Utilising learning analytics for study success in higher education: A systematic review

Dirk Ifenthaler
University of Mannheim, Germany &
Curtin University, Australia

Jane Yin-Kim Yau
University of Mannheim, Germany

This study examined the utilisation of learning analytics to support study success in higher education. The main research question was to identify whether there is a link between learning analytics and the respective intervention measures to increase study success at higher education institutions. The systematic review included empirical studies conducted during the past five years. Search terms identified 6,220 articles from various scientific sources. After duplicated articles were removed, there were 3,163 articles remaining. Each of the articles were screened and the inclusion criteria (e.g., peer-reviewed, rigorous research findings) limited the key studies to 41 articles. This paper presents an overview of the results of this systematic review. It is concluded that evidence can be found supporting the use of learning analytics to support study success in higher education. However, study success may not be exclusively the result of the use of learning analytics but also some additional means of technological or institutional support. The findings also suggest a wider adoption of learning analytics systems as well as work towards standardisation of learning analytics procedures which can be integrated into existing digital learning environments.

Keywords: Learning analytics, study success, dropout, systematic review

Introduction

Mining data for insights to improve education enables an additional level of evidence-based research into learning and teaching. Currently, promising learning analytics applications are being developed which utilise data produced in the educational context (e.g., Pistilli & Arnold, 2010; Gašević, Dawson, Rogers, & Gašević, 2016). From a holistic point of view, learning analytics use static and dynamic information from digital learning environments, administrative systems, and social platforms for real-time modelling, prediction, and optimisation of learning processes, learning environments, and educational decision-making (Ifenthaler, 2015). Accordingly, learning analytics are expected to provide benefits for all involved stakeholders (i.e., students, teachers, designers, administrators, etc.) at higher education institutions. Various research methodologies and techniques are currently being implemented on different categories of learning analytics (such as descriptive, predictive and prescriptive) and offer different insights into the design and deployment at higher education institutions (Berland, Baker, & Bilkstein, 2014). Descriptive analytics use data obtained from sources such as course assessments, surveys, student information systems, learning management system activities, and forum interactions mainly for reporting purposes. Predictive analytics utilise similar data from those sources and attempts to measure onward learning success or failure. Prescriptive analytics deploy algorithms to predict commonly the study success and whether students retain on their courses as well as suggesting immediate interventions (Baker & Siemens, 2015). The main motivations of utilising learning analytics for higher education institutions include (a) improving students' learning and their motivation in learning, hence, retaining their studies on courses and reducing dropout (or inactivity), as well as (b) attempting to improve the learner's learning process by providing personalised and adaptive learning pathways (toward specific goals set by the teacher or student). However, the success of learning analytics in improving higher education students' learning has yet to be proven systematically and based on rigorous empirical findings. Only a few works have tried to address this but limited evidence is shown (Suchithra, Vaidhehi, & Iyer, 2015). The current study aims to form a systematic review of empirical evidence demonstrating how learning analytics have been successful in facilitating study success in continuation and completion of students' university courses. The overriding research question is as follows: Is it possible to identify a link between learning analytics and related prevention and intervention measures to increase study success in international empirical studies?

Study success and learning analytics

Study success includes the successful completion of a first degree in higher education to the largest extent, and the successful completion of individual learning tasks to the smallest extent (Sarrico, 2018). The essence here is to capture any positive learning satisfaction, improvement, or experience during learning. As some of the more common and broader definitions of study success include terms such as retention, persistence, graduation rate and the opposing terms include withdrawal, dropout, non-completion, attrition and failure (Mah, 2016).

Learning analytics show promise to enhance study success in higher education (Pistilli & Arnold, 2010). For example, students often enter higher education academically unprepared and with unrealistic perceptions and expectations of academic competencies for their studies. Both, the inability to cope with academic requirements as well as unrealistic perceptions and expectations of university life, in particular with regard to academic competencies, are important factors for leaving the institution prior to degree completion (Mah, 2016). However, Sclater and Mullan (2017) reported on the difficulty to isolate the influence of the use of learning analytics, as often they are used in addition to wider initiatives to improve student retention and academic achievement.

An extensive systematic literature review of empirical evidence on the benefits of learning analytics as well as the related field of educational data mining was conducted by Papamitsiou and Economides (2014). They classified the findings from case studies focussing on student behaviour modelling, prediction of performance, increase self-reflection and self-awareness, prediction of dropout as well as retention. Their findings suggest that large volumes of educational data are available and that pre-existing algorithmic methods are applied. Further, learning analytics enable the development of precise learner models for guiding adaptive and personalised interventions. Additional strengths of learning analytics include the identification of critical instances of learning, learning strategies, navigation behaviours, and patterns of learning (Papamitsiou & Economides, 2014). Another related systematic review on learning analytics was conducted by Kilis and Gülbahar (2016). They conclude from the reviewed studies that log data of student's behaviour needs to be enriched with additional information (e.g., actual time spent for learning, semantic rich information) for better supporting learning processes. Hence, learning analytics for supporting study success requires rich data about students' efforts and performance as well as detailed information about psychological, behavioural and emotional states.

Method

Preparation and literature search

The preparation of the systematic review followed the eight steps proposed by Okoli and Schabram (2010). First, the goal for the systematic review, i.e., an overview of international studies utilising learning analytics to support study success was defined. Second, the inclusion criteria for the studies were determined. Third, international databases including Google Scholar, ACM Digital Library, Web of Science, Science Direct, ERIC and DBLP were searched. Search terms included "learning analytics" in combination with "study success", "retention", "dropout prevention", "course completion", and "attrition". In addition, specific journals such as Journal of Learning Analytics, Computers in Human Behaviour, Computers & Education, Australasian Journal of Educational Technology and British Journal of Educational Technology were searched. Searches were conducted using the above-mentioned terms, matched to the databases' subject headings and as keywords in the title and abstract. Fourth, the selection criteria, i.e., clear theoretical foundation, research methodology and implications were defined. Fifth, benchmarks for quality assessment, i.e., presentation of findings, sampling technique, methodological procedure were set. Sixth, relevant information from the individual studies (e.g., empirical evidence, implications) were extracted. Seventh, the extracted information from the individual studies were merged together to draw conclusions. Finally, the findings of the systematic review are being reported in this contribution.

Inclusion criteria

Retrieved articles were restricted to studies that (a) were situated in the higher education context, (b) were published between January 2013 and December 2017, (c) were published in English language, (d) had an abstract available, (e) presented either qualitative or quantitative analyses and findings, and (f) were peer-reviewed. 6,220 articles were located and after duplicated papers were removed, 3,163 articles were remaining. The number of key studies identified was 374 (in the first round) then limited to 41 (due to substantiality of empirical evidence).

Results

Summary of key studies

The 41 key studies included in this systematic review were conducted in USA (*n* = 13), Australia (*n* = 6), Brazil (*n* = 2), Ireland (*n* = 2), UK (*n* = 2), Taiwan (*n* = 2), India (*n* = 2), South Korea (*n* = 2), Sweden (*n* = 1), Israel (*n* = 1), The Netherlands (*n* = 1), Pakistan (*n* = 1), Columbia (*n* = 1), France (*n* = 1), Spain (*n* = 1), Japan (*n* = 1), and Saudi Arabia (*n* = 1). Most articles were published in 2017 (17) followed by five articles in 2016, 2015 (7), 2014 (8), and four articles in 2014.

The key studies utilised adequate data analytics methods such as binary logistic regression, decision tree analysis, support vector machines, logistic regression and classification systems. Many of the key studies applied several statistical methods in order to determine which one can achieve the most accurate prediction of study success and/or dropout. The main predictions forecasted in the key studies were on course completion, grades to be obtained, and dropout. In addition, empirical evidence from articles which were not eligible to form the key studies in this systematic review (due to incomplete work/lack of depth) are available. Still, these studies provide additional supporting evidence in the ways learning analytics can be used to increase study success.

Evidence of learning analytics supporting study success

Table 1 shows example findings of the systematic review (the complete table of findings will be part of the conference presentation) including the bibliographic information, origin of the study, sample size, data collection and analysis, as well as key indicators.

Table 1: Example findings of selected Australian key studies

Author(s) & Year	Country	Sample size	Data collection sources	Data analysis methods	Key indicators
Chai & Gibson (2015)	Australia	23,291	University datasets	Cross-validation technique	Three types of machine learning techniques were tested with a focus on retention: Logistic regression, decision trees and random forests. The models were evaluated using precision and recall metrics. Logistic regression gained the best performance and user utility (*67% precision, 29% recall*).
Dawson, Jovanovic, Gasevic, & Pardo (2017)	Australia/ UK	11,160	Student information system, LMS interactions and assessment	Common statistical methods	Positive association between the intervention and student retention was identified using common statistical analysis. Higher variability in the data (over 99%) can be achieved using more advanced statistical methods, e.g., mixed-effect methods.
Rogers, Colvin, & Chiera (2014)	Australia	2,332	Variables from online systems (demographic, performance)	Regression	An index method was utilised which could make accurate predictions of dropout.

The 41 key studies provided the following evidence of learning analytics for supporting study success in higher education:

1. Study success can be achieved by students who utilised learning analytics interventions.
2. Engagement of students is a predictor of study success.
3. Recommender systems produce positive effects toward study success.
4. GPA and financial status characterise study success.
5. Predictive power / prediction accuracy can be manifested and increased through data on course completion, dropouts, achievement level, study achievement, total study time, interaction with colleagues, frequency of

regular learning intervals, and number of downloads from the learning environment.
6. Prediction accuracy for study success increases over time (80% from week 12 of a semester).
7. Reduction of dropout rates and prognosis of dropouts can be based on the specific courses attended.
8. Strong correlation between CGPA and pre-set grades. CGPA serves as a study success indicator.
9. Small positive relationship between student satisfaction with the use of the learning analytics dashboard and their study success.
10. Students who completed a course expected better performance feedback.
11. Various online learning systems are reliable and can produce solid predictions for study success.
12. Indexing methods can be utilised which produce accurate predictions of dropout and study success.

In addition, the findings of the integrative review provide insights into predictors to indicate study success, available predictive models (algorithms), pedagogical models toward personalized learning and integration of data visualisation for study success.

Discussion

Attrition and course incompletion rates often do not inform whether the student dropped out due to personal, financial, academic or course quality reasons. Different measures and intervention strategies need to be set in place and to individualise student support services for various learners due to the different reasons of dropout and also different intervention strategies may work for some and not for others (Mah, 2016; Sclater & Mullan, 2017). A high completion does not necessary inform of the quality of the course, whether students' learning outcomes were achieved and how motivated the students were intrinsically or extrinsically, causing more difficulties to support the student to complete the course.

The findings of this systematic review on learning analytics and study success highlights the demand of personalised learning environments as well as tailored education packages offered by higher education institutions. This requires rich data about the student's personal profile which includes information such as socio-demographic background, previous qualifications and academic achievements, engagement in the recruitment journey as well as dispositions about learning and motivation (Ifenthaler & Widanapathirana, 2014). In addition, the findings obtained in this systematic review so far suggest that there is a considerable number of sophisticated learning analytics tools which utilise effective techniques in predicting study success and at-risk students of dropping out. However, standards for implementation in productive digital learning environments at higher education institutions are missing. Further, as learning analytics are of growing interest for higher education institutions, it is important to understand students' expectations of learning analytics features to be able to align them with learning theory and technical possibilities before implementing them (Schumacher & Ifenthaler, 2018; Marzouk et al., 2016). One suggestion is to leverage existing learning theory by clearly designing (quasi-)experimental studies based on theoretical frameworks and connect learning analytics research with decades of previous research in education (Marzouk et al., 2016). The systematic review also indicated that over the past two years, several reviews and reports have been published which document policy recommendations established for policy-makers, administrators and course conveners. This line of research demands more work on ethical and privacy guidelines supporting learning analytics when implementing learning analytics at higher education institutions (e.g., West, Huijser, & Heath, 2016). Another demand is a well-facilitated change management within the higher education institution including institution-wide acceptance of learning analytics, the integration of all stakeholders as well as rigorous guidelines and policies focussing on data protection and ethics for learning analytics applications (Ifenthaler, 2017).

Conclusion

The findings obtained in this systematic review suggest that there is a considerable number of sophisticated learning analytics tools which utilise effective techniques in supporting study success and at-risk students of dropping out. Limitations of this study include the difficulty in comparing results of different studies as various techniques and algorithms, research questions and aims were used. Although much empirical evidence is documented in these articles, many studies are still works-in-progress, experimental studies, and limited in external validity. The key studies discuss how learning analytics can work to predict study success. However, the implications how to support stakeholders at higher education institutions in utilising learning analytics to support study success are under-documented. The questions raised are for example: Will students be able to respond positively and proactively when informed that their learning progress is hindered or inactivated?; Will instructors be able to influence the at-risk students positively so that they will re-engage with the studies? In addition, ethical dimensions regarding descriptive, predictive and prescriptive LA need to be addressed with further empirical studies and linked to study success indicators (West et al., 2016).

To sum up, empirical evidence on a large scale to support the effectiveness of learning analytics actually retaining students onto courses are still lacking. It is therefore imperative to leverage existing learning theory, psychological methods and connecting them to advances of learning analytics research for designing (quasi-)experimental studies including theoretical frameworks and sound empirical methodologies.

References

Baker, R. S., & Siemens, G. (2015). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2 ed., pp. 253–272). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

Berland, M., Baker, R. S., & Bilkstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19*(1–2), 205–220. doi:10.1007/s10758-014-9223-7

*Chai, K. E. K., & Gibson, D. C. (2015). Predicting the risk of attrition for undergraduate students with time based modelling. In D. G. Sampson, J. M. Spector, D. Ifenthaler, & P. Isaias (Eds.), *Proceedings of cognition and exploratory learning in the digital age* (pp. 109–116). Maynooth, Ireland: IADIS Press.

*Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In I. Molenaar, X. Ochoa, & S. Dawson (Eds.), *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 474–478). New York, NY: ACM.

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education, 28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Thousand Oaks, CA: Sage.

Ifenthaler, D. (2017). Are higher education institutions prepared for learning analytics? *TechTrends, 61*(4), 366–371. doi:10.1007/s11528-016-0154-0

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1–2), 221–240. doi:10.1007/s10758-014-9226-4

Kilis, S., & Gulbahar, Y. (2016). *Learning analytics in distance education: a systematic literature review*. Paper presented at the 9th European Distance and E-learning Network (EDEN) Research Workshop, Oldenburg, Germany.

Mah, D.-K. (2016). Learning analytics and digital badges: potential impact on student retention in higher education. *Technology, Knowledge and Learning, 21*(3), 285–305. doi:10.1007/s10758-016-9286-8

Marzouk, Z., Rakovic, M., Liaqat, A., Vytasek, J., Samadi, D., Stewart-Alonso, J., . . . Nesbit, J. C. (2016). What if learning analytics were based on learning science? *Australasian Journal of Educational Technology, 32*(6), 1–18. doi:10.14742/ajet.3058

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems, 10*(26).

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.

Pistilli, M. D., & Arnold, K. E. (2010). Purdue Signals: Mining real-time academic data to enhance student success. *About campus: Enriching the student learning experience, 15*(3), 22–24.

*Rogers, T., Colvin, C., & Chiera, B. (2014). Modest analytics: using the index method to identify students at risk of failure. In A. Pardo & S. D. Teasley (Eds.), *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 118–122). New York, NY: ACM.

Sarrico, C. S. (2018). Completion and retention in higher education. In T. P. & S. J. (Eds.), *Encyclopedia of international higher education systems and institutions*. Dordrecht: Springer.

Sclater, N., & Mullan, J. (2017). *Learning analytics and student success – assessing the evidence*. Bristol: JISC.

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior, 78*, 397–407. doi:10.1016/j.chb.2017.06.030

Suchithra, R., Vaidhehi, V., & Iyer, N. E. (2015). survey of learning analytics based on purpose and techniques for improving student performance. *International Journal of Computer Applications, 111*(1), 22–26.

West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. *Educational Technology Research and Development, 64*(5), 903–922. doi:10.1007/s11423-016-9464-3

Please cite as: Ifenthaler, D. & Yau, J. (2018). Utilising learning analytics for study success in higher education: A systematic review. In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri & J. Tai (Eds.), Open Oceans: Learning without borders. Proceedings ASCILITE 2018 Geelong (pp. 406-411). https://doi.org/10.14742/apubs.2018.1958