ASCILITE 2020
ASCILITE's first Virtual Conference

Understanding Learning Analytics Indicators for Predicting Study Success

Dirk Ifenthaler
University of Mannheim & Curtin University

Jane Yin-Kim Yau
University of Mannheim

Common factors, which are related to study success include students' sociodemographic factors, cognitive capacity, or prior academic performance, and individual attributes as well as course related factors such as active learning and attention or environmental factors related to supportive academic and social embeddedness. The aim of this research is to gain a deeper understanding of not only if learning analytics can support study success, but which aspects of a learner's learning journey can benefit from the utilisation of learning analytics. We, therefore, examined different learning analytics indicators to show which aspect of the learning journey they were successfully supporting. Key indicators may include GPA, learning history, and clickstream data. Depending on the type of higher education institution, and the mode of education (face-to-face and/or distance), the chosen indicators may be different due to them having different importance in predicting the learning out-comes and study success.

Keywords: learning analytics; study success, indicator, analytics method

Introduction

Research focusing on learning analytics is still rapidly evolving with most of the respective implementations being located in UK, USA and Australia (Ifenthaler, Mah, & Yau, 2019; Sclater, Peasgood, & Mullan, 2016). Although in the last five years, there has been an increase of the number of related research, large-scale empirical evidence regarding the effectiveness of learning analytics remain to be seen (Mah, Yau, & Ifenthaler, 2019; J. Wong et al., 2019). The field arose originally as a result of the increasing availability of educational data, and the phenomenon that a significant proportion of first year university students do not complete their courses (Prieto, Rodríguez-Triana, Martínez-Maldonado, Dimitriadis, & Gašević, 2019). A number of benefits arising from learning analytics include the identification of at-risk students (Azcona, Hsiao, & Smeaton, 2019; Lawson, Beer, Rossi, Moore, & Fleming, 2016), the possibility of constructing adaptive support of students' learning journeys (Ifenthaler, 2015; Ma, Adespope, Nesbit, & Liu, 2014) or providing students with additional support for coping with academic requirements and expectations (Mah & Ifenthaler, 2017, 2020). Therefore, study success is conceptualised as 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).

However, small-scale empirical evidence regarding the effectiveness of learning analytics for supporting study success has been located as presented in a recent systematic review (Ifenthaler & Yau, 2020) as well as in several other review articles (Banihashem, Aliabadi, Pourroostaei Ardakani, Delaver, & Nili Ahmadabadi, 2018; Larsen, Kornbeck, Larsen, Kristensen, & Sommersel, 2013; Papamitsiou & Economides, 2014; Vieira, Parsons, & Byrd, 2018). The aim of this research is to gain a deeper understanding of not only if learning analytics can support study success, but which aspects of a learner's learning journey can benefit from the utilisation of learning analytics. We, therefore, examined different learning analytics indicators to show which aspect of the learning journey they were successfully supporting. The guiding research question is therefore: *Which learning analytics indicators work best for what purposes within a learner's learning journey?*

Background

The success of students at higher education institutions has been a global concern for many years (Tinto, 2005). Even though many academic support programs have been implemented (Padgett, Keup, & Pascarella, 2013), and research on study success is extensive (Attewell, Lavin, Domina, & Levey, 2006; Bijsmans & Schakel, 2018; Morosanu, Handley, & O'Donovan, 2010; Schmied & Hänze, 2015), dropout rates in higher education remain at about 30% in the Organization for Economic Cooperation and Development member countries (OECD, 2019). Factors that contribute to student success, which may influence a student's decision to discontinue higher education are various and complex (Tinto, 1982, 2005). Important factors for dropouts that have been consistently found in international studies include the choice of the wrong study program, lack of motivation, personal circumstances, an unsatisfying first-year experience, lack of university support services, and academic unpreparedness (Heublein, 2014; Thomas, 2002; Willcoxson, Cotter, & Joy, 2011; Yorke & Longden, 2008).

Common factors, which are related to study success include students' sociodemographic factors (e.g., gender, ethnicity, family background), cognitive capacity, or prior academic performance (e.g., grade point average [GPA]), and individual attributes (e.g., personal traits, and motivational or psychosocial contextual influences) as well as course related factors such as active learning and attention or environmental factors related to supportive academic and social embeddedness (Bijsmans & Schakel, 2018; Brahm, Jenert, & Wagner, 2017; Remedios, Clarke, & Hawthorne, 2008; Tinto, 2017). The possibility to collect and store data for the above mentioned factors and combining them in (near) real-time analysis opens up advanced evidence-based opportunities to support study success utilising meaningful interventions referred to as learning analytics (Pistilli & Arnold, 2010).

The concept of learning analytics has been used in various contexts and with various focal points, resulting in a lack of clarity and precise definition. For instance, B. T. M. Wong (2017) presents several case studies utilising learning analytics for (a) improving student retention, (b) supporting informed decision making, (c) increasing cost-effectiveness, (d) helping to understand learning behaviour, (e) providing personalised assistance, and (f) delivering feedback and interventions. Further, an extensive diversification of the initial learning analytics approaches can be documented (Prieto et al., 2019). These learning analytics approaches apply various methodologies, such as descriptive, predictive, and prescriptive analytics to offer different insights into learning and teaching (Berland, Baker, & Bilkstein, 2014). Verbert, Duval, Klerkx, Govaerts, and Santos (2013) identified two distinctive approaches of learning analytics: (1) data visualisation (i.e., showing data for improving students learning), and (2) predictive analytics (i.e., highlighting areas for improvement that forecast study success/retention). In order to overcome the current conceptual ambiguity, learning analytics with a specific focus on higher education and their link to study success have been defined as the use, assessment, elicitation and analysis of static and dynamic information about learners and learning contexts, for the near real-time modelling, prediction and optimisation of learning processes, and learning environments, as well as for educational decision-making (Ifenthaler, 2015, p. 447). Still, studies completed on which learning analytics indicators would best fit the different purposes of learning success prediction such as student grades, student engagement, student behaviour, student performance and course completion are scarce.

Method

This paper presents a secondary analysis approach of a previously conducted systematic review which followed the eight steps proposed by Okoli and Schabram (2010). Hence, from our previous completed systematic review of studies derived from high-quality academic journals and conference proceedings (Ifenthaler et al., 2019), we formulated a list of 49 studies to inform whether there is empirical evidence that the general use of learning analytics could improve study success. Even though there were 3,163 articles that contained the required search terms "learning analytics" in combination with "study success", "retention", "dropout prevention", "course completion", and "attrition", the actual number of articles which fitted our inclusion criteria a) higher education context, b) published between January 2013 and December 2019, c) written in English language, d) had substantial qualitative or quantitative analyses and findings, and e) were peer-reviewed. The findings showed that only a small number of identified articles were implemented into higher educational institutions successfully with a tangible positive increase of study success. The secondary analysis of these articles specifically focusses on learning analytics indicators utilised for supporting study success. The research team developed a research protocol, which described the individual steps of conducting the secondary analysis and validated the research protocol in a training session focusing on database handling, reviewing, and note-taking techniques. The full text analysis of the remaining publications focused on the theoretical rigor of the key publications. This synthesis of key publications followed the triangulation approach, as the final studies included quantitative and qualitative studies (Okoli, 2015).

Results

From the 49 studies, five categories of predictions were formulated: (1) student grades, (2) student social learning behaviour/engagement, (3) at-risk/low-performers, (4) student performance, and (5) course completion.

The applied learning analytics indicators for the five categories are presented in the following subsections.

1. Two out of the 49 studies aimed to predict the exact grades/answers of students' assignments. As the targeted information to be obtained is very precise, a full detailed range of information is also required. For example, Thompson (2013) used transcription, extraction and analysis of video and audio recordings utilising discourse captured in video and audio recordings. Key indicators: Video, audio and digital pen input data.
2. Eight out of the 49 studies aimed to obtain more generic social learning behaviour such as their engagement/participation, any relating study patterns. These indicators can also be used for checking and confirming student attendance. Methods such as social network analysis, latent class analysis, descriptive statistics and correlation analysis, data mining techniques, group behaviour analysis, mean-generation task, visualisation and multi-level modelling were popular. For example, Hu, Lo, and Shih (2014) administered data mining techniques, classification and regression tree, system usability survey (self-report questionnaire) techniques utilising a dataset of completed learning activities. Key indicators: login, total reading time, homework delay and forum activity.
3. The majority of studies (N = 20) were focused on locating at-risk students. Techniques in combination with various datasets were utilised such as binary classification problem, basic and extended pass-fail classifier, cross-validation techniques, data examination, logistic regression, sequence model, feature vector model, binary classifiers, probabilistic models, chi-squared and machine learning. One example stems from Bukralia, Deokar, and Sarnikar (2014) who administered binary classification problem, descriptive statistics and data mining techniques (obtained from the Variables from Student Information Systems and Course Management System). Key indicators include: academic ability, financial support, academic goals, technology preparedness, demographics, course engagement and motivation, and course characteristics.
4. Thirteen out of the 49 studies focused on the overall student performance and achievement. Similar methods as in subsection 4.3 were used in this category. For example, Conijn, Snijders, Kleingeld, and Matzat (2017) computed correlation analysis and multi-level analyses with cross-random effects, multiple linear regressions techniques on datasets of students' online behaviour (from Moodle LMS). Key indicators: LMS data and assessment data (including in-between grades, final exam grades and overall course grade).
5. Six out of the 49 studies were focused on overall student course completion. Similar indicators identified in section 4.3 and 4.4 were utilised. An example is reported by Aulck et al. (2017) using machine learning experiments with university databases containing demographic and pre-college entry information, e.g., standardised test scores, high school grades, parents' educational attainment, application zip code and complete transcript records (these variables also forming the key indicators).

Discussion

The following five recommendations for specific indicators focus on (1) task related predictions, (2) social, learning or engagement behaviour, (3) low-performing or dropout students, (4) general or overall performance of students, and (5) course completion.

1. For predicting the correctness of answers/grades – indicators such as videos and clickstream data are useful. Methods such as transcription, extraction and analysis of video and audio recordings are helpful.
2. For predicting social learning behaviour – indicators such as study-related, social behaviour, lecture attendance, material and forum activity, and study patterns are useful. Methods such as social network analysis, latent class analysis, descriptive statistics and correlation analysis, data mining techniques, group behaviour analysis, mean-generation task, visualisation and multi-level modelling mostly utilising datasets are helpful.
3. For predicting at-risk students – indicators such as online activity, academic ability and goals, motivation, interaction with other students, socioeconomic status, test performance, study load, demographic information are useful. Methods such as binary classification problem, basic and extended pass-fail classifier, cross- validation techniques, data examination, logistic regression, sequence model, feature vector model, binary classifiers, probabilistic models, chi-squared and machine learning are helpful.
4. For predicting student performance – indicators such as exam grades, interaction with others, forum activity, completed assignments, dashboard usage frequency, learning achievement and academic history. Methods such as statistical, correlational analyses, support vector machines, multi-regression model, multiple linear regression, recurrent neural network, cross-validation techniques, risk detection algorithms are useful.
5. For predicting student course completion – indicators such as frequency of posts, LMS engagement,

students’ activity and forum interactions are useful. Methods such as binary logistic regression, machine learning, common statistical analysis, stepwise logistic regression techniques and confirmatory factor analysis are helpful.

A number of research/implementation directions were made clear and concluded from our study. These include (as) the standardisation of learning analytics systems ready for institutions to adopt without the need for each one to implement their own; (b) additional personalised prevention and intervention strategies for different study programmes fitting to different requirements in various institutions with the awareness that the standardised system may need to be adjusted; (c) elaborating from (b), individual tailored learning packages optimised for each learner based on their profile (e.g., geo-social demographic backgrounds, qualifications, learning journey engagement, website activities, search information); (d) more work on privacy and ethical guidelines; (e) quality assurance of learning analytics systems and related recommendations including an accreditation body; rigorous multidisciplinary research focussing on (quasi)experimental studies and longitudinal designs for producing robust findings regarding the effectiveness of learning analytics for learning and teaching.

Acknowledgement

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany (BMBF, project number 16DHL1038).

References

Attewell, P. A., Lavin, D. E., Domina, T., & Levey, T. (2006). New evidence on college remediation. *The Journal of Higher Education, 77*(5), 886–924. doi:10.1353/jhe.2006.0037

Aulck, L., Aras, R., Li, L., L’Heureux, C., Lu, P., & West, J. (2017). *Stemming the tide: Predicting STEM attrition using student transcript data*. Paper presented at the Machine Learning for Education, Halifax, Nova Scotia, Canada.

Azcona, D., Hsiao, I., & Smeaton, A. F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students’ digital footprints. *User Modeling and User-Adapted Interaction, 29*, 759–788. doi:10.1007/s11257-019-09234-7

Banihashem, S. K., Aliabadi, K., Pourroostaei Ardakani, S., Delaver, A., & Nili Ahmadabadi, M. (2018). Learning analytics: a systematic literature review. *Interdisciplinary Journal of Virtual Learning in Medical Sciences, 9*(2), 1–10. doi:10.5812/ijvlms.63024

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

Bijsmans, P., & Schakel, A. H. (2018). The impact of attendance on first-year study success in problem-based learning. *Higher Education, 76*, 865–881. doi:10.1007/s10734-018-0243-4

Brahm, T., Jenert, T., & Wagner, D. (2017). The crucial first year: a longitudinal study of students’ motivational development at a Swiss Business School. *Higher Education, 73*(3), 459–478.

Bukralia, R., Deokar, A., & Sarnikar, S. (2014). Using academic analytics to predict dropout risk in e-Learning courses. In L. S. Iyer & D. J. Power (Eds.), *Reshaping Society through Analytics, Collaboration and Decision Support* (pp. 67–93). Cham: Springer.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: a comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17–29.

Heublein, U. (2014). Student drop-out from German higher education institutions. *European Journal of Education, 49*(4), 497–513.

Hu, Y., Lo, C., & Shih, S. (2014). Developing early warning systems to predict students’ online learning performance. *Computers in Human Behavior, 36*, 469–478.

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., Mah, D.-K., & Yau, J. Y.-K. (2019). Utilising learning analytics for study success. Reflections on current empirical findings. In D. Ifenthaler, J. Y.-K. Yau, & D.-K. Mah (Eds.), *Utilizing Learning Analytics to Support Study Success* (pp. 27–36). Cham: Springer.

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development, 68*(4), 1961–1990. doi:10.1007/s11423-020-09788-z

Larsen, M. S., Kornbeck, K. P., Larsen, M. B., Kristensen, R. M., & Sommersel, H. (2013). *Dropout phenomena at universities: What is dropout? Why does dropout occur? What can be done by the universities to prevent or reduce it? A systematic review*. Copenhagen: Danish Clearinghouse for Educational Research.

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development, 64*(5), 957–968. doi:10.1007/s11423-016-9459-0 Ma, W., Adespope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: a meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918.

Mah, D.-K., & Ifenthaler, D. (2017). Academic staff perspectives on first-year students' academic competencies. *Journal of Applied Research in Higher Education, 9*(4), 630–640. doi:10.1108/JARHE-03- 2017-0023

Mah, D.-K., & Ifenthaler, D. (2020). What do first-year students need? Digital badges for academic support to enhance student retention. *Journal of Applied Research in Higher Education, 12*(1), 86–96. doi:10.1108/JARHE-12-2018-0258

Mah, D.-K., Yau, J. Y.-K., & Ifenthaler, D. (2019). Future directions on learning analytics to enhance study success. In D. Ifenthaler, J. Y.-K. Yau, & D.-K. Mah (Eds.), *Utilizing Learning Analytics to Support Study Success* (pp. 313–321). Cham: Springer.

Morosanu, L., Handley, K., & O'Donovan, B. (2010). Seeking support: Researching first-year students' experiences of coping with academic life. *Higher Education Research & Development, 29*(6), 665–678. doi:10.1080/07294360.2010.487200

OECD. (2019). *Education at a glance 2019: OECD indicators*. Paris: OECD Publishing.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems Research, 37*(43), 879–910. doi:10.17705/1CAIS.03743

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), 1–49.

Padgett, R. D., Keup, J. R., & Pascarella, E. T. (2013). The impact of first- year seminars on college students' life-long learning orientations. *Journal of Student Affairs Research and Practice, 50*(2), 133–151.

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.

Prieto, L. P., Rodríguez-Triana, M. J., Martínez-Maldonado, R., Dimitriadis, Y., & Gašević, D. (2019). Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology, 35*(4), 14–33. doi:10.14742/ajet.4314

Remedios, L., Clarke, D., & Hawthorne, L. (2008). The silent participant in small group collaborative learning contexts. *Active Learning in Higher Education, 9*(3), 201–216.

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.

Schmied, V., & Hänze, M. (2015). The effectiveness of study skills courses: Do they increase general study competences? *Zeitschrift für Hochschulentwicklung, 10*(4), 176–187.

Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education: A review of UK and international practice*. Bristol: JISC.

Thomas, L. (2002). Student retention in higher education: The role of institutional habitus. *Journal of Education Policy, 17*(4), 423–442. doi:10.1080/02680930210140257

Thompson, K. (2013). *Using micro-patterns of speech to predict the correctness of answers to mathematics problems: an exercise in multimodal learning analytics*. Paper presented at the ICMI '13: Proceedings of the 15th ACM on International conference on multimodal interaction, Sydney, NSW, Australia.

Tinto, V. (1982). Limits of theory and practice in student attrition. *The Journal of Higher Education, 53*(6), 687–700.

Tinto, V. (2005). Reflections on student retention and persistence: moving to a theory of institutional action on behalf of student success. *Studies in Learning, Evaluation, Innovation and Developmental Psychology, 2*(3), 89–97.

Tinto, V. (2017). Through the eyes of students. *Journal of College Student Retention: Research, Theory and Practice, 19*(3), 254–269.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509. doi:10.1177/0002764213479363

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education, 122*, 119–135.

doi:10.1016/j.compedu.2018.03.018 Willcoxson, L., Cotter, J., & Joy, S. (2011). 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.
Wong, B. T. M. (2017). Learning analytics in higher education: an analysis of case studies. *Asian Association of Open Universities Journal, 12*(1), 21–40. doi:10.1108/AAOUJ-01-2017-0009
Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., . . . Paas, F. G. (2019). Educational theories and learning analytics: from data to knowledge. In D. Ifenthaler, J. Y.-K. Yau, & D.-K. Mah (Eds.), *Utilizing learning analytics to support study success* (pp. 3–25). Cham: Springer.
Yorke, M., & Longden, B. (2008). *The first-year experience of higher education in the UK*. York: The Higher Education Academy.

Ifenthaler, D., & Yin-Kim Yua, J. (2020). Understanding Learning Analytics Indicators for Predicting Study Success. In S. Gregory, S. Warburton, & M. Parkes (Eds.), ASCILITE's First Virtual Conference. Proceedings ASCILITE 2020 in Armidale (pp. 36–41). https://doi.org/10.14742/ascilite2020.0101