesis projects

Table 1. Descriptive statistics

Variable	N	M	SD	Range	Min	Max
Student average study grade	775	3.53	.83	4.00	1.00	5.00
Student average thesis grade	775	2.16	2.04	5.00	.00	5.00
Supervisor average days to complete	772	417.44	183.85	1415.00	157.00	1572.00
Supervisor average thesis grade	772	3.72	.51	3.50	1.50	5.00
Supervisor no. of scientific publications	760	59.82	48.69	271.00	1.00	272.00
Supervisor no. of incomplete thesis projects	774	11.11	7.80	33.00	.00	33.00
Supervisor no. of started thesis projects	774	29.27	16.14	77.00	1.00	78.00
Supervisor ratio incomplete thesis projects	774	38.25	17.65	83.33	.00	83.33
Supervisor average no. of publication per year	774	4.84	3.67	13.25	.00	13.25
Student no. of course credits within program	769	62.65	45.25	206.00	.00	206.00
Completed vs non-completed thesis projects	775	.57	.50	1.00	.00	1.00

All grade-related variables on a 6-item scale ranging from F = 0 to A = 5

incomplete thesis projects of supervisors ($r = -0.36$, $p < 0.01$), students' average grade in their study program at the university ($r = 0.28$, $p < 0.01$), supervisors total number of incomplete thesis projects ($r = -0.17$, $p < 0.01$), the average time it takes for supervisors to complete thesis projects ($r = -0.14$, $p < 0.01$), the ratio of supervisor thesis ideas ($r = 0.10$, $p < 0.05$), and supervisors average thesis grade ($r = 0.09$, $p < 0.04$). As can be noted, the ratio and total amount of unfinished thesis projects by supervisors presented the strongest correlations with thesis dropout, together with students' average grade during the educational program.

Multiple independent t-tests were also performed in order to explore differences between completers and dropouts with regards to many student and supervisor variables. See Table 3 for a full presentation of the t-test results. Based on these tests, the following can be concluded:

- there is a significant difference between completers ($M = 3.73$, $SD = 0.85$) and non-completers ($M = 3.26$, $SD = 0.74$) regarding their average grade during their studies in the program they are seeking to graduate in, $t(-8.26) = 1.07$, $p < 0.01$;
- there is a significant difference between completers ($M = 399.99$, $SD = 166.02$) and non-completers ($M = 440.92$, $SD = 203.33$) in terms of their supervisors' average days to complete thesis projects, $t(3.08) = 5.57$, $p < 0.01$;
- there is significant difference between completers ($M = 3.76$, $SD = 0.49$) and non-completers ($M = 3.66$, $SD = 0.53$) in terms of their supervisors average thesis grade, $t(-2.51) = 0.32$, $p < 0.05$;
- there is a significant difference between completers ($M = 9.98$, $SD = 7.45$) and non-completers ($M = 12.63$, $SD = 8.01$) in terms of their supervisors' total number of incomplete thesis projects $t(4.75) = 2.28$, $p < 0.01$;
- there is a significant difference between completers ($M = 32.71$, $SD = 15.90$) and non-completers ($M = 45.65$, $SD = 17.17$) in terms of their supervisors' ratio of incomplete thesis projects, $t(4.75) = 2.28$, $p < 0.01$, and

Significant differences were, however, not revealed concerning the total number of scientific publications published by supervisors, the total number of thesis projects supervised by the supervisors, or the total course credits received by students within the educational program prior the master thesis (Table 2).

Table 2. Significant differences between completed and non-completed thesis projects (t-test)

Variable		M	SD	F	t	p
Student average study grade	Complete	3.73	0.85	1.07	-8.26	<0.01
	Incomplete	3.26	0.74			
Supervisor average days to complete	Complete	399.99	166.02	5.57	3.08	<0.01
	Incomplete	440.92	203.33			
Supervisor average thesis grade	Complete	3.76	0.49	0.32	-2.51	<0.05
	Incomplete	3.66	0.53			
Supervisor no. of incomplete thesis projects	Complete	9.98	7.45	2.28	4.75	<0.01
	Incomplete	12.63	8.01			
Supervisor ratio incomplete thesis projects	Complete	32.71	15.90	2.28	4.75	<0.01
	Incomplete	45.65	17.17			

4.1 Predicting Completion and Non-completion

Then predictive analytics was performed using several machine learning models (Naive Bayes, Logistic Regression, Deep Learning, Decision Tree, Random Forest and Gradient Boosted Trees) in order to predict the completion/non-completion variable using the features described in Table 1. The performance across the models showed AUC values between 0.74 and 0.83 (see Table 3).

Table 3. Prediction accuracy and ROC

Model	AUC	Accuracy	F-measure	Recall
Gradient Boosted Trees	0.83	71.6%	0.66	67.0%
Logistic Regression	0.82	73.5%	0.65	58.9%
Naive Bayes	0.80	70.7%	0.58	49.7%
Deep Learning	0.76	70.3%	0.63	64.2%
Random Forest	0.74	63.5%	0.21	12.4%
Decision Tree	0.50	59.0%	na	4.2%

The Gradient Boosted Trees model proved to perform best concerning accuracy and AUC, with almost 72% accuracy in predicting completers and non-completers. The actual non-completers could be predicted with a 66% precision and 67% class recall; while the completers could be predicted with 76% precision and 75% class recall (see Table 4).

As can be seen from Table 5, the features with most weight were the ratio of unfinished thesis projects of supervisors, students' average grade during university studies, supervisors' total number of incomplete projects and the average time it takes for supervisors to complete a thesis project.

Table 4. Prediction of completers and non-completers using Gradient Boosted Trees

	True completer	True non-completer	Class precision
Predicted Completer	96	31	76%
Predicted Non-completer	32	63	66%
Class Recall	75%	67%	

Table 5. Weights of selected features

Features	Weight
Supervisor ratio incomplete thesis projects	1.0
Student average study grade	0.78
Supervisor no. of incomplete thesis projects	0.45
Supervisor average days to complete	0.31
Supervisor average thesis grade	0.23
Supervisor average no. of publications per year	0.17

5 Discussion

Not finishing a master thesis is a devastating personal experience for students that costs precious time, loss of money and energy. Non-completion also results in a vast waste of faculty time and institutional resources, and a societal loss of high skilled workers [2, 4, 5]. This study took as a departure point to address this problem by using large amounts of thesis-related data generated in thesis management systems in order to create

data-driven insights about the factors that influence completion and non-completion of master thesis projects. Such a learning analytics approach led us to identify factors that have not been reported on in the research literature.

The analysis of the data resulted in the identification of five central factors that influence students' completion and non-completion of master thesis projects. The strongest factor and predictor for non-completion, not reported on before, showed to be supervisors' history of incomplete thesis projects. This factor correlated *more* with incomplete thesis projects *than* student's academic performance before starting the thesis (which was the second strongest predictor) and was the factor/feature that had most information gain (weight) in the predictive models produced by the employed machine learning algorithms. Three additional factors/predictors were identified related to the supervisor, namely: (1) the average days it takes for supervisors to complete thesis projects; (2) the average grade of thesis works supervised, and (3) the average number of scientific publications produced by the supervisor per year.

Thus, the supervisor's historical thesis supervision performance and their performance as researchers, together with students' academic performance prior to the thesis, to a high extent determine success and failure of master thesis projects. While previous research mainly through qualitative studies has demonstrated that supervisors indeed play a significant role in the thesis process, by in particular pointing out *how* successful teachers supervise, this quantitative study identify actual predictors related to the supervisor and demonstrate the significant effect of supervisor historical performance on master thesis completion and non-completion, which constitute a central contribution of this study. However, the finding that students' academic performance prior to the thesis is a significant predictor has been reported on before and corroborate previous research [12, 13, 28].

Besides the identified factors and predictors that explain master thesis completion and non-completion, this study contributes to demonstrating how educational data and learning analytics can produce actionable data-driven insights. In this case, gained insights can be utilized to, on the one hand, inform and optimize how supervisors and students are matched, and on the other hand, stimulate targeted training and capacity building of supervisors.

Future research work can build upon this study and bridge its limitations by adding more contextual factors to the analysis, such as students' internal conditions and dispositions [41–43]. By dispositions, we mean behavioral and cognitive factors such as motivation (to write a master thesis, for instance), engagement, self-regulation skills, strategies and attitudes [43, 44]. Such an approach would most likely increase the probability of finding additional factors that influence the master thesis process and increase accuracy, replicability and transferability of prediction models [27, 45, 46].

References

1. Kamler, B., Thomson, P.: The failure of dissertation advice books: toward alternative pedagogies for doctoral writing. *Educ. Res.* **37**(8), 507–514 (2008)
2. Rauf, F.A.: Challenges of thesis work: towards minimizing the non-completion rate in the postgraduate degree program. *Eur. J. Bus. Manag.* **8**(7), 113–124 (2016)

3. Wong, P.T.P.: Meaning making and the positive psychology of death acceptance. *Int. J. Existential Psychol. Psychother.* **3**(2), 73–82 (2010)
4. Baum, S., Ma, J., Payea, K.: Education pays. *The Benefits of Higher Education for Individuals and Society* (2013)
5. Ho, J.C., Wong, P.T., Wong, L.C.: What helps and what hinders thesis completion: a critical incident study. *Int. J. Existential Psychol. Psychother.* **3**(2) (2010)
6. Bourke, S., et al.: Attrition, completion and completion times of PhD candidates. In: *AARE Annual Conference, Melbourne* (2004)
7. Agu, N., Oluwatayo, G.K.: Variables attributed to delay in thesis completion by postgraduate students. *J. Emerg. Trends Educ. Res. Policy Stud.* **5**(4), 435–443 (2014)
8. Ferrer, F.P.: Determinants of performance in thesis: evidence from selected filipino graduate students. *Int. J. Educ. Res.* **2**(10), 189–202 (2014)
9. Morton, K.R., Worthley, J.S.: Psychology graduate program retention, completion and employment outcomes. *J. Instr. Psychol.* (1995)
10. Rennie, D.L., Brewer, L.: A grounded theory of thesis blocking. *Teach. Psychol.* **14**(1), 10–16 (1987)
11. Chin, W.Y., et al.: Analyzing the factors that influencing the success of post graduates in achieving graduate on time (GOT) using analytic hierarchy process (AHP). In: *AIP Conference Proceedings*. AIP Publishing (2017)
12. House, J.D., Johnson, J.J.: Predictive validity of Graduate Record Examination scores and undergraduate grades for length of time to completion of degree. *Psychol. Rep.* **71**(3), 1019–1022 (1992)
13. Jiranek, V.: Potential predictors of timely completion among dissertation research students at an Australian faculty of sciences. *Int. J. Doctoral Stud.* **5**(1), 1–13 (2010)
14. Pitchforth, J., et al.: Factors affecting timely completion of a PhD: a complex systems approach. *J. Scholarsh. Teach. Learn.* **12**(4), 124–135 (2012)
15. van de Schoot, R., et al.: What took them so long? Explaining PhD delays among doctoral candidates. *PLoS ONE* **8**(7), e68839 (2013)
16. Can, E., et al.: Supervisors' perspective on medical thesis projects and dropout rates: survey among thesis supervisors at a large German university hospital. *BMJ Open* **6**(10), e012726 (2016)
17. Van Ours, J.C., Ridder, G.: Fast track or failure: a study of the graduation and dropout rates of Ph D students in economics. *Econ. Educ. Rev.* **22**(2), 157–166 (2003)
18. DesJardins, S.L., Kim, D.-O., Rzonca, C.S.: A nested analysis of factors affecting bachelor's degree completion. *J. Coll. Student Retention Res. Theor. Pract.* **4**(4), 407–435 (2003)
19. Graham, L.D.: Predicting academic success of students in a master of business administration program. *Educ. Psychol. Measur.* **51**(3), 721–727 (1991)
20. Herzog, S.: Estimating student retention and degree-completion time: decision trees and neural networks vis-à-vis regression. In: *New Directions for Institutional Research*, vol. 131, pp. 17–33 (2006)
21. Ishitani, T.T.: Studying attrition and degree completion behavior among first-generation college students in the United States. *J. High. Educ.* **77**(5), 861–885 (2006)
22. Kovacic, Z.: Early prediction of student success: mining students' enrolment data. In: *Informing Science + Information Technology Education Joint Conference, Cassino, Italy* (2010)
23. Zwick, R., Sklar, J.C.: Predicting college grades and degree completion using high school grades and SAT scores: the role of student ethnicity and first language. *Am. Educ. Res. J.* **42**(3), 439–464 (2005)
24. Hansén, P., Hansson, H.: Optimizing student and supervisor interaction during the SciPro thesis process – concepts and design. In: Li, F.W.B., Klamma, R., Laanpere, M., Zhang, J.,

- Manjón, B.F., Lau, R.W.H. (eds.) ICWL 2015. LNCS, vol. 9412, pp. 245–250. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25515-6_23
25. Rapp, C., Ott, J.: Learning analytics in academic writing instruction—opportunities provided by Thesis Writer (TW). In: Igel, C., Ullrich, C., Martin, W. (eds.) *Bildungsräume* 2017, pp. 391–392. Gesellschaft für Informatik, Bonn (2017)
 26. Rapp, C., Kauf, P.: Scaling academic writing instruction: evaluation of a scaffolding tool (Thesis Writer). *Int. J. Artif. Intell. Educ.* **28**, 1–26 (2018)
 27. Ifenthaler, D., Mah, D.-K., Yau, J.Y.-K.: *Utilizing Learning Analytics to Support Study Success*. Springer, Cham (2018)
 28. Wright, T., Cochrane, R.: Factors influencing successful submission of PhD theses. *Stud. High. Educ.* **25**(2), 181–195 (2000)
 29. Siegel, L.: A study of Ph. D. Completion at Duke University. *CGS Communicator*, XXXVIII **1**(1), p. 2 (2005)
 30. Pascarella, E.T., Terenzini, P.T.: Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: a path analytic validation of Tinto's model. *J. Educ. Psychol.* **75**(2), 215 (1983)
 31. Castro, V., et al.: The road to doctoral success and beyond. *Int. J. Doctoral Stud.* **6**, 51–78 (2011)
 32. Manathunga, C.: Early warning signs in postgraduate research education: a different approach to ensuring timely completions. *Teach. High. Educ.* **10**(2), 219–233 (2005)
 33. Retalis, S., et al.: Towards networked learning analytics—a concept and a tool. In: Banks, S., Vivien, H., Jones, C., Kemp, B., McConnell, D., Smith, C. (eds.) *Fifth International Conference on Networked Learning*, pp. 1–8. Lancaster University (2006)
 34. Koskenoja, M.: Factors supporting and preventing master thesis progress in mathematics and statistics. *Int. Electron. J. Math. Educ.* (2019)
 35. Styles, I., Radloff, A.: The synergistic thesis: student and supervisor perspectives. *J. Further High. Educ.* **25**(1), 97–106 (2001)
 36. de Kleijn, R.A., et al.: Master's thesis supervision: relations between perceptions of the supervisor–student relationship, final grade, perceived supervisor contribution to learning and student satisfaction. *Stud. High. Educ.* **37**(8), 925–939 (2012)
 37. Lindsay, S.: What works for doctoral students in completing their thesis? *Teach. High. Educ.* **20**(2), 183–196 (2015)
 38. de Kleijn, R.A., et al.: Master's thesis projects: student perceptions of supervisor feedback. *Assess. Eval. High. Educ.* **38**(8), 1012–1026 (2013)
 39. Jayaprakash, S.M., et al.: Early alert of academically at-risk students: an open source analytics initiative. *J. Learn. Anal.* **1**(1), 6–47 (2014)
 40. Kashyap, A., Nayak, A.: Different machine learning models to predict dropouts in MOOCs. In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE (2018)
 41. Gašević, D., Dawson, S., Siemens, G.: Let's not forget: learning analytics are about learning. *TechTrends* **59**(1), 64–71 (2015)
 42. Schumacher, C., Ifenthaler, D.: The importance of students' motivational dispositions for designing learning analytics. *J. Comput. High. Educ.* **30**(3), 599–619 (2018)
 43. Tempelaar, D., et al.: Student profiling in a dispositional learning analytics application using formative assessment. *Comput. Hum. Behav.* **78**, 408–420 (2018)
 44. Papamitsiou, Z., Economides, A.A.: Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *J. Educ. Technol. Soc.* **17**(4), 49–64 (2014)

45. Sedrakyan, G., et al.: Linking learning behavior analytics and learning science concepts: designing a learning analytics dashboard for feedback to support learning regulation. *Comput. Hum. Behav.* (2018)
46. Nouri, J., Larsson, K., Saqr, M.: Bachelor thesis analytics: using machine learning to predict dropout and identify performance factors. *Int. J. Learn. Anal. Artif. Intell. Educ.* **1**(1) (2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.