

# Reflections on Different Learning Analytics Indicators for Supporting Study Success

<https://doi.org/10.3991/ijai.v2i2.15639>

Jane Yin-Kim Yau

University of Mannheim, Mannheim, Germany

Dirk Ifenthaler <sup>(✉)</sup>

University of Mannheim, Mannheim, Germany

Curtin University, Perth, Australia

[dirk@ifenthaler.info](mailto:dirk@ifenthaler.info)

**Abstract**—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. In addition, there are various stages of a learner's learning journey from the beginning when commencing learning until its completion, as well as different indicators or variables that can be examined to gauge or predict how successfully that journey can or will be at different points during that journey, or how successful learners may complete the study and thereby acquiring the intended learning outcomes. 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 outcomes and study success.

**Keywords**—Learning analytics; study success; analytics methods; student-at-risk; dropout

## 1 Introduction

Research focusing on learning analytics is still rapidly evolving with most of the respective implementations being located in UK, USA and Australia [1, 2]. 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 [3, 4]. 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 [5]. Hence, a number of benefits arising from learning analytics include the identification of at-risk students [6, 7], the possibility of constructing adaptive support of students' learning journeys [8, 9] or providing students with additional support for coping with academic requirements and expectations [10, 11]. Accordingly, 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 [12]. However, only small-scale empirical evidence regarding the effectiveness of learning analytics for supporting study success has been located as presented in a recent systematic review [13] as well as in several other review articles [14-17].

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. Following a data profiles approach [18], the following research questions guides the present study: Which learning analytics indicators help to determine and support the study success in higher education while classifying them into student, learning, and curriculum profiles?

The remainder of the paper is organised as follows – Section 2 presents a literature review focusing on study success and learning analytics; Section 3 presents the research methodology we undertook for this paper; Section 4 presents the results; Section 5 presents a discussion and recommendations; and finally Section 6 presents a conclusion and future work.

## **2 Literature Review**

The success of students at higher education institutions has been a global concern for many years [19]. Even though many academic support programmes have been implemented [20], and research on study success is extensive [21-24], dropout rates in higher education remain at about 30% in the Organization for Economic Cooperation and Development member countries [25]. Factors that contribute to student success, which may influence a student's decision to discontinue higher education are various and complex [19, 26]. Important factors for dropouts that have been consistently found in international studies include the choice of the wrong study programme, lack of motivation, personal circumstances, an unsatisfying first-year experience, lack of university support services, and academic unpreparedness [27-30].

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 [24, 31-33]. 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 [34].

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, Wong [35] 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 [5]. These learning analytics approaches apply various methodologies, such as descriptive, predictive, and prescriptive analytics to offer different insights into learning and teaching [36]. 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 [8].

From a data management perspective, three distinctive data profiles have been identified [18]: student profile, learning profile, and curriculum profile (see Figure 1).

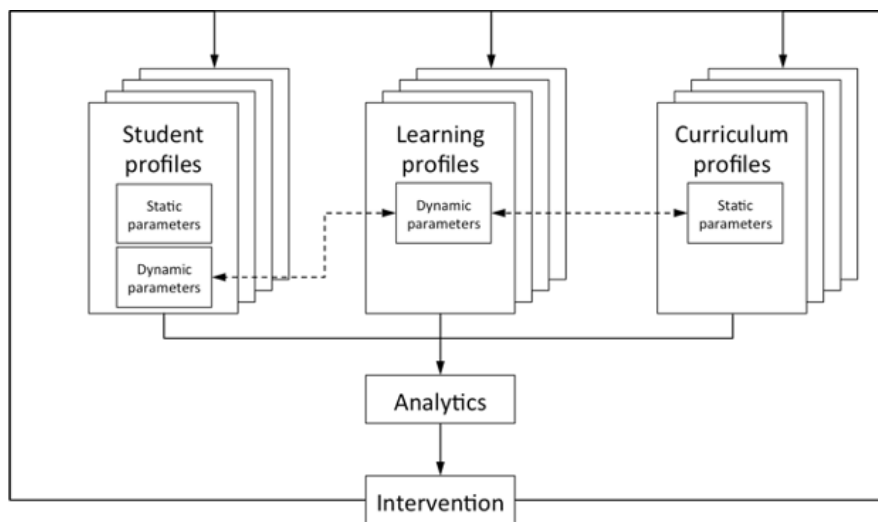


Fig. 1. Distinctive data profiles for learning analytics applications

The student profile includes static and dynamic indicators. Static indicators include gender, age, education level and history, work experience, current employment status, etc. Dynamic indicators include interest, motivation, response to reactive inventories (e.g., learning strategies, achievement motivation, emotions), computer and social media competencies, enrolments, drop-outs, pass/fail rate, academic performance, etc.

The learning profile includes indicators reflecting the current behaviour and performance within the learning environment (e.g., learning management system). Dynamic

indicators include trace data such as time specific information (e.g., time spent on learning environment, time per session, time on task, time on assessment). Other indicators of the learning profile include login frequency, task completion rate, assessment activity, assessment outcome, learning material activity (upload/download), discussion activity, support access, ratings of learning material, assessment, support, effort, etc.

The curriculum profile includes indicators reflecting the expected and required performance defined by the learning designer and course creator. Static indicators include course information such as facilitator, title, level of study, and prerequisites. Individual learning outcomes are defined including information about knowledge type (e.g., content, procedural, causal, meta cognitive), sequencing of materials and assessments, as well as required and expected learning activities.

The available data from all data profiles are analysed using pre-defined analytic models allowing summative, real-time, and predictive comparisons. The results of the comparisons are used for specifically designed interventions which are returned to the corresponding profiles. The (semi-)automated interventions include reports, dashboards, prompts, and scaffolds for different stakeholders (e.g., teachers, students, administrators). Additionally, stakeholders receive customised messages for following up with critical incidents (e.g., students at risk, assessments not passed, satisfaction not acceptable, etc.).

The above described data profiles have been utilised in different learning analytics applications and systems [1]. However, 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. Accordingly, the following research seeks to investigate international research studies with regard to the effectiveness of learning analytics indicators for determining and supporting higher education students learning journey while classifying them based on the three data profiles.

### **3 Method**

This article presents a secondary analysis approach of a previously conducted systematic review which followed the eight steps proposed by Okoli and Schabram [37]. Hence, from our previous completed systematic review of studies derived from high-quality academic journals and conference proceedings [2], 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. In other words, learning analytics were effective in retaining students and decreasing student dropout.

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 focussing 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. The research team used a quantitative and qualitative content analysis as well as reflective exchange to extract the findings of the key studies. This synthesis of key publications followed the triangulation approach, as the final studies included quantitative and qualitative studies [38]. The final step of conducting the secondary analysis included the dissemination of the findings through the writing of this paper, which documents the findings and discussion of implications as well as obvious limitations.

## **4 Results**

From the 49 studies, five categories of predictions were formulated: (1) student answers/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. The results are summarised in Table 1 which includes indicators of the five categories mapped to the three data profiles (student, learning, curriculum) as described above.

**Table 1.** Summary of learning analytics indicators mapped to three data profiles

	<b>Student profile</b>	<b>Learning profile</b>	<b>Curriculum profile</b>
Students answers/grades	N/A	Content access (video/ audio trace data) pen trace data (self-)assessment (score, grade, completion) data	N/A
Students social learning behaviour/engagement	Prior academic performance prior competence/skills demographic background social behaviour trait self-report survey current workload study pattern	Course access (login) content access discussion/forum (length, quality) trace data engagement trace data (self-)assessment (score, grade, completion) data	N/A
At-risk/ low-performers	Prior academic performance prior competence/skills demographic background socioeconomic background academic goals technology preparedness Completed/ withdrawn courses motivation/interest prior learning behaviour prior academic institutions enrolment history/ mode/ load	Course access (login) content access assignment submission engagement trace data discussion/forum (length, quality) trace data (Self-)assessment (score, grade, completion) data final grade reflection/ feedback access social network usage	Course characteristics course survey
Student performance	Prior academic performance demographic background socioeconomic background enrolment history/ mode/ load counselling activities psychological test outcomes	(Self-)assessment (score, grade, completion) data final grade course access content access discussion/forum (length, quality) trace data engagement trace data	N/A
Course completion	Prior academic performance demographic background completed/ withdrawn courses enrolment history/ mode/ load	Course access (login) content access discussion/forum (length, quality) trace data engagement trace data (self-)assessment (score, grade, completion) data	N/A

The following sections outline the individual studies from which the indicators were drawn from as well as additional information regarding the utilised data analytics methods.

#### 4.1 Indicators for predicting the correctness of answers/grades

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.

1. Thompson [39] 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 trace data.
2. Yang, et al. [40] applied statistical analysis (such as means and standard deviation) utilising clickstream data for video. Key indicators: trace data (clickstream) and assessment data.

#### **4.2 Indicators for predicting students' social/learning behaviour including engagement**

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.

1. Bydzovska and Popelinsky [41] applied social network analysis utilising student datasets including study-related, social behaviour and data concerning previously passed courses (key indicators being the two aforementioned variables).
2. Carroll and White [42] applied latent class analysis utilising datasets on lecture attendance, tutorial attendance, online scheduled access, print access, online full access to learning materials (key indicators being the five aforementioned variables).
3. Gong, et al. [43] used descriptive statistics and correlation analysis with quantitative self-report (student engagement questionnaire) and quantitative observation measures (number of viewing records and posts to a discussion board). Key indicators: student engagement and achievement.
4. Hu, et al. [44] 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.
5. Labarthe, et al. [45] conducted group behaviour analysis, i.e., a recommender panel was integrated into the experimental users' interface, which enabled them to manage contacts, send instant messages or consult their profiles. They utilised learning traces as interaction logs and demographic information from questionnaires. Key indicators: attendance, completion, scores and participation.
6. Lu, et al. [46] applied statistical analysis such as t-test utilising questionnaires and skills post-tests. Key indicators: programming skills, and number, length and quality of discussion.
7. Nam, et al. [47] used a mean-generation task via log data (a total of 1,500 items including free-text responses). Key indicators: familiarity concerning the learning tasks, previous grades and level of skill.
8. Nguyen, et al. [48] applied visualisation and multi-level modelling utilising online VLE. Key indicators: time spent on the VLE, actual workload in hours, study patterns and performance level.

### **4.3 Indicators for predicting at-risk/low-performing students**

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.

1. Aguiar, et al. [49] used feature selection methods (information gain, gain ratio, chi-squared & Pearson's correlation) and classification methods (naïve Bayes, decision trees, logistic regression, random forest). These methods utilised electronic portfolios. Key indicators: logins, submissions and hits.
2. Bukralia, et al. [50] 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.
3. Cambuzzi, et al. [51] present a case study utilising datasets from distance education. Key indicators: interactions between students in forums.
4. Casey and Azcona [52] used basic and extended pass-fail classifier, linking key-stroke metrics techniques. Key indicators: number of successful and failed compilations, number of connections, time spent, and slides coverage.
5. Chai and Gibson [53] deployed cross-validation techniques. Key indicators: logins for materials access, whether assignments were submitted, course average, participation in teacher and unit feedback survey, and number of completed and withdrawn units.
6. Cohen [54] applied data examination techniques for calculating student activity variables, and analysing student actions utilising website log files. Key indicators: course website usage, number of student actions, final grade and students' learning status.
7. Darlington [55] computed logistic regression, stepwise logistic regression utilising datasets from LMS (number of submissions, how often the student has viewed feedback etc.). Key indicators: motivation (interest, relevance, knowledge), time allocation, attendance, reading, study/preference and prior access.
8. Dorodchi, et al. [56] used sequence model and feature vector model techniques. Key indicators: student background information (age group, gender, major), performance scores (grades for quizzes), reflections and self-assessments.
9. Guarrin [57] developed new variables standardization, normalization, categorisation techniques utilising information of bi-annual admission process, options for enrolment, socio-demographic attributes, data regarding students' academic life and loss of academic status. Key indicators: demographic data (student's age at enrolment), socioeconomic status, admission scores, type of school attended, option for enrolment and grades.
10. He, et al. [58] applied supervised binary classifiers, regularised logistic regression, SVM, random forest, decision tree, naïve Bayes and Bayes Net techniques on student engagement datasets with video lectures and assignments data, and their



performance on assignments. Key indicators: percentage of lectures viewed, assignments completed, and score on completed assignment.

11. Ikbal, et al. [59] applied a cross-validation technique utilising datasets containing student information (enrolment, demographics, course history, test scores), and data concerning teachers, learning and assessment activities. Key indicators: students' demographics, enrolment history, test performance, learning disability (if any), student course history, discipline history (if any), absenteeism (if any), test details.
12. Lacave, et al. [60] used probabilistic models and Bayesian networks techniques on databases containing records regarding students' information. Key indicators: academic and social data such as gender, enrolment age, prior subject choice, first enrolment year, grade, scholarship, highest course enrolled.
13. Manrique, et al. [61] applied global feature-based, local feature-based, time series techniques utilising student information datasets from two courses. Key indicators: grade averages, number of credits enrolled, and number of lost courses.
14. Sales, et al. [62] computed F-measures, gain ratio algorithm, specific model selection techniques with students' academic records. Key indicators: course information, assessment information, student situation, and subject type.
15. Saqr, et al. [63] used binary logistic model technique with online activity data in LMS. Key indicators: logins, views/hits, forums visited, time spent online, and formative assessment.
16. Seidel and Kutieleh [64] applied Chi-squared automatic iterative detection model techniques on student demographics, university application, enrolment characteristics, academic performance, logins to LMS and registrations. Key indicators: number of recent class registrations, forum access and study load.
17. Srilekshmi, et al. [65] used data mining techniques utilising datasets on first year MITx and HarvardX courses on the edX platform. Key indicators: personal information, academic records, and learning background.
18. Thammasiri, et al. [66] applied data mining techniques utilising institutional student datasets. Key indicators: academic, financial and demographic information.
19. Whitehill, et al. [67] deployed machine learning techniques utilising 40 MOOCs datasets from Harvard X. Key indicators: clickstream features and social network measures.
20. Wolff, et al. [68] used General Unary Hypothesis Automation technique utilising user's activity in VLE. Key indicator: number of clicks.

#### 4.4 Indicators for predicting general performance of students

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.

1. Carter, et al. [69] used statistical analysis and machine learning techniques on programming log data and course grades. Key indicators: students' grades students' overall assignment average, students' final grades.
2. Conijn, et al. [70] 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).
3. Conijn, et al. [71] used descriptive analyses, Pearson correlational analyses and multiple linear regression, backward stepwise regression techniques utilising datasets on a MOOC-course provided on Coursera. Key indicators: platform logfiles (trace data).
  4. Daud, et al. [72] deployed Support Vector Machines, C4.5, classification and regression tree, Bayes network, Naïve Bayes techniques. Key indicators: family expenditure, family income, student personal information and family assets.
  5. Elbadrawy, et al. [73] used a multi-regression model utilising student data, course data and learning activities data. Key indicators: performance-specific features, activity/course-specific features and Moodle interaction features.
  6. Gkontzis, et al. [74] applied regression analysis techniques (random forest, linear regression, neural network, adaboost, sim and kin) utilising each student's interactions extracted from Moodle. Key indicators: gender, logins in module, logins in forum, forum replies, dedication time, main quizzes, MCQ per week, self-assessment quizzes.
  7. Jo, et al. [75] computed a multiple linear regression analysis utilising web-log data. Key indicators: total login frequency, studying time, irregularity of learning interval, interactions with content, peers and instructor(s), and total number of completed assignments/assessments.
  8. Kim, et al. [76] calculated dashboard usage frequency, summation of students' scores, t-test, multiple regression analysis techniques utilising log data extracted from LMS, survey data and final assessment scores. Key indicators: dashboard usage frequency, dashboard satisfaction, and learning achievement.
  9. Mitra and Goldstein [77] used cross-validation techniques utilising online survey datasets. Key indicators: demographic factors, academic history and records, work-related factors, course-related factors and academic self-concept factors.
  10. Nespereira, et al. [78] applied risk-detection algorithm, time-series, temporal decomposition techniques utilising datasets from the Moodle platform (such as course contents, students' personal data, grades, and students' interactions with the platform). Key indicators: number of completed assignments/courses, blog activity (if any), and participation in the viewing of resources, forums and quizzes.
  11. Okubo, et al. [79] deployed recurrent neural network utilising log data in educational systems. Key indicators: attendance, quizzes completed, report data, course & slide views, utilization of markers, memos, actions, and word count in forums.
  12. Rogers, et al. [80] computed regression analysis utilising variables from online systems (demographic and performance-based). Key indicators: grade, gender, age, current academic load, completed courses, GPA, whether they had withdrawn any courses previously, whether they are enrolled next year for any courses, counselling activities (if any), and any previously notice given for poor progresses.
  13. Sarker [81] used categorical principal component analysis and logistic regression techniques on data from a questionnaire consisting of 49 items from the institutional internal databases items. Key indicators: academic background, environmental variables and psychological test outcomes.

#### **4.5 Indicators for predicting course completion**

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.

1. Andersson, et al. [82] used binary logistic regression for examining the trail left by students' activities on a discussion forum in online courses during three different points in time (50%, 75%, 100% of course completion) while studying a course. Key indicators: number, length and frequency of posts.
2. Aulck, et al. [83] reported machine learning experiments utilising 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).
3. Dawson, et al. [84] applied common statistical methods utilising student information systems, LMS interactions and assessment data. Key indicators: LMS engagement, attendance in class, and academic grades/outcomes.
4. Djulovic and Li [85] used Chi-squared tests, information gain, gain ratio and correlation analysis techniques with enrolment data (such as age, gender, GPA, sat reading score, math score, writing score and student term-specific financial balance). Key indicators: Pre-enrolment variables, semester-specific variables and financial aid status.
5. Guerrero-Higuera, et al. [86] deployed the Model Evaluator to investigate different machine learning models. Cross-validation analysis was also deployed in this study utilising learners' interaction with the version control system – GIT repository. Key indicators: students' activity, students' interactions - such as comments, days spent, comments per day, additions, deletions, number of issues, and authorship proof - in online forums.
6. Zimmerman and Johnson [87] computed stepwise logistic regression and confirmatory factor analysis utilising datasets. Key indicators: expected grade, expected time commitment and first lesson quiz.

## **5 Discussion**

Technology-based innovations in education have significantly altered both the scale and resolution of measurements for complex learning processes [88-90]. Recent developments in this field have heightened the need for educational data mining, machine learning, and statistics to gain insights from the fine-grained process data generated in technology-rich learning environments [89-91]. Several perspectives on educational data and analytics have been identified: (1) The data-driven perspective utilises existing data, mostly stemming from database systems, for informing different stakeholders. While big datasets may be available, the purpose for collecting data may have been different in the first place, hence, being biased when utilised for other purposes. In contrast, the (2) data-demand perspective follows a specific analytics purpose and defines the data to be collected. This enables a well-directed analysis of educational data with

direct implications for learning and teaching. A third perspective may be a combination of the above-mentioned approaches.

Consequently, emerging analytics solutions are related to different data available for analysis, also dependent on the level or type of educational institution (e.g., university vs. distance learning vs. MOOC) [92]. Predictive models need to be trained using tutors' experience or by machine learning algorithms. Then they are required to make predictions for current students based on information that they have on the current presentation of the course as well as past information. Tendencies and traditions, many universities would like to be informed whether their students would finish their studies in the designated time. Models are trained from data from previous cohort and applied to current cohorts [92]. GPA was found to be the feature with the highest predictive power [93]. At university or higher education settings, models are trained for identify success or failure of current cohort, often only from demographic data [18]. In addition, institutions may define at-risk student differently and therefore currently no standard method of detecting at-risk students exist [53, 92].

In summary, student study history such as GPA or evaluating the learners' learning progress from assessment results are the most successful key indicators. However, for entry-level courses or first-year students, historic academic performance or study history may be unavailable [10, 11]. However, quiz results in the first week can be used to obtain this information to make the appropriate analysis. In MOOCs, clickstream data seem to be the key indicators, as performance and demographic data may not be available due to limited data collection and privacy issues [92].

Regarding the distribution of indicators mapped on the three data profiles (student, learning, curriculum) [18] it becomes apparent that indicator related to the curriculum profile are underrepresented. As such curriculum related indicators may function as benchmarks for formative feedback and (near) real-time scaffolds as well as for improving the learning design, it is suggested to further investigate the benefits and usability as well as the validity of such indicators.

## **5.1 Recommendation of methods**

The following five recommendations for specific indicators focus on (1) task related predictions, (2) social, learning or engagement behaviour, (3) low-performing or drop-out 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, visualization and multi-level modelling mostly utilizing 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.

## 5.2 Conclusions and future works

A number of research/implementation directions were made clear and concluded from our study. These include (1) the standardisation of learning analytics systems ready for institutions to adopt without the need for each one to implement their own; (2) 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; (3) elaborating from (2), 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); (4) more work on privacy and ethical guidelines; (5) 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.

## 6 Acknowledgement

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

## 7 References

- [1] N. Sclater, A. Peasgood, and J. Mullan, Learning analytics in higher education: A review of UK and international practice. Bristol: JISC, 2016.

- [2] D. Ifenthaler, D.-K. Mah, and J. Y.-K. Yau, "Utilising learning analytics for study success. Reflections on current empirical findings," in Utilizing learning analytics to support study success, D. Ifenthaler, J. Y.-K. Yau, and D.-K. Mah Eds. Cham: Springer, 2019, pp. 27–36. [https://doi.org/10.1007/978-3-319-64792-0\\_2](https://doi.org/10.1007/978-3-319-64792-0_2)
- [3] D.-K. Mah, J. Y.-K. Yau, and D. Ifenthaler, "Future directions on learning analytics to enhance study success," in Utilizing learning analytics to support study success, D. Ifenthaler, J. Y.-K. Yau, and D.-K. Mah Eds. Cham: Springer, 2019, pp. 313–321. [https://doi.org/10.1007/978-3-319-64792-0\\_17](https://doi.org/10.1007/978-3-319-64792-0_17)
- [4] J. Wong et al., "Educational theories and learning analytics: from data to knowledge," in Utilizing learning analytics to support study success, D. Ifenthaler, J. Y.-K. Yau, and D.-K. Mah Eds. Cham: Springer, 2019, pp. 3–25. <https://doi.org/10.1007/978-3-319-64792-0>
- [5] L. P. Prieto, M. J. Rodríguez-Triana, R. Martínez-Maldonado, Y. Dimitriadis, and D. Gašević, "Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level," *Australasian Journal of Educational Technology*, vol. 35, no. 4, pp. 14–33, 2019. <https://doi.org/10.14742/ajet.4314>
- [6] D. Azcona, I. Hsiao, and A. F. Smeaton, "Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints," *User Modeling and User-Adapted Interaction*, vol. 29, pp. 759–788, 2019, doi: 10.1007/s11257-019-09234-7. <https://doi.org/10.1007/s11257-019-09234-7>
- [7] C. Lawson, C. Beer, D. Rossi, T. Moore, and J. Fleming, "Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution," *Educational Technology Research and Development*, vol. 64, no. 5, pp. 957–968, 2016. <https://doi.org/10.1007/s11423-016-9459-0>
- [8] D. Ifenthaler, "Learning analytics," in *The SAGE encyclopedia of educational technology*, vol. 2, J. M. Spector Ed. Thousand Oaks, CA: Sage, 2015, pp. 447–451.
- [9] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, "Intelligent tutoring systems and learning outcomes: a meta-analysis," *Journal of Educational Psychology*, vol. 106, no. 4, pp. 901–918, 2014. <https://doi.org/10.1037/a0037123>
- [10] D.-K. Mah and D. Ifenthaler, "Academic staff perspectives on first-year students' academic competencies," *Journal of Applied Research in Higher Education*, vol. 9, no. 4, pp. 630–640, 2017. <https://doi.org/10.1108/jarhe-03-2017-0023>.
- [11] D.-K. Mah and D. Ifenthaler, "What do first-year students need? Digital badges for academic support to enhance student retention," *Journal of Applied Research in Higher Education*, vol. 12, no. 1, pp. 86–96, 2020. <https://doi.org/10.1108/jarhe-12-2018-0258>.
- [12] C. S. Sarrico, "Completion and retention in higher education," in *Encyclopedia of international higher education systems and institutions*, T. P. and S. J. Eds. Dordrecht: Springer, 2018. [https://doi.org/10.1007/978-94-017-9553-1\\_87-1](https://doi.org/10.1007/978-94-017-9553-1_87-1)
- [13] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics for study success: a systematic review," *Education Tech Research Dev*, in press. <https://doi.org/10.1007/s11423-020-09788-z>
- [14] S. K. Banihashem, K. Aliabadi, S. Pourroostaei Ardakani, A. Delaver, and M. Nili Ahmadabadi, "Learning analytics: a systematic literature review," *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, vol. 9, no. 2, pp. 1–10, 2018. <https://doi.org/10.5812/ijvllms.63024>
- [15] M. S. Larsen, K. P. Kornbeck, M. B. Larsen, R. M. Kristensen, and H. Sommersel, *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, 2013.

- [16] C. Vieira, P. Parsons, and V. Byrd, "Visual learning analytics of educational data: A systematic literature review and research agenda," *Computers & Education*, vol. 122, pp. 119–135, 2018, <https://doi.org/10.1016/j.compedu.2018.03.018>.
- [17] Z. Papamitsiou and A. Economides, "Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence," *Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, 2014.
- [18] D. Ifenthaler and C. Widanapathirana, "Development and validation of a learning analytics framework: Two case studies using support vector machines," *Technology, Knowledge and Learning*, vol. 19, no. 1–2, pp. 221–240, 2014, <https://doi.org/10.1007/s10758-014-9226-4>.
- [19] V. Tinto, "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*, vol. 2, no. 3, pp. 89–97, 2005.
- [20] R. D. Padgett, J. R. Keup, and E. T. Pascarella, "The impact of first-year seminars on college students' life-long learning orientations," *Journal of Student Affairs Research and Practice*, vol. 50, no. 2, pp. 133–151, 2013, <https://doi.org/10.1515/jsarp-2013-0011>.
- [21] P. A. Attewell, D. E. Lavin, T. Domina, and T. Levey, "New evidence on college remediation," *The Journal of Higher Education*, vol. 77, no. 5, pp. 886–924, 2006, <https://doi.org/10.1353/jhe.2006.0037>.
- [22] L. Morosanu, K. Handley, and B. O'Donovan, "Seeking support: Researching first-year students' experiences of coping with academic life," *Higher Education Research & Development*, vol. 29, no. 6, pp. 665–678, 2010, <https://doi.org/10.1080/07294360.2010.487200>.
- [23] V. Schmied and M. Hänze, "The effectiveness of study skills courses: Do they increase general study competences?," *Zeitschrift für Hochschulentwicklung*, vol. 10, no. 4, pp. 176–187, 2015, <https://doi.org/10.3217/zfhe-10-04/09>.
- [24] P. Bijmans and A. H. Schakel, "The impact of attendance on first-year study success in problem-based learning," *Higher Education*, vol. 76, pp. 865–881, 2018, <https://doi.org/10.1007/s10734-018-0243-4>.
- [25] OECD, *Education at a glance 2019: OECD indicators*. Paris: OECD Publishing, 2019.
- [26] V. Tinto, "Limits of theory and practice in student attrition," *The Journal of Higher Education*, vol. 53, no. 6, pp. 687–700, 1982, <https://doi.org/10.2307/1981525>.
- [27] U. Heublein, "Student drop-out from German higher education institutions," *European Journal of Education*, vol. 49, no. 4, pp. 497–513, 2014, <https://doi.org/10.1111/ejed.12097>.
- [28] L. Thomas, "Student retention in higher education: The role of institutional habitus," *Journal of Education Policy*, vol. 17, no. 4, pp. 423–442, 2002, <https://doi.org/10.1080/02680930210140257>.
- [29] L. Willcoxson, J. Cotter, and S. Joy, "Beyond the first-year experience: The impact on attrition of student experiences throughout undergraduate degree studies in six diverse universities," *Studies in Higher Education*, vol. 36, no. 3, pp. 331–352, 2011, <https://doi.org/10.1080/03075070903581533>.
- [30] M. Yorke and B. Longden, *The first-year experience of higher education in the UK*. York: The Higher Education Academy, 2008.
- [31] V. Tinto, "Through the eyes of students," *Journal of College Student Retention: Research, Theory and Practice*, vol. 19, no. 3, pp. 254–269, 2017, <https://doi.org/10.1177/1521025115621917>.
- [32] L. Remedios, D. Clarke, and L. Hawthorne, "The silent participant in small group collaborative learning contexts," *Active Learning in Higher Education*, vol. 9, no. 3, pp. 201–216, 2008, <https://doi.org/10.1177/1469787408095846>.

- [33] T. Brahm, T. Jenert, and D. Wagner, "The crucial first year: a longitudinal study of students' motivational development at a Swiss Business School," *Higher Education*, vol. 73, no. 3, pp. 459–478, 2017. <https://doi.org/10.1007/s10734-016-0095-8>
- [34] M. D. Pistilli and K. E. Arnold, "Purdue Signals: Mining real-time academic data to enhance student success," *About campus: Enriching the student learning experience*, vol. 15, no. 3, pp. 22–24, 2010. <https://doi.org/10.1002/abc.20025>
- [35] B. T. M. Wong, "Learning analytics in higher education: an analysis of case studies," *Asian Association of Open Universities Journal*, vol. 12, no. 1, pp. 21–40, 2017, <https://doi.org/10.1108/aaouj-01-2017-0009>
- [36] M. Berland, R. S. Baker, and P. Bilkstein, "Educational data mining and learning analytics: Applications to constructionist research," *Technology, Knowledge and Learning*, vol. 19, no. 1–2, pp. 205–220, 2014, <https://doi.org/10.1007/s10758-014-9223-7>.
- [37] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *Sprouts: Working Papers on Information Systems*, vol. 10, no. 26, pp. 1–49, 2010. <https://doi.org/10.2139/ssm.1954824>
- [38] C. Okoli, "A guide to conducting a standalone systematic literature review," *Communications of the Association for Information Systems Research*, vol. 37, no. 43, pp. 879–910, 2015, doi: 10.17705/1CAIS.03743.
- [39] K. Thompson, "Using micro-patterns of speech to predict the correctness of answers to mathematics problems: an exercise in multimodal learning analytics," presented at the ICMI '13: Proceedings of the 15th ACM on International conference on multimodal interaction, Sydney, NSW, Australia, 2013. <https://doi.org/10.1145/2522848.2533792>
- [40] T. Yang, C. Brinton, C. Wong, and M. Chiang, "Behavior-based grade prediction for MOOCs via time series neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 716–728, 2017. <https://doi.org/10.1109/jstsp.2017.2700227>
- [41] H. Bydzovska and L. Popelinsky, "The influence of social data on student success prediction," presented at the International Database Engineering & Applications Symposium, 2014. <https://doi.org/10.1145/2628194.2628199>
- [42] P. Carroll and A. White, "Identifying patterns of learner behaviour: What Business Statistics students do with learning resources," *INFORMS Transactions on Education*, vol. 18, no. 1, pp. 1–13, 2017. <https://doi.org/10.1287/ited.2016.0169>
- [43] L. Gong, Y. Liu, and W. Zhao, "Using learning analytics to promote student engagement and achievement in blended learning: an empirical study," presented at the International Conference on E-Education, E-Business and E-Technology, 2018. <https://doi.org/10.1145/3241748.3241760>
- [44] Y. Hu, C. Lo, and S. Shih, "Developing early warning systems to predict students' online learning performance," *Computers in Human Behavior*, vol. 36, pp. 469–478, 2014. <https://doi.org/10.1016/j.chb.2014.04.002>
- [45] H. Labarthe, F. Bouchet, R. Bachelet, and K. Yacef, "Does a peer recommender foster students' engagement in MOOCs?," presented at the International Conference on Educational Data Mining, 2016.
- [46] O. Lu, J. Huang, A. Huang, and S. Yang, "Applying learning analytics for improving students' engagement and learning outcomes in an MOOCs enabled collaborative programming course," *Interactive Learning Environments*, vol. 25, no. 2, pp. 220–234, 2017. <https://doi.org/10.1080/10494820.2016.1278391>
- [47] S. Nam, G. Frishkoff, and K. Collins-Thompson, "Predicting students' disengaged behaviors in an online meaning-generation task," *IEEE Transactions on Learning Technology*, vol. 11, no. 3, pp. 362–375, 2017. <https://doi.org/10.1109/tlt.2017.2720738>



- [48] Q. Nguyen, B. Rienties, L. Toetnel, R. Ferguson, and D. Whitelock, "Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates," *Computers in Human Behavior*, vol. 76, pp. 703–714., 2017. <https://doi.org/10.1016/j.chb.2017.03.028>
- [49] E. Aguiar, N. Chawla, J. Brockman, G. Ambrose, and V. Goodrich, "Engagement vs performance: using electronic portfolios to predict first semester engineering student retention," in *Proceedings of the fourth international learning analytics & knowledge conference*, I. Molenaar, X. Ochoa, and S. Dawson Eds. New York, NY: ACM, 2014, pp. 103–112. <https://doi.org/10.18608/ila.2014.13.3>
- [50] R. Bukralia, A. Deokar, and S. Sarnikar, "Using academic analytics to predict dropout risk in e-Learning courses," in *Reshaping society through analytics, collaboration and decision support*, L. S. Iyer and D. J. Power Eds. Cham: Springer, 2014, pp. 67–93. [https://doi.org/10.1007/978-3-319-11575-7\\_6](https://doi.org/10.1007/978-3-319-11575-7_6)
- [51] W. Cambrozzi, S. Rigo, and J. Barbosa, "Dropout prediction and reduction in distance education courses with the learning analytics. Multi-trail approach," *Journal of Computer Science*, vol. 21, no. 1, pp. 23–47, 2015.
- [52] K. Casey and D. Azcona, "Utilising student activity patterns to predict performance," *Journal of Educational Technology in Higher Education*, vol. 14, no. 4, pp. 1–15, 2017, doi: 10.1186/s41239-017-0044-3.
- [53] K. E. K. Chai and D. C. Gibson, "Predicting the risk of attrition for undergraduate students with time based modelling," in *Proceedings of cognition and exploratory learning in the digital age*, D. G. Sampson, J. M. Spector, D. Ifenthaler, and P. Isaias Eds. Maynooth, Ireland: IADIS Press, 2015, pp. 109–116.
- [54] A. Cohen, "Analysis of student activity in web-supported courses as a tool for predicting dropout," *Educational Technology Research and Development*, vol. 65, no. 5, pp. 1–20, 2017. <https://doi.org/10.1007/s11423-017-9524-3>
- [55] W. Darlington, "Predicting underperformance from students in upper level engineering courses," Rochester Institute of Technology, 2017.
- [56] M. Dorodchi, A. Benedict, D. Desai, M. Mahzoon, S. Macneil, and N. Dehbozorgi, "Design and implementation of an activity-based introductory Computer Science Course (CS1) with periodic reflections validated by learning analytics," presented at the *IEEE Frontiers In Education*, 2018. <https://doi.org/10.1109/fie.2018.8659196>
- [57] C. Guarrin, "Data mining model to predict academic performance at the Universidad Nacional de Colombia," University of Colombia, 2013.
- [58] J. He, J. Bailey, B. Rubinstein, and R. Zhang, "Identifying at-risk students in Massive Open Online Courses," presented at the *International Conference on Artificial Intelligence*, 2015.
- [59] S. Ikbal, A. Tamhane, A. Sengupta, M. Chetlur, S. Ghosh, and J. Appleton, "On early prediction of risks in academic performance for students," *IBM Journal of Research and Development*, vol. 59, no. 6, pp. 1–14, 2015. <https://doi.org/10.1147/jrd.2015.2458631>
- [60] C. Lacave, A. Molina, and J. Cruz-Lemus, "Learning analytics to identify dropout factors of Computer Science studies through Bayesian networks," *Behaviour & Information Technology*, vol. 37, no. 10–11, pp. 993–1007, 2018, <https://doi.org/10.1080/0144929x.2018.1485053>.
- [61] R. Manrique, B. Nunes, O. Marino, M. Casanova, and T. Nurmikko-Fuller, "An analysis of student representation, representative features and classification algorithms to predict degree dropout," presented at the *International Conference on Learning Analytics and Knowledge*, 2018. <https://doi.org/10.1145/3303772.3303800>

- [62] A. Sales, L. Balby, and A. Cajueiro, "Exploiting academic records for predicting student drop out: a case study in Brazilian higher education," *Journal of Information and Data Management*, vol. 7, no. 2, pp. 166–181, 2016.
- [63] M. Saqr, U. Fors, and M. Tedre, "How learning analytics can early predict under-achieving students in a blended medical education course," *Medical Teacher*, vol. 39, no. 7, pp. 757–767, 2017. <https://doi.org/10.1080/0142159x.2017.1309376>
- [64] E. Seidel and S. Kutieleh, "Using predictive analytics to target and improve first year student attrition," *Australasian Journal of Educational Technology*, vol. 61, no. 2, pp. 200–218, 2017. <https://doi.org/10.1177/0004944117712310>
- [65] M. Srilekshmi, S. Sindhumol, S. Chatterjee, and K. Bijani, "Learning analytics to identify students at-risk in MOOCs " presented at the International Conference on Technology for Education, 2016. <https://doi.org/10.1109/t4e.2016.048>
- [66] D. Thammasiri, D. Delen, P. Meesad, and N. Kasap, "A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition," *Expert Systems with Applications*, vol. 41, no. 2, pp. 321–330, 2014. <https://doi.org/10.1016/j.eswa.2013.07.046>
- [67] J. Whitehill, K. Mohan, D. Seaton, Y. Rosen, and D. Tingley, "MOOC dropout prediction: how to measure accuracy?," presented at the ACM Conference on Learning @ Scale, 2017. <https://doi.org/10.1145/3051457.3053974>
- [68] A. Wolff, Z. Zdrahal, N. Andriy, and M. Pantucek, "Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment," in *Proceedings of the third international learning analytics & knowledge conference*, I. Molenaar, X. Ochoa, and S. Dawson Eds. New York, NY: ACM, 2013, pp. 145–149. <https://doi.org/10.1145/2460296.2460324>
- [69] A. Carter, C. Hundhausen, and O. Adesope, "The normalized programming state model: predicting student performance in computing courses based on programming behavior," presented at the International Conference on Computing Education Research, 2015. <https://doi.org/10.1145/2787622.2787710>
- [70] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: a comparison of 17 blended courses using Moodle LMS," *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 17–29, 2017. <https://doi.org/10.1109/tlt.2016.2616312>
- [71] R. Conijn, A. Van den Beemt, and P. Cuijpers, "Predicting student performance in a blended MOOC," *Journal of Computer Assisted Learning*, vol. 34, pp. 615–628, 2018, <https://doi.org/10.1111/jcal.12270>.
- [72] A. Daud, N. Aljohani, R. Abbasi, M. Lytras, F. Abbas, and J. Alowibdi, "Predicting student performance using advanced learning analytics," presented at the Conference on World Wide Web Companion, 2017. <https://doi.org/10.1145/3041021.3054164>
- [73] A. Elbadrawy, R. Studham, and G. Karypis, "Collaborative multi-regression models for predicting students' performance in course activities," in *Proceedings of the fifth international learning analytics & knowledge conference*, I. Molenaar, X. Ochoa, and S. Dawson Eds. New York, NY: ACM, 2015, pp. 103–107. <https://doi.org/10.1145/2723576.2723590>
- [74] A. Gkontzis, S. Kotsiantis, R. Tsoni, and V. Verykios, "An effective LA approach to predict student achievement," presented at the Pan-Hellenic Conference, 2018. <https://doi.org/10.1145/3291533.3291551>
- [75] I. Jo, T. Yu, H. Lee, and Y. Kim, "Relations between student online learning behavior and academic achievement in higher education: A learning analytics approach," in *Emerging issues in smart learning*, G. Chen, V. Kumar, Kinshuk, R. Huang, and S. Kong Eds. Berlin: Springer, 2014, pp. 275–287. [https://doi.org/10.1007/978-3-662-44188-6\\_38](https://doi.org/10.1007/978-3-662-44188-6_38)

- [76] J. Kim, I.-H. Jo, and Y. Park, "Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement," *Asia Pacific Education Review*, vol. 17, no. 1, pp. 13–24, 2016. <https://doi.org/10.1007/s12564-015-9403-8>
- [77] S. Mitra and Z. Goldstein, "Designing early detection and intervention techniques via predictive statistical models—A case study on improving student performance in a business statistics course," *Communications in Statistics: Case Studies, Data Analysis and Applications*, vol. 1, 1, pp. 9–21, 2015. <https://doi.org/10.1080/23737484.2015.1063409>
- [78] C. Nespereira, A. Vilas, and R. Redondo, "Am I failing this course?: risk prediction using e-learning data," presented at the Conference on Technological Ecosystems for enhancing Multiculturalism, 2015. <https://doi.org/10.1145/2808580.2808621>
- [79] F. Okubo, T. Yamashita, A. Shimada, and H. Ogata, "A neural network approach for students' performance prediction," presented at the International Conference on Learning Analytics and Knowledge, 2017. <https://doi.org/10.1145/3027385.3029479>
- [80] T. Rogers, C. Colvin, and B. Chiera, "Modest analytics: using the index method to identify students at risk of failure," in *Proceedings of the fourth international conference on learning analytics and knowledge*, A. Pardo and S. D. Teasley Eds. New York, NY: ACM, 2014, pp. 118–122. <https://doi.org/10.1145/2567574.2567629>
- [81] F. Sarker, "Linked data technologies to support higher education challenges: student retention, progression and completion," Dissertation, University of Southampton, 2014.
- [82] U. Andersson, T. Arvemo, and M. Gellerstedt, "How well can completion of online courses be predicted using binary logistic regression?," in *Information Systems Research Seminar in Scandinavia*, Ljungskile, 07-08-2016, 2016, pp. 1–12.
- [83] L. Aulck, R. Aras, L. Li, C. L'Heureux, P. Lu, and J. West, "Stemming the tide: Predicting STEM attrition using student transcript data," presented at the Machine Learning for Education, Halifax, Nova Scotia, Canada, 2017.
- [84] S. Dawson, J. Jovanović, D. Gašević, and A. Pardo, "From prediction to impact: Evaluation of a learning analytics retention program," in *Proceedings of the seventh international learning analytics & knowledge conference*, I. Molenaar, X. Ochoa, and S. Dawson Eds. New York, NY: ACM, 2017, pp. 474–478. <https://doi.org/10.1145/3027385.3027405>
- [85] A. Djulovic and D. Li, "Towards freshman retention prediction: a comparative study," *International Journal of Information and Educational Technology*, vol. 3, no. 5, pp. 494–500, 2013.
- [86] A. M. Guerrero-Higuera, N. DeCastro-Garci, V. Matellan, and M. Conde-Gonzalez, "Predictive models of academic success: a case study with version control systems," presented at the International Conference on Technological Ecosystems for Enhancing Multiculturalism, 2018. <https://doi.org/10.1145/3284179.3284235>
- [87] W. Zimmerman and G. Johnson, "Exploring factors related to completion of an online undergraduate-level introductory statistics course," *Online Learning*, vol. 21, no. 3, pp. 19–205, 2017. <https://doi.org/10.24059/olj.v21i3.1017>
- [88] D. C. Gibson and D. Ifenthaler, "Preparing the next generation of education researchers for big data in higher education," in *Big data and learning analytics: Current theory and practice in higher education*, B. Kei Daniel Ed. New York, NY: Springer, 2017, pp. 29–42. [https://doi.org/10.1007/978-3-319-06520-5\\_4](https://doi.org/10.1007/978-3-319-06520-5_4)
- [89] D. Ifenthaler, "Designing effective digital learning environments: toward learning analytics design," *Technology, Knowledge and Learning*, vol. 22, no. 3, pp. 401–404, 2017, <https://doi.org/10.1007/s10758-017-9333-0>.

- [90] D. Ifenthaler, "Supporting higher education students through analytics systems," *Journal of Applied Research in Higher Education*, vol. 12, no. 1, pp. 1–3, 2020, <https://doi.org/10.1108/jarhe-07-2019-0173>.
- [91] D. Ifenthaler, D. G. Sampson, and J. M. Spector, "Linking analytics data and digital systems for supporting cognition and exploratory learning in 21st Century," *Computers in Human Behavior*, vol. 78, pp. 348–350 2018, <https://doi.org/10.1016/j.chb.2017.10.016>.
- [92] M. Hlosta, Z. Zdrahal, and J. Zendulka, "Ouroboros: Early identification of at-risk students without models based on legacy data," presented at the Seventh International Conference on Learning Analytics & Knowledge, Vancouver, Canada, 2017. <https://doi.org/10.1145/3027385.3027449>
- [93] H. Lakkaraju et al., "A machine learning framework to identify students at risk of adverse academic outcomes " presented at the KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 2015. <https://doi.org/10.1145/2783258.2788620>

## 8 Authors

**Dirk Ifenthaler** is Professor and Chair of Learning, Design and Technology at the University of Mannheim, Germany and UNESCO Deputy Chair of Data Science in Higher Education Learning and Teaching at Curtin University, Australia. His research focuses on the intersection of cognitive psychology, educational technology, data analytics and organisational learning ([www.ifenthaler.info](http://www.ifenthaler.info)).

**Jane Yin-Kim Yau** is a Researcher in Learning Analytics & Mobile Learning at the Chair for Learning, Design and Technology at the University of Mannheim, Germany. [jyau@mail.uni-mannheim.de](mailto:jyau@mail.uni-mannheim.de)

Article submitted 2020-05-19. Resubmitted 2020-06-04. Final acceptance 2020-06-04. Final version published as submitted by the authors.