Received: 10 March 2022 | Accepted: 22 September 2022

DOI: 10.1111/bjet.13282

REVIEW ARTICLE

British Journal of Educational Technology BERA

Which log variables significantly predict academic achievement? A systematic review and meta-analysis

Qin Wang | **Amin Mousavi**

Department of Educational Psychology & Special Education, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Correspondence
Amin Mousavi, Department of Educational Psychology & Special Education, University of Saskatchewan, 28 Campus Drive, Saskatoon, SK S7N 0X1, Canada.
Email: amin.mousavi@usask.ca

Abstract

Technologies and teaching practices can provide a rich log data, which enables learning analytics (LA) to bring new insights into the learning process for ultimately enhancing student success. This type of data has been used to discover student online learning patterns, relationships between online learning behaviors and assessment performance. Previous studies have provided empirical evidence that not all log variables were significantly associated with student academic achievement and the relationships varied across courses. Therefore, this study employs a systematic review with meta-analysis method to provide a comprehensive review of the log variables that have an impact on student academic achievement. We searched six databases and reviewed 88 relevant empirical studies published from 2010 to 2021 for an in-depth analysis. The results show different types of log variables and the learning contexts investigated in the reviewed studies. We also included four moderating factors to do moderator analyses. A further significance test was performed to test the difference of effect size among different types of log variables. Limitations and future research expectations are provided subsequently.

KEYWORDS
learning analytics, log analysis, online learning, academic achievement, meta-analysis

Practitioner notes

What is already known about this topic

- Significant relationship between active engagement in online courses and academic achievement was identified in a number of previous studies.
- Researchers have reviewed the literature to examine different aspects of applying LA to gain insights for monitoring student learning in digital environments (eg, data sources, data analysis techniques).

What this paper adds

- Presents a new perspective of the log variables, which provides a reliable quantitative conclusion of log variables in predicting student academic achievement.
- Conducted subgroup analysis, examined four potential moderating variables and identified their moderating effect on several log variables such as regularity of study interval, number of online sessions, time-on-task, starting late and late submission.
- Compared the effect of generic and course-specific, basic and elaborated log variables, and found significant difference between the basic and elaborated.

Implications for practice and/or policy

- A depth of understanding of these log variables may enable researchers to build robust prediction models.
- It can guide the instructors to timely adjust teaching strategies according to their online learning behaviors.

INTRODUCTION

In recent years, learning analytics (LA) has emerged as a field aiming to provide solutions for questions related to teaching and learning with technology, such as the ways to explore online learning and get an accurate description of learning process (Larusson & White, 2014). For the purpose of understanding and optimizing digital learning and the environments in which it occurs, LA ideally attempts to collect and analyze data that exists in educational repositories such as LMS to assess the behavior of educational communities (Romero & Ventura, 2010). Researchers have synthesized the literature regarding the data analyzed in LA studies and found that log records of learners' interaction with and participation in LMSs was the main data source of LA. For example, Saqr et al. (2018) conducted a systematic review of six empirical studies on LA published before 2017 in the field of medical education. Results showed that most reviewed studies collected data from LMSs or online learning resources. Algayres and Triantafyllou (2020) conducted a scoping review of 49 articles on LA in flipped learning environments. They found that LMS data was the main data source. Log variables such as total login time, time spent on online activities, regularity and engagement were usually extracted from LMS log traces. The analysis of log data, also known as data logging, is a process of making sense of computer-generated records (logs). Log analysis has had extensive adoption in the field of LA for some time, and empirical implications have witnessed its potential for providing valuable feedback for improving the effectiveness of online education. More specifically, it helps instructors understand students' online learning behaviors (Breslow et al., 2013; Cooper & Sahami, 2013; Daradoumis et al., 2013), provide feasible feedback and to adjust instructional strategies (Dietz-Uhler & Hurn, 2013).

The application of artificial intelligence (AI) in assessment has enabled a more continuous view of individual's ongoing engagement with an online learning environment, rather than discrete snapshots of performance provided by traditional assessments (Swiecki et al., 2022). AI techniques have been applied to different assessment tasks and evidence, such as electronic assessment platforms, stealth assessment, latent knowledge estimation and learning processes. By analyzing data generated from these approaches, previous research has investigated the following: test-taken behaviors (eg, time-on-task, answering and revising behavior during exams) (Lee et al., 2019), formative assessment using stealth methods (Yang et al., 2021), knowledge tracing (Molenaar et al., 2021) and analyzing multi-channel data (eg, clickstreams, eye-tracking, mouse movements) in multimodal LA with different AI techniques such as process mining and network analysis (de Marcos et al., 2016; Saqr et al., 2020). A wealth of such research has made student academic performance analysis and prediction become two widely explored research topics in LA.

Log variables investigated in the previous studies can be divided into basic and elaborated types. Basic log variables are those extracted from raw log data and are not specific measurements of previously outlined concepts, such as simple frequency and time counts. This type of log variables (eg, the number of clicks, total time spent online) is the most typical measure used to predict student learning performance. For instance, total login time was found to be positively related to final course grade (eg, Conijn et al., 2017; Wei et al., 2015). However, researchers suggested to extract and aggregate meaningful and elaborated indicators from log data, rather than basic frequency measures of online events (Hadwin et al., 2007; Huang & Fang, 2013; You, 2016). Huang and Fang (2013) claimed that merely adding more basic variables does not improve the predictability of mathematical models. Therefore, researchers need to develop significant indicators that effectively capture online engagement. For example, the variable of regular study (ie, the degree to which a student consistently accesses the learning materials) was generated in some studies based on the notion that self-regulated learners show a typical characteristic of studying on a regular basis (eg, Conijn et al., 2017; Jo et al., 2015; You, 2016). Results in these studies showed learners who regularly logged into the LMS throughout the course showed better performance. Such elaborated time-based indicators can serve as leading factors of student access time and study patterns simultaneously (You, 2016). These indicators explain learners' sustained endeavors and awareness of their learning status better than either login time or login frequency. A similar log variable of distributed learning was examined in Theobald et al. (2018), which was a measure of the number of weeks in which each student had accessed the LMS irrespective of the actual amount of time students spent online. Higher values suggested a more distributed and continual engagement with the course content. It was found that distributed learning was associated with better exam grades.

Log variables can also be classified as generic variables (eg, total login time, total number of clicks) and course-specific variables generated from interactions with specific online activities required in course syllabus (eg, number of weeks of high engagement with summative exercises, weekly use of course videos for the pre-class activities). For better investigating the relationship between online participation and academic performance, some studies have used both kinds of variables into analysis (eg, Jovanović et al., 2019, 2021; Wei et al., 2015). For course-specific log variables, different learning designs used in the courses can potentially lead to different activities in LMSs thus resulting in different LMS usage. As a result, these course-specific predictors cannot be compared across courses. In the present review study, meta regression analysis was conducted only on generic log variables.

Many previous studies analyzed LMS data of one or only a few courses and learning tasks, which makes it difficult to compare study results and draw generalizable conclusions in the ways of using LMS data for predictive modeling (Conijn et al., 2017). Some studies performed prediction modelling on several courses and found that the effect of student LMS behaviors on students' learning performance differs across courses (eg, Conijn et al., 2017; Gašević et al., 2016). In Conijn et al. (2017)'s study, several log variables, such as the total number of views and clicks, had a positive relationship with students' grade in some courses while showed a negative relationship in others. These contradicting results may be explained by the fact that the courses differed in characteristics such as type, theme and learning design. For example, students taking fully online courses show more online interactions with LMSs compared with blended courses, which might have a great possibility of contributing to the effect of log variables on student academic performance. Furthermore, the dependent variable used in prediction models were not all the same. Some studies performed regression analysis to investigate the relationships between log variables and total course score (eg, Bravo-Agapito et al., 2021; Jovanović et al., 2021), while others used final exam score or post-test score (eg, Schumacher & Ifenthaler, 2021; Ulfa & Fatawi, 2021). Final exam score and post-test score are the one-time exams and tests which assess learners' knowledge acquired through the courses or learning modules. Total course score is the sum of all assessment parts of the course, which typically covers both final exam score (if there is) and the grade weights of the course design such as assignments, discussion forums, and quizzes. Therefore, the current study examines whether the predictive power of log variables on final exam score and total course score differs.

In the current empirical studies on LA, researchers investigated the predictive power of log variables in different learning contexts, mainly including learning type (fully online or blended), learning theme (eg, STEM, culture and arts), and the type of the dependent variable in prediction models (total course score or final exam score). Furthermore, the study characteristic variable of sample size that has been examined in many published meta-analysis articles was also considered in the current study. Therefore, a total of four potential moderators (ie, sample size, learning type, learning theme, the type of the dependent variable) would be examined in this meta-analysis. We aim to provide a review with meta-analysis of log variables that have been found to be significant predictors for student academic performance, compare the effect size of basic and elaborated, generic and course-specific log variables and investigate whether the effect size of generic log variables will change according to the four moderating factors.

RESEARCH SIGNIFICANCE AND OBJECTIVES

Improving academic performance is considered one of the crucial issues for education. In the existing literature, numerous studies have adopted LA approach to explore the relationships between log variables and student academic performance in the hope of providing guidance for instructors to make decisions. For example, instructors can encourage students in engaging online learning activities if they are less active during a longer period.

Recently, researchers have reviewed the literature to examine different aspects of applying LA to gain insights for monitoring student learning in digital environments, such as data sources, data analysis techniques, purposes and LA applications on some topics (eg, evaluation and assessment of student academic performance). Among them, several studies reviewed certain log variables used in predicting academic performance. For example, Namoun and Alshanqiti (2021) conducted a systematic literature review of 62 studies between 2010 and 2020 to investigate the applications of data mining and LA techniques in predicting student performance. Results showed that most studies employed regression and supervised ML models to predict student performance. Online engagement in learning activities, term assessment grades, and student academic emotions were the most evident predictors of learning performance. Ifenthaler and Yau (2020) reviewed 46 empirical studies published from 2013 to 2018 to investigate the effective role of LA in facilitating study success in the context of higher education. Results showed that one set of predictors for student success was variables extracted from online log traces which represented student online interactions and engagement, such as login frequency and submission of assignments.

Although some review studies found log data to be the main data sources of LA and summarized several log variables for predicting student academic performance, few studies further systematically provided a fine-grained summary of the influential log variables. Furthermore, despite the increase of LA research, there is no consensus to date on how LA might be implemented, eg, which data is useful, what different considerations have to be made regarding course characteristics, etc. (Agudo-Peregrina et al., 2014). It indicates that there is a need for further studies to investigate the issue of the generalizability of prediction models. Especially regarding the predictors, there is still a lack of comprehensive review of the log variables used in the context of predicting student academic achievement, especially, whether the effect of log variables on student academic achievement varies according to different learning contexts.

Thus, this study sets out to fill the gaps by providing a systematic review with a meta-analysis of log variables and their performance in predicting student academic achievement. In this regard, this paper investigates the studies from the body of research published in the most recent decade. The study aims to:

1. Perform a systematic review of influential log variables and moderating factors (ie, sample size, the course type and theme, the type of dependent variable used in prediction models).
2. Conduct meta-regression analyses on most frequently generic log variables and investigate whether the effect size of generic log variables will change according to the above mentioned four moderators.
3. Compare the effect size of basic and elaborated, generic and course-specific log variables.

METHOD

This study employs a systematic review to provide a comprehensive examination of the log variables that have a significant impact on student academic achievement. We extracted the standardized regression coefficients of the statistically significant log variables in the prediction models to do a meta-analysis. The meta-analysis of standardized regression coefficients has the potential to yield a more accurate estimate of the effect of a predictor variable on a dependent variable after controlling for other variables that might also be related to the outcome variable (Fernández-Castilla et al., 2018). As previous studies involved two kinds of dependent variables in regression analysis, that is, continuous (assessment scores) and categorical (eg, pass/fail, student dropout), we only focused on the regression analysis of assessment scores. We use academic achievement to indicate assessment scores in the following analysis.

This systematic review is conducted based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework proposed by Page et al., 2021. We systematically searched the empirical studies related to using log variables to predict student academic achievement. We used preselected keywords to search peer-reviewed scholarly studies through main electronic databases. The output of the research results created the primary collection of the studies which were then imported to Endnote, a reference management software. The metadata of the searched studies was extracted in the MS Excel worksheet. Then, the title and abstract of the studies were screened by two reviewers based on inclusion and exclusion criteria. Next, the full texts of the studies were evaluated based on the eligibility criteria.

Search strategy

We searched articles published from 2010 to 2021 on six databases that cover popular journals of interest in education and data science: Scopus, IEEE Xplore, ERIC, Web of science, ScienceDirect, and ACM Digital Library. The search was performed between November 2021 and December 2021.

The keywords are terms relating to LA (eg, log data, log analysis, educational data mining, learning analytics), and “assessing academic performance” (eg, assess*, student assess*, study success, learning performance, academic achievement). The keywords or the synonyms within each term were paired with Boolean operator OR and two terms of keywords were paired with AND. Because our main research objective is to examine the effect size of log variables in predicting student academic achievement, we used AND to add the keyword “regression OR coefficient” to Boolean expressions. An example search query was: (“learning analytics” OR “log data” OR “log analysis” OR “log file” OR “educational data mining” OR “log variable”) AND (assessment OR “educational assess*” OR “academic performance” OR “academic achievement” OR “study success” OR “academic success” OR “learning performance”) AND (regression OR coefficient).

Inclusion and exclusion criteria

The summary of the inclusion and exclusion criteria is given in Table 1. The search process was limited to complete full text articles published on journals and conference proceedings from 2010 to 2021 and written in English. Other types such as notes and book chapters were excluded during the search process. Only empirical studies were included.

Selection process

Article selection process was carried out by following the recommendations from the PRISMA framework as shown in Figure 1. The search work output a total of 3717 articles from the six digital databases. Three hundred and eighty-nine duplicates were excluded in the first round of article selection process. In the second round, we removed 2984 studies that were not relevant to the setting or context of this review. Finally, in the third round, we excluded studies that did not meet the inclusion criteria (n = 256). This round involved the inclusion of related studies and exclusion of non-related studies according to these eligibility criteria: (1) Does the studies involve log analysis and regression analysis? (2) Is assessment score

TABLE 1 Inclusion and exclusion criteria

Inclusion	Exclusion
Published from 2010 to 2021	Published outside the period of 2010 to 2021
Written in English	Not written in English
Primary studies	Not primary studies
Journal articles, conference papers	Books, Slides, Notes, Posters, Reports
Empirical studies	Not empirical studies
Complete full-text studies	Incomplete full-text studies
Directly focus on using online log variables in regression analysis for predicting student academic achievement and report regression coefficients	Studies that do not directly use online log variables in regression analysis for predicting student academic achievement and report regression coefficients

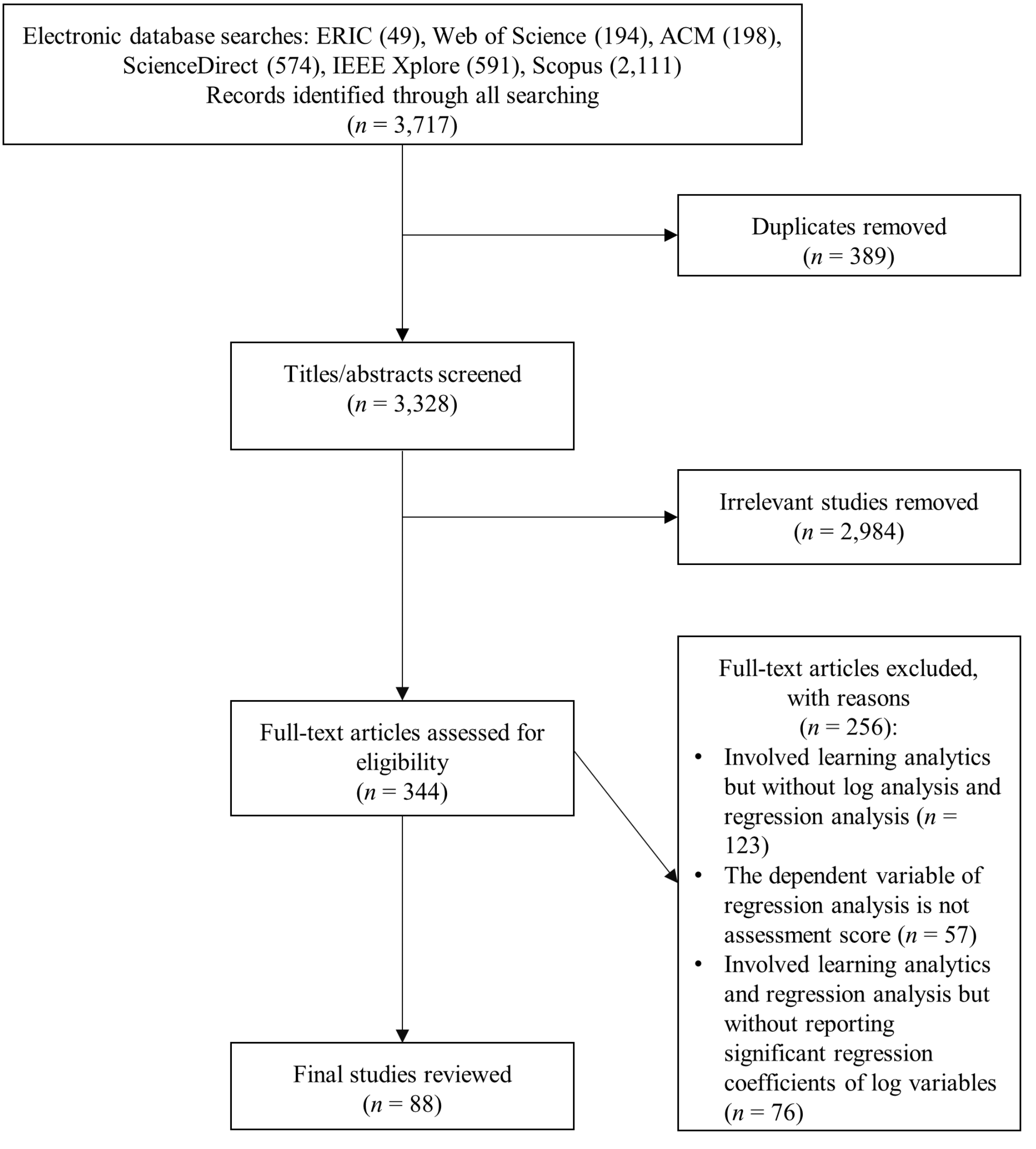

FIGURE 1 Flow diagram of the article search and selection process.

the dependent variable of the regression analysis? (3) Are the significant regression coefficients of log variables reported in the study? A total of 123 studies were eliminated from the selection as they did not perform log analysis or regression analysis. Then, 57 studies were excluded because their regression models were not used for predicting assessment scores. At last, 76 studies were excluded because no regression coefficients were reported. Therefore, 88 studies were kept from the final selection.

Data extraction and analysis

The final selection of the studies was examined by two reviewers for achieving the objectives of the study. The metadata of the selected studies was tabulated by using an Excel worksheet. The metadata includes author information, publication year, source title, conference rank and journal rank by Journal Impact Factor (JIF) and Journal Citation Indicator (if JIF is not available) (Clarivate, 2021) (Table 2).

Coding procedures

Outcome variable

We focus on the effect of log variables on student academic achievement. Therefore, the significant regression coefficients β of log variables and the sample size of each study were recorded. Regression coefficients were coded based on each independent regression model, and separately coded if a study performed several independent regression models. Besides, if a study included repeated regression analysis at different time, the results retrieved from the last regression analysis would be chosen. The coding was conducted by two researchers and checked interchangeably. Disputes that had occurred in the coding process were resolved.

Potential moderating variables

Four potential moderating variables were examined in the meta-analysis. These 33 studies were coded by two reviewers for these four variables.

1. Sample size. Sample size was coded as five ranges, "≤50", "51–100", "101–500", "501–1000", "≤1001".
2. Learning type. The learning type was coded as "fully online" and "blended".
3. Learning theme. The learning theme was coded as "Education", "Business", "STEM (Science, Technology, Engineering, Mathematics)", "Culture and Arts (CA)", "Life Sciences and Chemistry", "Medicine", "Society", "K-12", and "Multiple themes". The code of "Multiple themes" means that some studies built a regression model by using log data from multiple courses.
4. The type of the dependent variable. The type of the dependent variable was coded as "total score" and "exam score".

Statistical analyses

The statistical analyses include two parts, meta-regression analysis of generic log variables and independent sample *T* test for testing the significant difference between the effect sizes of generic and course-specific, basic and elaborated log variables.

As course-specific log variables were measured differently across courses, we only performed meta-regression analysis on generic log variables. We used the package *metafor* in R to conduct the meta-analysis. The *metafor* package (Viechtbauer, 2010) provides functions for conducting meta-analyses in R and includes the required methods for conducting moderator analyses without limitations compared with other packages. Users can fit

TABLE 2 Metadata of the reviewed studies

Year	Authors	Source title	Rank
2021	Bravo-Agapito et al. (2021)	Computers in Human Behavior	Q1
2021	Dewar et al. (2021)	Medical Teacher	Q1
2021	Galikyan et al. (2021)	Computers and Education	Q1
2021	Huang et al. (2021)	Proceedings of the 29th International Conference on Computers in Education	B - CORE
2021	Jost et al. (2021)	Education and Information Technologies	Q1
2021	Jovanović et al. (2021)	Computers and Education	Q1
2021	Maier (2021)	Computers and Education	Q1
2021	Mangaroska et al. (2021)	IEEE Transactions on Learning Technologies	Q1
2021	Mills (2021)	Heliyon	Q2
2021	Ober et al. (2021)	Computers and Education	Q1
2021	Schumacher and Ifenthaler (2021)	The Internet and Higher Education	Q1
2021	Smith et al. (2021)	Proceedings of the 52nd ACM Technical Symposium on Computer Science Education	A - CORE
2021	Tan et al. (2021)	International Conference on Human System Interaction, HSI, 2021	C - CORE
2021	Ulfa and Fatawi (2021)	International Journal of Emerging Technologies in Learning	Q2
2021	Wu et al. (2021)	IEEE 3rd International Conference on Computer Science and Educational Informatization	Not ranked
2021	Xu et al. (2021)	IEEE Transactions on Education	Q2
2021	Yang et al. (2021)	Assessment for Effective Intervention	Q2
2021	Chan et al. (2021)	European Journal of Dental Education	Q3
2020	Han and Ellis (2020)	Australasian Journal of Educational Technology	Q1
2020	Han et al. (2020)	Educational Technology Research and Development	Q1
2020	Li et al. (2020)	The Internet and Higher Education	Q1
2021	Papamitsiou and Economides (2021)	Journal of Computer Assisted Learning	Q1
2020	Pei et al. (2020)	International Journal of Science Education	Q2
2020	Saqr et al. (2020)	BMC Medical Education	Q2
2020	Sharma et al. (2020)	Studies in Higher Education	Q1
2020	Stadler et al. (2020)	Computers in Human Behavior	Q1
2020	Summers et al. (2020)	Assessment & Evaluation in Higher Education	Q1
2021	Tacoma, et al. (2021)	Journal of Computer Assisted Learning	Q1
2020	Tacoma et al. (2020)	Computers in Human Behavior	Q1
2020	Tempelaar et al. (2020)	PloS One	Q2
2020	Zarrabi and Bozorgian (2020)	Computers and Composition	Not available
2019	Chen et al. (2019)	2019 IEEE International Conference on Engineering, Technology and Education	Not ranked
2019	Foung and Chen (2019)	Electronic Journal of E-Learning	Q2
2019	Gu and Xu (2019)	Journal of Educational Computing Research	Q1

TABLE 2 (Continued)

Year	Authors	Source title	Rank
2019	Jokhan et al. (2019)	Studies in Higher Education	Q1
2019	Jovanović et al. (2019)	Computers and Education	Q1
2019	Koh et al. (2019)	Education Sciences	Q1
2019	Lee et al. (2019)	Science Education	Q1
2019	Musabirov et al. (2019)	International Journal of Emerging Technologies in Learning	Q2
2019	Ramirez-Arellano et al. (2019)	Journal of Educational Computing Research	Q1
2019	Soffer and Cohen (2019)	Journal of Computer Assisted Learning	Q1
2019	Tian et al. (2019)	Proceedings—International Joint Conference on Information, Media, and Engineering, IJCIME 2019	Not ranked
2018	Chiu and Hew (2018)	Australasian Journal of Educational Technology	Q1
2018	Conijn et al. (2018)	Journal of Computer Assisted Learning	Q1
2018	Jiang et al. (2018)	Contemporary Educational Psychology	Q1
2018	Li and Baker (2018)	Computers and Education	Q1
2018	Li et al. (2018)	Research and Practice in Technology Enhanced Learning	Q2
2018	Liu et al. (2018)	Journal of Information Science and Engineering	Q4
2018	Ruipérez-Valiente et al. (2018)	Expert Systems	Q2
2018	Saqr et al. (2018)	PloS One	Q2
2018	Tan et al. (2018)	Proceedings of 2018 International Symposium on Educational Technology	Not ranked
2018	Theobald et al. (2018)	Learning and Individual Differences	Q1
2017	Conijn et al. (2017)	IEEE Transactions on Learning Technologies	Q1
2017	Ellis et al. (2017)	Educational Technology & Society	Q2
2017	Grubišić et al. (2017)	Proceedings of the 9th International Conference on Computer Supported Education	B - CORE
2017	Jo et al. (2017)	Journal of Computer Assisted Learning	Q1
2017	Li et al. (2017)	Proceedings of the 5th International Conference on Educational Innovation through Technology	Not ranked
2017	Lin et al. (2017)	Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics	Not ranked
2017	Mwalumbwe and Mtebe (2017)	Electronic Journal of Information Systems in Developing Countries	Q2
2017	Pardo et al. (2017)	IEEE Transactions on Learning Technologies	Q1
2017	Scheffel et al. (2017)	IEEE Transactions on Learning Technologies	Q1
2017	Widyahastuti et al. (2017)	International Conference on e-Learning	B4 - Qualis
2016	Choi et al. (2016)	Proceedings of the 9th International Conference on Educational Data Mining	B - CORE
2016	de Marcos et al. (2016)	Computers in Human Behavior	Q1
2016	Goggins and Xing (2016)	Computers and Education	Q1
2016	Naumann and Salmerón (2016)	International Review of Research in Open and Distance Learning	Q2
2016	Strang (2016)	Education and Information Technologies	Q1

(Continues)

TABLE 2 (Continued)

Year	Authors	Source title	Rank
2016	Yamada et al. (2016)	ICCE 2016 - 24th international conference on computers in education: think global act local - main conference proceedings	B - CORE
2016	You (2016)	The Internet and Higher Education	Q1
2015	Joksimović, Gašević, Kovanović, et al. (2015)	Journal of Computer Assisted Learning	Q1
2015	Joksimović, Gašević, Loughin, et al. (2015)	Computers and Education	Q1
2015	Junco and Clem (2015)	Internet and Higher Education	Q1
2015	Jo et al. (2015)	Educational Technology & Society	Q2
2015	Kennedy et al. (2015)	Proceedings of the 5th International Conference on Learning Analytics & Knowledge (LAK 2015)	A - CORE
2015	Pardo et al. (2015)	Proceedings of the 5th International Conference on Learning Analytics & Knowledge (LAK 2015)	A - CORE
2015	Svihla et al. (2015)	Journal of Learning Analytics	Q1
2015	Wei et al. (2015)	Computers and Education	Q1
2015	You (2015)	Educational Technology & Society	Q2
2015	Zacharis (2015)	The Internet and Higher Education	Q1
2014	Agudo-Peregrina et al. (2014)	Computers in Human Behavior	Q1
2014	Yoo and Kim (2014)	International Journal of Artificial Intelligence in Education	Q3
2014	Yu and Jo (2014)	Proceedings of the 4th International Conference on Learning Analytics & Knowledge (LAK 2014)	A - CORE
2013	Gijlers and de Jong (2013)	Journal of the Learning Sciences	Q1
2013	Lin and Chiu (2013)	Proceedings - 2013 IEEE 13th International Conference on Advanced Learning Technologies	B - CORE
2013	Miller and Soh (2013)	Proceedings - Frontiers in Education Conference	C - CORE
2013	Ritter et al. (2013)	Proceedings of the 6th International Conference on Educational Data Mining	B - CORE
2012	Bernacki et al. (2012)	Contemporary Educational Psychology	Q1
2012	Romero-Zaldivar et al. (2012)	Computers and Education	Q1

Note: According to CORE 2021 summary: A*—7.22% of 803 ranked venues; A—16.06% of 803 ranked venues; B—37.11% of 803 ranked venues; Australasian B—1.62% of 803 ranked venues; C—36.24% of 803 ranked venues; Australasian C—1.74% of 803 ranked venues; Other—167 total. Qualis: This conference ranking has been published by the Brazilian ministry of education and uses the H-index as performance measure for conferences. Based on the H-index percentiles, the conferences are grouped into performance classes that range from A1 (=best), A2, B1, …, B5 (=worst).

meta-regression models to examine the influence of one or more moderator variables on the outcomes. It can handle both continuous and categorical moderator variables. Furthermore, it includes functions for fitting fixed-effects and random-effects models.

The meta-regression analysis in the current study includes the following steps. Firstly, each regression coefficient was transformed into a standard normal metric, Fisher's *Z* score, as an effect size. Secondly, Cochran's *Q*-Test (Cochran, 1950) and the I^2 statistic (Higgins et al., 2002) were used for the heterogeneity test. Heterogeneity in meta-analysis refers to the variation in study outcomes between studies. Cochran's *Q* is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies, with the weights being those used in the pooling method. The I^2 statistic describes the percentage of variation across studies that is due to heterogeneity rather than chance. I^2 values of 25%, 50% and 75%, correspond to small, moderate and large amounts of heterogeneity among studies. Thirdly, the funnel plot and Egger regression test were used to test whether the results were biased due to different publication sources. Finally, moderator analyses were performed.

Log variables can be classified into generic and course-specific, basic and elaborated variables. We used independent samples *T* test to test whether the effect sizes of generic and course-specific, basic and elaborated log variables were significantly different.

RESULTS

This section reports the findings and discoveries by considering the research objectives of this review study.

Soundness and quality assurance of the dataset of the selected studies

Out of 69 journal publications and 19 conference studies, 78% of the studies are published in Q1 and Q2 ranking journals and A level conferences. This finding shows the soundness of the selected studies that all studies are thorough research by technology and analytical field. Table 2 shows that the highest number of selected studies that were published in one journal is 10, and they are published in *Computers and Education*, followed by *Journal of Computer Assisted Learning*, which provided 6 publications in the selected studies.

Distribution of the reviewed studies over time

The current review focused on studies from January 2010 to December 2021. Over that time, we noticed a gradual rise in the number of selected studies that met our research objectives. Figure 2 shows the highest number of studies published in 2021 (n = 17) and no studies that met our inclusion criteria published before 2012.

Sample size of the reviewed studies

Table 3 provides the sample size, the learning type and theme investigated in the reviewed studies. Figure 3 shows sample sizes of the reviewed studies. It reveals a great variation in recruitment. There was a total of 106 samples investigated in the reviewed studies. Most models (n = 45) were performed on a sample size of between 100 and 500. Some models (n = 21) were performed on a sample size of between 500 and 1000. It indicates that most reviewed studies have fairly large sample sizes, which enables researchers to validate the regression models with more solid evidence.

Learning type

We coded the learning type in the reviewed studies as “fully online” and “blended”. Typically, online learning is the use of web-based technologies to provide out-of-class learning in the

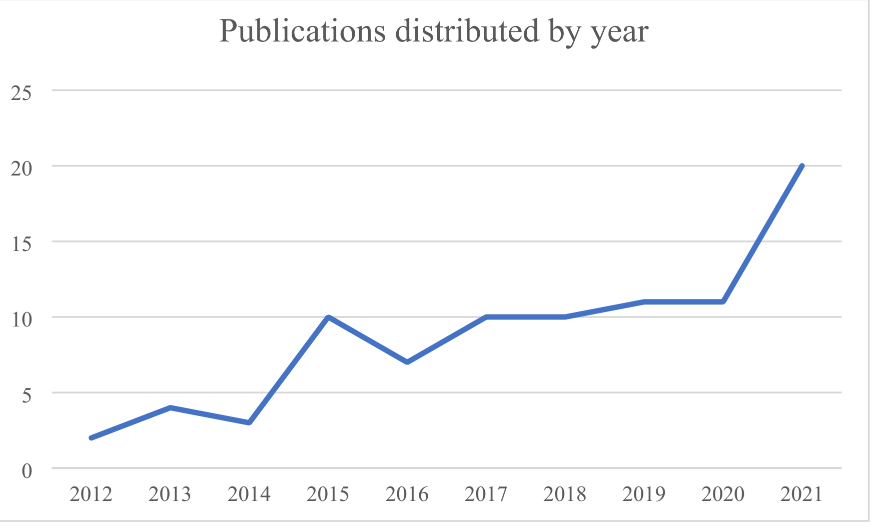

FIGURE 2 Publications distributed by year.

absence of the physical classroom, which enables learning without time, place and pace constraints (Bernard et al., 2014; Chigeza & Halbert, 2014; Israel, 2015). It is launched through LMSs or virtual learning environments (VLE) such as Moodle and Blackboard. Blended learning refers to the integration of traditional face-to-face learning and online learning, which is seen as a way of combining the benefits of two formats (Adams et al., 2015). In the current study, we coded the course or learning task that integrated face-to-face class session and an online learning tool or platform as blended. As shown in Appendix A, apart from 4 studies that did not specify the course type, 44 courses or learning tasks investigated in the reviewed studies were delivered fully online, while the other 59 were blended learning.

Learning theme

Learning theme investigated in the reviewed studies can be classified into 9 groups as we mentioned in the previous section. Figure 4 shows that the highest number of courses or learning tasks were related to the field of STEM, followed by K-12 education.

The type of dependent variable in the regression models

We found that there were two kinds of dependent variables used in the regression models, total course score and exam score. We coded them as “total score” and “exam score”. Appendix B provides the information of dependent variables and log variables, and regression coefficients of log variables. There were totally 112 independent regression models, in which 68 models used exam score as the dependent variable.

The type of log variables in the regression models

A total 328 log variables were found to significantly predict student academic achievement. There are 161 generic and 167 course-specific, 236 basic and 92 elaborated log variables

TABLE 3 Twelve log variables in the reviewed studies

Log variables	Description	Frequency	Study	Regression coefficients
Login time	Overall time spent online	13	Jovanović et al. (2021)	4.077
			Mills (2021)	0.110
			Sharma et al. (2020)	0.591
			Li et al. (2018)	0.104
			Conijn et al. (2017)[2]	−0.190
			Conijn et al. (2017)[6]	0.250
			Conijn et al. (2017)[9]	0.340
			Conijn et al. (2017)[10]	−0.340
			Conijn et al. (2017)[11]	−0.180
			Conijn et al. (2017)[12]	0.760
			Conijn et al. (2017)[14]	−0.230
			Joksimović, Gašević, Loughin, et al. (2015)	0.030
			Yu and Jo (2014)	0.238
Login frequency	Total number of learners' online access	6	Sharma et al. (2020)	0.287
			Li et al. (2017)	−0.044
			Strang (2016)	0.137
			You (2016)[1]	0.150
			You (2016)[2]	−0.180
			Wei et al. (2015)	0.060
Regularity of study time	Standard deviation of study time	5	Jost et al. (2021)	−0.310
			Conijn et al. (2017)[6]	0.260
			Conijn et al. (2017)[7]	−0.570
			Conijn et al. (2017)[10]	0.170
			Conijn et al. (2017)[13]	−0.400
Regularity of study interval	Standard deviation of study interval	9	Li et al. (2018)	−0.158
			Conijn et al. (2017)[1]	−0.370
			Conijn et al. (2017)[2]	−0.320
			Conijn et al. (2017)[4]	−0.190
			Conijn et al. (2017)[8]	−0.520
			Conijn et al. (2017)[10]	−0.220
			Conijn et al. (2017)[15]	−0.530
			Jo et al. (2017)	−0.270
			Jo et al. (2015)	−0.590

(Continues)

TABLE 3 (Continued)

Log variables	Description	Frequency	Study	Regression coefficients
Frequency of viewing course pages	Total number of learners' viewing course pages	16	Schumacher and Ifenthaler (2021)	0.218
			Han and Ellis (2020)	0.220
			Soffer and Cohen (2019)	0.220
			Tan et al. (2018)	0.275
			Conijn et al. (2017)[1]	0.310
			Conijn et al. (2017)[2]	0.330
			Conijn et al. (2017)[3]	0.350
			Conijn et al. (2017)[5]	0.630
			Conijn et al. (2017)[8]	−1.380
			Conijn et al. (2017)[9]	−0.420
			Conijn et al. (2017)[10]	−2.280
			Conijn et al. (2017)[11]	−0.520
			Ellis et al. (2017)	0.880
			Pardo et al. (2017)	0.850
			Joksimović, Gašević, Loughin, et al. (2015)	−0.090
			Scheffel et al. (2017)	−0.287
Number of clicks	Total number of clicks on online learning platform	10	Tempelaar et al. (2020)[1]	0.080
			Tempelaar et al. (2020)[2]	0.061
			Conijn et al. (2017)[1]	−0.200
			Conijn et al. (2017)[2]	−0.390
			Conijn et al. (2017)[3]	−0.360
			Conijn et al. (2017)[7]	−0.300
			Conijn et al. (2017)[8]	1.300
			Conijn et al. (2017)[10]	0.140
			Conijn et al. (2017)[11]	0.210
			Conijn et al. (2017)[12]	1.910
Time-on-task	Time spent on completing the learning task	5	Stadler et al. (2020)	0.090
			Tacoma et al. (2021)	0.190
			Tacoma et al. (2020)	0.280
			Tempelaar et al. (2020)[2]	−0.127
			Zarrabi and Bozorgian (2020)	0.750
Average time per online session	Average time studying each session	4	Conijn et al. (2017)[2]	0.220
			Conijn et al. (2017)[7]	0.350
			Conijn et al. (2017)[11]	0.130
			Conijn et al. (2017)[15]	−0.180

TABLE 3 (Continued)

Log variables	Description	Frequency	Study	Regression coefficients
Number of online sessions	Total number of active online learning sessions	7	Conijn et al. (2017)[2]	0.240
			Conijn et al. (2017)[9]	0.550
			Conijn et al. (2017)[10]	0.440
			Conijn et al. (2017)[11]	0.240
			Conijn et al. (2017)[13]	0.600
			Conijn et al. (2017)[14]	0.460
			Lin and Chiu (2013)	0.414
Starting late	Time until the first learning activity	6	Conijn et al. (2017)[2]	−0.100
			Conijn et al. (2017)[4]	−0.270
			Conijn et al. (2017)[7]	−0.580
			Conijn et al. (2017)[8]	−0.270
			Conijn et al. (2017)[15]	−0.120
			Foung and Chen (2019)	−0.078
Late submission	Number of learners' failure to submit assignments on time	4	You (2016)[1]	−0.360
			You (2016)[2]	−0.180
			You (2015)[1]	−0.400
			You (2015)[2]	−0.210
Largest period of inactivity	Largest period of not active online learning activities	10	Conijn et al. (2017)[1]	0.160
			Conijn et al. (2017)[2]	0.150
			Conijn et al. (2017)[3]	−0.250
			Conijn et al. (2017)[7]	−0.560
			Conijn et al. (2017)[8]	0.560
			Conijn et al. (2017)[10]	0.320
			Conijn et al. (2017)[11]	0.190
			Conijn et al. (2017)[13]	0.330
			Conijn et al. (2017)[15]	0.500
			Conijn et al. (2017)[16]	0.720

(Appendix B). It indicates that most studies investigated the frequency and time counts measures. Only a small number of log variables were aggregated and elaborated indicators.

Meta-regression analyses

We performed meta-regression analyses on 161 generic log variables. According to Fu et al. (2011), the sizes of the included studies should be at least 6 to 10 studies for a continuous study level variable; and for a (categorical) subgroup variable, each subgroup should have a minimum of 4 studies. Therefore, we summarized 12 log variables that were frequently found to be significant in the reviewed studies, including login time, login frequency, regularity of study time and interval, frequency of viewing course pages, number of clicks, average time per online session, time-on-task, number of online sessions, starting late, late submission

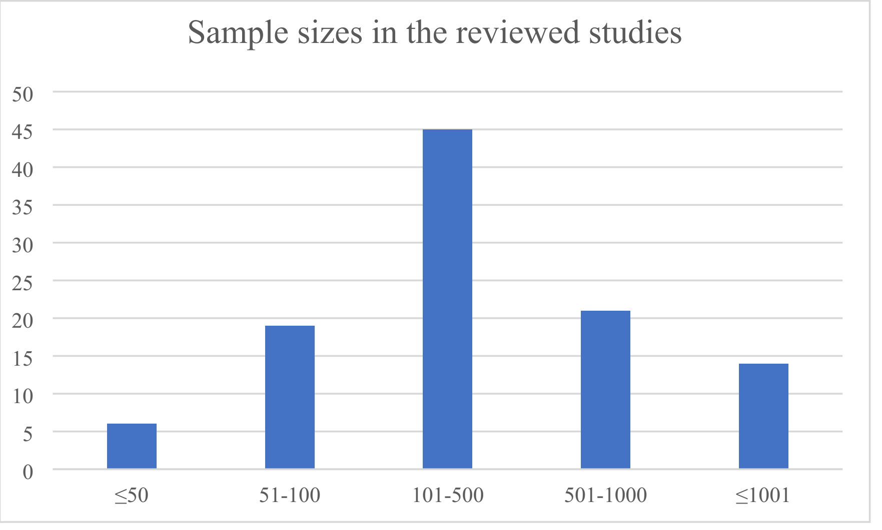

FIGURE 3 Sample sizes in the reviewed studies.

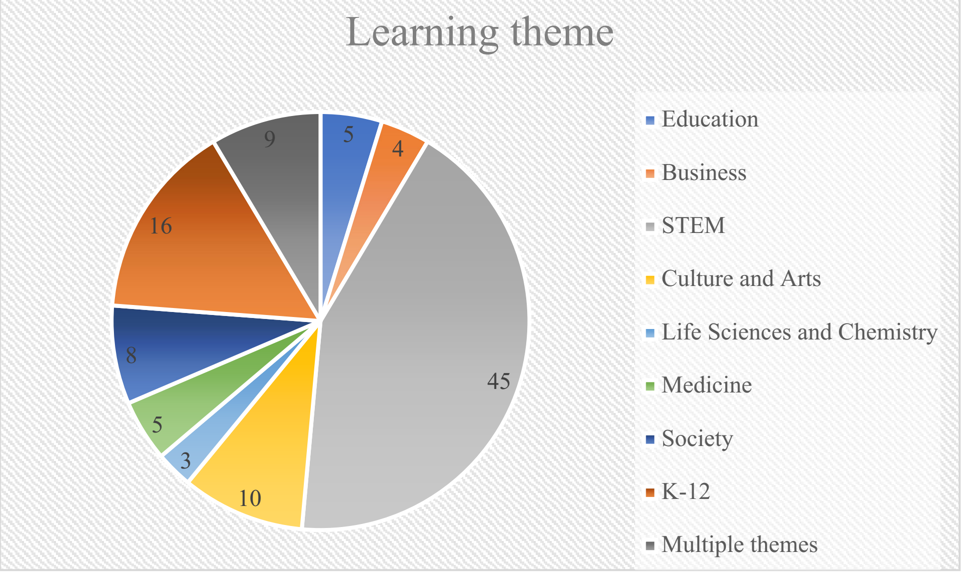

FIGURE 4 Learning theme investigated in the reviewed studies.

and largest period of inactivity (Table 3). We conducted separate meta-analyses, one for each subgroup. Because the metadata of average time per online session shows only one level on all the moderators, we did not perform meta-regression on this log variable. Not all subgroups show differences on all the moderating variables. We only did meta-analyses on the moderating variables with varying levels.

Heterogeneity

Since each subgroup gets its own separate meta-analysis, estimates of the heterogeneity will also differ from subgroup to subgroup. When the number of studies in a subgroup is small, it is likely that the estimate of heterogeneity will be imprecise (Borenstein et al., 2009). Therefore, in practice, the estimate of heterogeneity is pooled across subgroups. The heterogeneity test results were Q = 10,100.70 (df = 145, p<0.001, I^2 = 98.56%), so the random effects model was chosen.

Publication bias

Figure 5 shows that all the 160 effect sizes of generic log variables are evenly distributed on both sides and gather at the middle and upper part of the plot. Some studies have statistically significant effect sizes (the gray areas), others do not (the white area). The Egger regression reveals no significant bias with z = 0.96 (p>0.05). Therefore, we can conclude that the results were not biased due to the publication sources.

Mean effect size

For login time, the integrated results show a significantly positive effect on student academic achievement (β = 0.17, z = 14.77, p<0.001, 95% CI = [0.15, 0.19]), which means students whose login time one standard deviation above the mean would have a grade that is 0.17 of a standard deviation above the average grade. For login frequency, the integrated results show a significantly positive effect on student academic achievement (β = 0.04, z = 2.53, p<0.05, 95% CI = [0.01, 0.06]), which means students whose login frequency one standard deviation above the mean would have a grade that is 0.04 of a standard deviation above the average grade. For regularity of study time, the integrated results show no significant effect on student academic achievement (β = −0.04, z = 25.31, p>0.05). For regularity of study interval, the integrated results show a significantly negative effect on student academic achievement (β = −0.28, z = −22.02, p<0.001, 95% CI = [−0.30, −0.25]), which means students

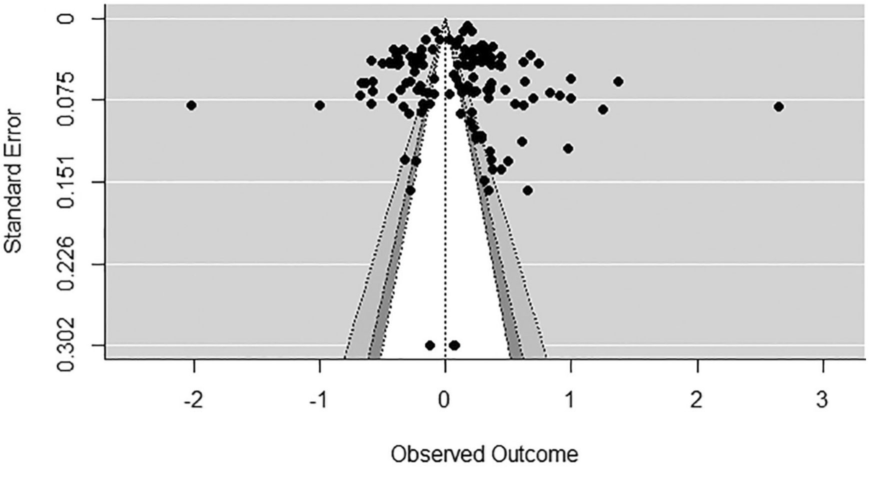

FIGURE 5 Contour-enhanced funnel plot.

whose standard deviation of study interval one standard deviation above the mean would have a grade that is 0.28 of a standard deviation below the average grade. For frequency on course pages, the integrated results show a significantly positive effect on student academic achievement ($\beta = 0.29$, $z = 22.49$, $p < 0.001$, 95% CI = [0.26, 0.31]), which means students whose frequency on course pages one standard deviation above the mean would have a grade that is 0.29 of a standard deviation above the average grade. For the number of clicks, the integrated results show a significantly negative effect on student academic achievement ($\beta = -0.27$, $z = -15.80$, $p < 0.001$, 95% CI = [−0.30, −0.23]), which means students whose number of clicks one standard deviation above the mean would have a grade that is 0.27 of a standard deviation below the average grade. For time-on-task, it shows a significantly positive effect on student academic achievement ($\beta = 0.24$, $z = 0.03$, $p < 0.001$, 95% CI = [0.19, 0.29]), which means students whose time-on-task one standard deviation above the mean would have a grade that is 0.24 of a standard deviation above the average grade. For the number of online sessions, it shows a significantly positive effect on student academic achievement ($\beta = 0.41$, $z = 22.24$, $p < 0.001$, 95% CI = [0.37, 0.44]), which means students whose number of online sessions one standard deviation above the mean would have a grade that is 0.41 of a standard deviation above the average grade. For starting late, the integrated results show a significantly negative effect on student academic achievement ($\beta = -0.12$, $z = -11.94$, $p < 0.001$, 95% CI = [−0.14, −0.10]), which means students who start learning late one standard deviation above the mean would have a grade that is 0.12 of a standard deviation below the average grade. For late submission, the integrated results show a significantly negative effect on student academic achievement ($\beta = -0.21$, $z = -9.91$, $p < 0.001$, 95% CI = [−0.25, −0.17]), which means students whose level of procrastination one standard deviation above the mean would have a grade that is 0.21 of a standard deviation below the average grade. For largest period of inactivity, the integrated results show a significant positive effect on student academic achievement ($\beta = 0.14$, $z = 8.99$, $p < 0.001$, 95% CI = [0.11, 0.17]), which means students whose largest period of inactivity one standard deviation above the mean would have a grade that is 0.14 of a standard deviation above the average grade. The above results indicate that the moderator analysis was suitable for the log variables of login time, login frequency, regularity of study interval, number of clicks, frequency on course pages, time-on-task, number of online sessions, starting late, late submission and largest period of inactivity.

Moderator analyses

Sample range

As shown in Table 4, the moderator test of sample range was found to be significant in four log variables, including regularity of study interval ($Q = 45.65$, $df = 4$, $p < 0.001$), time-on-task ($Q = 82.15$, $df = 4$, $p < 0.001$), number of online sessions ($Q = 39.66$, $df = 4$, $p < 0.001$), and starting late ($Q = 9.57$, $df = 3$, $p < 0.05$). We can see that different sample ranges show significantly different effects on the effect size of these four log variables. It indicates that sample size has an unsteady impact on the regression analysis in the reviewed studies.

Learning type

Table 5 shows that the moderator test of learning type was only found to be significant in the log variable of regularity of study interval ($Q = 29.89$, $df = 2$, $p < 0.001$). From Table 5, we can see that fully online learning shows a stronger effect on the effect size of regularity of study interval than blended learning ($\beta_{fully\ online} = -0.40$, $\beta_{blended} = -0.37$). It indicates that the

TABLE 4 Moderator analysis of sample range

Log variables	Q	df	Sample range	Estimate	Z	CI
Login time	695.82	12				
Login frequency	112.92	5				
Regularity of study interval	45.65***	4	≤50	−0.28		
			101–500	−0.52***	−5.81	[−0.69, −0.34]
			501–100	−0.29*	−2.46	[−0.52, −0.06]
			<1000	−0.24*	−2.09	[−0.47, −0.02]
Frequency on course pages	2.75	4				
Number of clicks	6.96	4				
Time-on-task	82.15***	4	≤50	−0.13		
			51–100	0.97***	7.41	[0.72, 1.23]
			101–500	0.09		
			501–1000	0.24***	5.06	[0.15, 0.33]
Number of online sessions	39.66***	4	51–100	0.50*	2.21	[0.06, 0.94]
			101–500	0.47***	4.14	[0.25, 0.69]
			501–1000	0.53***	3.99	[0.27, 0.79]
			<1000	0.25		
Starting late	9.57*	3	101–500	−0.35**	−2.76	[−0.61, −0.10]
			501–1000	−0.28	−1.27	[−0.70, 0.15]
			<1000	−0.08	−0.58	[−0.38, 0.21]
Largest period of inactivity	3.68	3				

TABLE 5 Moderator analysis of learning type

Log variables	Q	df	Learning type	Estimate	Z	CI
Login time	635.32	12				
Login frequency	109.19	5				
Regularity of study interval	29.89***	2	Fully online	−0.40***	−2.81	[−0.69, −0.12]
			Blended	−0.37***	−4.69	[−0.52, −0.22]
Frequency on course pages	4.66	2				
Time-on-task	3.74	2				

impact of regular study on academic achievement is stronger when students take fully online courses or learning tasks.

Learning theme

As shown in Table 6, the moderator test of learning theme was found to be significant in four log variables, including regularity of study interval (Q = 49.86, df = 4, $p<0.001$), time-on-task (Q = 79.04, df = 3, $p<0.001$), number of online sessions (Q = 118.47, df = 3, $p<0.001$), and starting late (Q = 8.35, df = 2, $p<0.05$). We can see that the courses or learning tasks in the field of business indicate stronger effect on the effect size of regularity of study interval ($\beta_{business}$ = −0.68) and courses or learning tasks in the field of STEM show a weaker effect (β_{STEM} = −0.38). The effect size of time-on-task was stronger in the field of CA than that in

TABLE 6 Moderator analysis of learning theme

Log variables	*Q*	*df*	Learning theme	Estimate	*Z*	CI
Login time	660.92	12				
Login frequency	111.88	5				
Regularity of study interval	49.86***	4	Business	−0.68***	−3.95	[−1.01, −0.34]
			STEM	−0.38***	−5.63	[−0.51, −0.25]
			CA	−0.16		
			Society	−0.28		
Frequency on course pages	6.55	5				
Number of clicks	2.06	2				
Time-on-task	79.04***	3	CA	0.97***	7.37	[0.71, 1.23]
			Society	0.23***	4.83	[0.14, 0.33]
			K-12	0.09		
Number of online sessions	118.47***	3	STEM	0.34***	5.70	[0.22, 0.45]
			Society	0.65***	8.27	[0.50, 0.80]
			Multiple themes	0.44***	4.19	[0.23, 0.65]
Starting late	8.35*	2	STEM	−0.29**	−2.87	[−0.48, −0.09]
			CA	−0.08	−0.36	[−0.50, 0.35]
Largest period of inactivity	2.66	2				

Society (β_{CA} = 0.97, $\beta_{society}$ = 0.23). The courses or learning tasks in the field of society indicate stronger effect on the effect size of number of online sessions than those in the field of STEM ($\beta_{society}$ = 0.65, β_{STEM} = 0.34). The courses or learning tasks in the field of STEM indicate a significant effect on the effect size of starting late (β_{STEM} = 0.33), while courses or learning tasks in the field of CA show no significant effect.

The type of the dependent variable

For the type of the dependent variable, Table 7 shows that the moderator test was significant in four log variables, including regularity of study interval (Q = 31.87, df = 2, $p<0.001$), number of online sessions (Q = 38.44, df = 2, $p<0.001$), starting late (Q = 8.35, df = 2, $p<0.05$), and late submission (Q = 8.11, df = 2, $p<0.05$). We can see that the type of dependent variable in the regression models significantly moderates the effect size of regularity of study interval ($\beta_{total\ score}$ = −0.28, $\beta_{exam\ score}$ = −0.39). It also moderates the effect size of number of online sessions ($\beta_{total\ score}$ = 0.44, $\beta_{exam\ score}$ = 0.46). Using exam score as the dependent variable shows a significant effect on the effect size of starting late ($\beta_{exam\ score}$ = −0.29), while using total score as the dependent variable shows no significant effect. However, for the effect size of late submission, using exam score as the dependent variable indicates no significant effect, while using total score as the dependent variable shows a significant effect ($\beta_{exam\ score}$ = −0.40).

T test on the type of log variables

The results of independent samples *T* test show that there is no significant difference between the effect size of the generic and course-specific log variables (t = −1.62, df = 291, $p>0.05$). Significant difference was found between the effect size of basic and elaborated

TABLE 7 Moderator analysis of the type of the dependent variable

Log variables	Q	df	Dependent variable type	Estimate	Z	CI
Login time	731.32	12				
Login frequency	115.99	5				
Regularity of study interval	31.87***	2	Total score	−0.28*	−1.12	[−0.26, −0.76]
			Exam score	−0.39***	−5.53	[−0.52, −0.25]
Frequency on course pages	4.67	2				
Number of online sessions	38.44***	2	Total score	0.44*	2.35	[0.07, 0.81]
			Exam score	0.46***	5.74	[0.30, 0.61]
Starting late	8.35*	2	Total score	−0.08	−0.36	[−0.50, 0.35]
			Exam score	−0.29**	−2.87	[−0.48, −0.09]
Late submission	8.11*	2	Total score	−0.40**	−2.85	[−0.68, −0.12]
			Exam score	−0.02	−0.11	[−0.29, 0.26]

log variables ($t = 2.269$, $df = 100$, $p < 0.001$). We also calculated Cohen's D to examine the extent of the significant difference. The result was 0.33, which indicates a small effect.

DISCUSSION

As the first meta-analysis of the log variables in predicting academic achievement, this paper presents a synthesis of the effect sizes related to the log variables based on empirical research. Four potential moderators were examined (ie, sample size, learning type, learning theme, the type of the dependent variable), and the moderating effect was found to be different across different log variables. Furthermore, the effect sizes of different types of log variables were compared, in order to provide a deep understanding of how LA studies investigated log data.

Influential log variables

Based on the systematic review, we found that there is no wide variation in the learning type, the dependent variable of regression models, and the generic or course-specific log variables investigated in the reviewed studies. However, most reviewed studies investigated STEM courses and learning tasks. Furthermore, most log variables were generated from the basic frequency and time counts. Although only a small number of elaborated log variables were found to be significantly associated with academic achievement, the vigorous development of AI affords opportunities to improve the assessment of processes and enables the various possibility of the extraction and aggregation of log data: further calculation (eg, ratio, entropy), theory-driven LA-based variables (eg, regularity, procrastination), and network analysis-based variables (eg, density, centrality). Furthermore, variables are extracted from multichannel data, based on activities on not just LMSs but also AI-based tools (eg, e-book, video annotation application, notetaking, highlighting and bookmarks), which implies that promising directions for assessing learning processes are being developed with different AI techniques. The review results also show that the generation of some log variables varies across studies. For example, some studies investigated regularity of study time or regularity of study interval, which was calculated based on standard deviation. However, the log variable of regular study in the work of You (2016) was calculated based on the virtual attendance score. It indicates a lack of a uniform paradigm in generating complex log variables.

Among 12 generic log variables that were most frequently found to be significant in predicting academic achievement, most of them showed positive effects in some courses and learning tasks but negative in others. Only the positive effect of number of online sessions and negative effects of starting learning late and regularity of study interval were found in all the reviewed studies. A session was defined as the sequence of behavior from the first click after the login to the LMS until the last click before logging out, or the last click before staying inactive for at least 40 minutes (Conijn et al., 2017). The number of online sessions is a basic indicator that reflects student active interactions with the LMSs. Starting late is a measure of the time until the first activity, which is a more complex predictor relating to time management. Regularity of study interval is calculated by standard deviation of the study intervals. Therefore, this variable technically means the "irregular study". It indicates that the association between log variables that reflect regular study and time management behaviors and academic achievement was steadier across different learning contexts. Positive effect of regular study on student learning has been confirmed in many studies (eg, Conijn et al., 2017; You, 2016). You (2016) pointed out that successful students would actively participate in their learning, such as regularly accessing online courses and completing assignments in a timely manner. Furthermore, previous studies also demonstrated the detrimental effects of starting to learn late on learning success (eg, Levy & Ramin, 2012; Michinov et al., 2011). Therefore, students who start the course early, regularly access the course and actively participate in the learning activities would be more likely to have a higher grade.

Moderator analyses

We performed a meta-regression on eleven generic log variables. Only regularity of study time was not suitable for the moderator analysis. The significant effect was found on other ten log variables. Mean effect sizes of these log variables show that login time, frequency on course pages, number of online sessions and largest period of inactivity positively impact student academic achievement, while starting late, regularity of study interval and late submission have a negative impact. Not all moderators influence the effect sizes of these log variables on all levels. For the moderator role of sample range, only regularity of study interval, time-on-task, number of online sessions and starting late are significantly moderated. For the learning type, it only significantly moderates the effect size of regularity of study interval. Students who regularly participate in fully online learning would have higher grades than studying blended courses. This result is logical because student academic achievement in fully online courses basically depends on how students perform online learning activities. The learning theme only significantly moderates the effect sizes of regularity of study interval, time-on-task, and number of online sessions, and starting late. The type of dependent variable in the regression models significantly moderates the effect size of regularity of study interval, starting late and late submission. These results suggest that the effect sizes of log variables relating to regular study, time management and active interactions vary across different courses and learning tasks. For building generalizable prediction models for multiple courses, more theoretical reasoning is needed to aggregate effective and meaningful log predictors that accurately reflect underlying concepts and can be widely applied, for example, optimizing the measurements of regular study and distributed learning across courses based on learning theories and a more fine-grained disentanglement of learning content. In order to do so, we need to better understand what the measurements are actually measuring, what the effect is and how to translate it into specific measures of previously defined theoretical concepts (Conijn et al., 2017). This necessitates the introduction of theories into LA. As suggested by Gašević et al. (2015), Gašević et al. (2016), Wise and Shaffer (2015), theory-driven LA can help researchers gain more dynamic and replicable insights into the learning process rather than the static prediction of academic outcomes in a single case.

Difference in effect size

We found that there was no difference of the effect sizes between generic and course-specific log variables, while a significant difference between the basic and elaborated was found. Even though the generic and course-specific log variables have no significant difference in the prediction power, previous research has found that predictive models with only generic indicators were able to explain only a small portion of the overall variability in the students' course performance (Jovanović et al., 2019). It has also emphasized the quality of learning behaviors rather than the quantity of learning (Jo & Kim, 2013; You, 2015; You, 2016). As we mentioned in the previous section, theory-driven LA is needed for better conversion of log data into elaborated variables. For example, guided by the theory of self-regulated learning (SRL), You (2016) found that the variable of regular study was a more persuasive indicator than simple frequency measures. It further proves the crucial role theory plays in justifying selecting variables, developing models and interpreting results. Therefore, more complex variables based on well-defined theoretical frameworks warrants consideration.

LIMITATIONS AND FUTURE RESEARCH

Although the present review study provides insight into the effect of influential log variables on student academic achievement, there are still some limitations. First, we only focused on the statistically significant log variables in regression analysis. Future review studies can examine why some log variables were significant in some studies while insignificant in others. Second, not all categorical levels of some moderators were available in some generic log variables, which limits the meta-regression analysis. Future research could broaden the scope of the search to obtain more diverse metadata. Third, we did not examine the prediction for students at risk of failing a course or student dropout which deals with a categorical dependent variable. Further research can be conducted on this issue.

Regarding the direction of future research on predicting student academic achievement, most reviewed studies mainly examined frequency and time duration log variables. Additional elaborated indicators that reflect students' levels of engagement, motivation (eg, effort, persistence), and decision-making in online learning should be explored and tested in this field. This requires the integration of educational and learning theories and LA, such as, how well learners self-regulate their learning and how learners build connections with others. Furthermore, the application of AI in assessment brings a new set of challenges, for example, the sidelining of professional expertise in automated decision-making generated by AI approaches (Swiecki et al., 2022). Researchers can study on making AI-based assessments more explainable to the teacher for balancing between AI and teacher decision-making on teaching, learning, and assessment.

CONCLUSIONS

The present study reviewed 88 empirical studies on using log variables to predict student academic achievement. The main aim of this study is to explore the significant influential log variables and whether their effect size was moderated by factors reflecting different learning contexts. This research made three main contributions: (1) presents a new perspective of the log variables, which provides a reliable quantitative conclusion of log variables in predicting student academic achievement; (2) conducted subgroup analysis, examined four potential moderating variables, and identified varying degrees of moderating effect on several log variables including regularity of study interval, time-on-task, number of online sessions, starting late, and late submission; (3) compared the effect of generic and course-specific, basic and elaborated log variables, and found significant difference between the basic and elaborated. The findings help understand the role of log variables in predicting student academic achievement and make it clear that the effect size of certain generic log variables varies across the research contexts. For instructors, a depth understanding of log variables can help them obtain the continuous views of learners' engagement and infer student knowledge and learning in online learning, thus overcoming "one size fit all" approach and making appropriate decisions based on AI-based data outputs, for example, providing personalized teaching or remedial instruction for students. For researchers, the exploration of the effect size of log variables enables the generation of more robust log variables which best represent and reflect the true learning process of learners.

ACKNOWLEDGEMENTS
The authors thank the editor and reviewers for the careful and insightful review and providing valuable feedback.

FUNDING INFORMATION
This research received no external funding.

CONFLICT OF INTEREST
No potential conflict of interest was reported by the authors.

DATA AVAILABILITY STATEMENT
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ETHICS STATEMENT
This research didn't involve human subjects and animals.

ORCID
Amin Mousavi https://orcid.org/0000-0002-6920-2319

REFERENCES
Adams, A. E. M., Randall, S., & Traustadóttir, T. (2015). A tale of two sections: An experiment to compare the effectiveness of a hybrid versus a traditional lecture format in introductory microbiology. *CBE - Life Sciences Education*, *14*, 1–8.

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*, 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Algayres, M., & Triantafyllou, E. (2020). Learning analytics in flipped classrooms: A scoping review. *Electronic Journal of E-Learning*, *18*(5), 397–409. https://doi.org/10.34190/JEL.18.5.003

Bernacki, M. L., Byrnes, J. P., & Cromley, J. G. (2012). The effects of achievement goals and self-regulated learning behaviors on reading comprehension in technology-enhanced learning environments. *Contemporary Educational Psychology*, *37*(2), 148–161. https://doi.org/10.1016/j.cedpsych.2011.12.001

Bernard, M. B., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, *26*(1), 87–122.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Subgroup analyses. In S. T. River (Ed.), *Introduction to meta-analysis* (pp. 149–186). John Wiley & Sons. https://doi.org/10.1002/9780470743386.ch19

Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, *115*, 106595. https://doi.org/10.1016/j.chb.2020.106595

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, *8*, 13–25 http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf

Chan, A. K. M., Botelho, M. G., & Lam, O. L. T. (2021). The relation of online learning analytics, approaches to learning and academic achievement in a clinical skills course. *European Journal of Dental Education, 25*(3), 442–450. https://doi.org/10.1111/eje.12619

Chen, L., Goda, Y., Shimada, A., & Yamada, M. (2019). Factors investigation of learning behaviors affecting learning performance and self-regulated learning. In *TALE 2019 - 2019 IEEE International Conference on Engineering, Technology and Education [9225926] (TALE 2019 - 2019 IEEE International Conference on Engineering, Technology and Education)*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/TALE48000.2019.9225926

Chigeza, P., & Halbert, K. (2014). Navigating e-learning and blended Learning for pre-service teachers: Redesigning for engagement, access and efficiency. *Australian Journal of Teacher Education*, *39*(11), 133–146. https://doi.org/10.14221/ajte.204v39n11.8

Chiu, T. K. F., & Hew, T. K. F. (2018). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian Journal of Educational Technology*, *34*(4), 16–28. https://doi.org/10.14742/ajet.3240

Choi, H., Lee, J. E., Hong, W., Lee, K., Recker, M., & Walker, A. (2016). Exploring learning management system interaction data: Combining data-driven and theory-driven approaches. In *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016* (pp. 324–329). International Educational Data Mining Society (IEDMS).

Clarivate. (2021). 2020 Journal Citation Indicator. *Journal Citation Reports*.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3–4), 256–266. https://doi.org/10.1093/biomet/37.3-4.256

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. https://doi.org/10.1109/TLT.2016.2616312

Conijn, R., van den Beemt, A., & Cuijpers, P. (2018). Predicting student performance in a blended MOOC. *Journal of Computer Assisted Learning*, *34*(5), 615–628. https://doi.org/10.1111/jcal.12270

Cooper, S., & Sahami, M. (2013). Reflections on Stanford's MOOCs. *Communications of the ACM*, *56*(2), 28–30. https://doi.org/10.1145/2408776.2408787

Daradoumis, T., Bassi, R., Xhafa, F., & Caballé, S. (2013). A review on massive elearning (MOOC) design, delivery and assessment. In *2013 Eighth international conference on P2P, parallel, grid, cloud and internet computing* (pp. 208–213). IEEE. https://doi.org/10.1109/3PGCIC.2013.37

de Marcos, L., García-López, E., García-Cabot, A., Medina-Merodio, J. A., Domínguez, A., Martínez-Herráiz, J. J., & Diez-Folledo, T. (2016). Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance. *Computers in Human Behavior*, *60*, 312–321. https://doi.org/10.1016/j.chb.2016.02.052

Dewar, A., Hope, D., Jaap, A., & Cameron, H. (2021). Predicting failure before it happens: A 5-year, 1042 participant prospective study. *Medical Teacher*, *43*(9), 1039–1043. https://doi.org/10.1080/0142159X.2021.1908526

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, *12*(1), 17–26.

Ellis, R. A., Han, F., & Pardo, A. (2017). Improving learning analytics – Combining observational and self-report data on student learning. *Educational Technology & Society*, *20*(3), 158–169.

Fernández-Castilla, B., Aloe, A. M., Declercq, L., Jamshidi, L., Onghena, P., Natasha Beretvas, S., & Van den Noortgate, W. (2018). Concealed correlations meta-analysis: A new method for synthesizing standardized regression coefficients. *Behavior Research Methods*, *51*(1), 316–331. https://doi.org/10.3758/s13428-018-1123-7

Foung, D., & Chen, J. (2019). A learning analytics approach to the evaluation of an online learning package in a Hong Kong University. *Electronic Journal of E-Learning*, *17*(1), 11–24.

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, L., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J., & Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology*, *64*(11), 1187–1197. https://doi.org/10.1016/j.jclinepi.2010.08.010

Galikyan, I., Admiraal, W., & Kester, L. (2021). MOOC discussion forums: The interplay of the cognitive and the social. *Computers & Education*, *165*, 104133. https://doi.org/10.1016/j.compedu.2021.104133

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

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gijlers, H., & de Jong, T. (2013). Using concept maps to facilitate collaborative simulation-based inquiry learning. *The Journal of the Learning Sciences*, *22*(3), 340–374. https://doi.org/10.1080/10508406.2012.748664

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education*, *94*, 241–251. https://doi.org/10.1016/j.compedu.2015.11.002

Grubišić, A., Stankov, S., Žitko, B., Šarić, I., Tomaš, S., Brajković, E., Volarić, T., Vasić, D., & Dodaj, A. (2017). Knowledge tracking variables in intelligent tutoring systems. In *CSEDU 2017 - Proceedings of the 9th International Conference on Computer Supported Education, 1* (pp. 513–518). SciTePress. https://doi.org/10.5220/0006366905130518

Gu, X., & Xu, H. (2019). Missing piece in understanding student learning: Out-of-school computer use. *Journal of Educational Computing Research*, *57*(2), 320–342. https://doi.org/10.1177/0735633118755494

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2*, 107–124. https://doi.org/10.1007/s11409-007-9016-7

Han, F., & Ellis, R. A. (2020). Combining self-reported and observational measures to assess university student academic performance in blended course designs. *Australasian Journal of Educational Technology*, *36*(6), 1–14. https://doi.org/10.14742/ajet.6369

Han, J., Huh, S. Y., Cho, Y. H., Park, S., Choi, J., Suh, B., & Rhee, W. (2020). Utilizing online learning data to design face-to-face activities in a flipped classroom: A case study of heterogeneous group formation. *Educational Technology Research and Development*, *68*(5), 2055–2071. https://doi.org/10.1007/s11423-020-09743-y

Higgins, J., Thompson, S., Deeks, J., & Altman, D. (2002). Statistical heterogeneity in systematic reviews of clinical trials: A critical appraisal of guidelines and practice. *Journal of Health Services Research & Policy*, *7*(1), 51–61. https://doi.org/10.1258/1355819021927674

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, *61*, 133–145.

Huang, X. P., Yu, C. K., & Yang, S. J. H. (2021). Exploring the correlation between students' attention and learning performance. In *Proceedings of the 29th International Conference on Computers in Education* (pp. 515–520). Asia-Pacific Society for Computers in Education.

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. https://doi.org/10.1007/s11423-020-09788-z

Israel, M. J. (2015). Effectiveness of integrating MOOCs in traditional classrooms for undergraduate students. *International Review of Research in Open and Distributed Learning*, *16*(5), 102–118.

Jiang, Y., Clarke-Midura, J., Keller, B., Baker, R. S., Paquette, L., & Ocumpaugh, J. (2018). Note-taking and science inquiry in an open-ended learning environment. *Contemporary Educational Psychology*, *55*, 12–29. https://doi.org/10.1016/j.cedpsych.2018.08.004

Jo, I., Park, Y., & Lee, H. (2017). Three interaction patterns on asynchronous online discussion behaviours: A methodological comparison. *Journal of Computer Assisted Learning*, *33*(2), 106–122. https://doi.org/10.1111/jcal.12168

Jo, I. H., Kim, D., & Yoon, M. (2015). Constructing proxy variables to measure adult learners' time management strategies in LMS. *Educational Technology & Society*, *18*(3), 214–225.

Jo, I. H., & Kim, Y. (2013). Impact of learner's time management strategies on achievement in an e-learning environment: A learning analytics approach. *Journal of Educational Information and Media*, *19*(1), 83–107.

Jokhan, A., Sharma, B., & Singh, S. (2019). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, *44*(11), 1900–1911. https://doi.org/10.1080/03075079.2018.1466872

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, *31*(6), 638–654. https://doi.org/10.1111/jcal.12107

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education*, *87*, 204–217. https://doi.org/10.1016/j.compedu.2015.07.002

Jost, N. S., Jossen, S. L., Rothen, N., & Martarelli, C. S. (2021). The advantage of distributed practice in a blended learning setting. *Education and Information Technologies*, *26*(3), 3097–3113. https://doi.org/10.1007/s10639-020-10424-9

Jovanović, J., Mirriahi, N., Gašević, D., Dawson, S., & Pardo, A. (2019). Predictive power of regularity of pre-class activities in a flipped classroom. *Computers & Education*, *134*, 156–168. https://doi.org/10.1016/j.compedu.2019.02.011

Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, *172*, 104251. https://doi.org/10.1016/j.compedu.2021.104251

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education*, *27*, 54–63. https://doi.org/10.1016/j.iheduc.2015.06.001

Kennedy, G., Coffrin, C., De Barba, P., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. In *Proceedings of the 5th International Conference on Learning Analytics & Knowledge (LAK 2015)* (pp. 136–140). ACM Press. https://doi.org/10.1145/2723576.2723593

Koh, E., Jonathan, C., & Tan, J. P.-L. (2019). Exploring conditions for enhancing critical thinking in networked learning: Findings from a secondary school learning analytics environment. *Education Sciences*, *9*(4), 287. https://doi.org/10.3390/educsci9040287

Larusson, J. A., & White, B. (2014). *Learning analytics: From research to practice*. Springer-Verlag. https://doi.org/10.1007/978-1-4614-3305-7

Lee, H.-S., Pallant, A., Pryputniewicz, S., Lord, T., Mulholland, M., & Liu, O. L. (2019). Automated text scoring and real-time adjustable feedback: Supporting revision of scientific arguments involving uncertainty. *Science Education*, *103*(3), 590–622. https://doi.org/10.1002/sce.21504

Levy, Y., & Ramin, M. M. (2012). A study of online exams procrastination using data analytics techniques. *Interdisciplinary Journal of E-Learning and Learning Objects*, *8*, 97–113.

Li, H., Flanagan, B., Konomi, S., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research and Practice in Technology Enhanced Learning*, *13*(1), 19–12. https://doi.org/10.1186/s41039-018-0087-7

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers & Education*, *127*, 41–65. https://doi.org/10.1016/j.compedu.2018.08.005

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, *45*, 100727. https://doi.org/10.1016/j.iheduc.2020.100727

Li, S., Yu, C., Hu, J., & Zhong, Y. (2017). Exploring the effect of behavioral engagement on learning achievement in online learning environment: Learning analytics of non-degree online learning data. In *Proceedings - 5th International Conference on Educational Innovation through Technology, EITT 2016* (pp. 246–250). IEEE. https://doi.org/10.1109/EITT.2016.56

Lin, C.-C., & Chiu, C.-H. (2013). Correlation between course tracking variables and academic performance in blended online courses. In *Proceedings - 2013 IEEE 13th International Conference on Advanced Learning Technologies, ICALT 2013* (pp. 184–188). IEEE. https://doi.org/10.1109/ICALT.2013.57

Lin, F.-C., Chen, C.-M., & Wang, W.-F. (2017). Learning process analysis based on sequential pattern mining and lag sequential analysis in a web-based inquiry science environment. In *Proceedings – 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017* (pp. 655–660). https://doi.org/10.1109/IIAI-AAI.2017.57

Liu, M.-C., Yu, C.-H., Wu, J., Liu, A.-C., & Chen, H.-M. (2018). Applying learning analytics to deconstruct user engagement by using log data of MOOCs. *Journal of Information Science and Engineering*, *34*(5), 1175–1186. https://doi.org/10.6688/JISE.201809_34(5).0004

Maier, U. (2021). Self-referenced vs. reward-based feedback messages in online courses with formative mastery assessments: A randomized controlled trial in secondary classrooms. *Computers & Education*, *174*, 104306. https://doi.org/10.1016/j.compedu.2021.104306

Mangaroska, K., Vesin, B., Kostakos, V., Brusilovsky, P., & Giannakos, M. N. (2021). Architecting analytics across multiple e-learning systems to enhance learning design. *IEEE Transactions on Learning Technologies*, *14*(2), 173–188. https://doi.org/10.1109/TLT.2021.3072159

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, *56*, 243–252.

Miller, L. D., & Soh, L.-K. (2013). Significant predictors of learning from student interactions with online learning objects. In *Proceedings - 2013 Frontiers in Education Conference, FIE* (pp. 203–209). IEEE. https://doi.org/10.1109/FIE.2013.6684817

Mills, N. (2021). ALEKS constructs as predictors of high school mathematics achievement for struggling students. *Heliyon*, *7*(6), e07345. https://doi.org/10.1016/j.heliyon.2021.e07345

Molenaar, I., Horvers, A., & Baker, R. S. (2021). What can moment-by-moment learning curves tell about students' self-regulated learning? *Learning and Instruction*, *72*, 101206. https://doi.org/10.1016/j.learninstruc.2019.05.003

Musabirov, I., Pozdniakov, S., & Tenisheva, K. (2019). Predictors of academic achievement in blended learning: The case of data science minor. *International Journal of Emerging Technologies in Learning*, *14*(5), 64–74. https://doi.org/10.3991/ijet.v14i05.9512

Mwalumbwe, I., & Mtebe, J. S. (2017). Using learning analytics to predict students' performance in moodle learning management system: A case of mbeya university of science and technology. *The Electronic Journal of Information Systems in Developing Countries*, *79*(1), 1–13. https://doi.org/10.1002/j.1681-4835.2017.tb00577.x

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, *11*, 237. https://doi.org/10.3390/app11010237

Naumann, J., & Salmerón, L. (2016). Does navigation always predict performance? Effects of navigation on digital reading are moderated by comprehension skills. *International Review of Research in Open and Distance Learning*, *17*(1), 42–59. https://doi.org/10.19173/irrodl.v17i1.2113

Ober, T. M., Hong, M. R., Rebouças-Ju, D. A., Carter, M. F., Liu, C., & Cheng, Y. (2021). Linking self-report and process data to performance as measured by different assessment types. *Computers & Education*, *167*, 104188. https://doi.org/10.1016/j.compedu.2021.104188

Page, M. J., JE, M. K., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*, n71. https://doi.org/10.1136/bmj.n71

Papamitsiou, Z., & Economides, A. A. (2021). The impact of on-demand metacognitive help on effortful behaviour: A longitudinal study using task-related visual analytics. *Journal of Computer Assisted Learning*, *37*(1), 109–126. https://doi.org/10.1111/jcal.12472

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82–92. https://doi.org/10.1109/TLT.2016.2639508

Pardo, A., Zhao, Y., Mirriahi, N., Zhao, A., Dawson, S., & Gašević, D. (2015). Identifying learning strategies associated with active use of video annotation software. In *Proceedings of the 5th International Conference on Learning Analytics & Knowledge (LAK 2015)* (pp. 255–259). ACM Press. https://doi.org/10.1145/2723576.2723611

Pei, X., Jin, Y., Zheng, T., & Zhao, J. (2020). Longitudinal effect of a technology-enhanced learning environment on sixth-grade students' science learning: The role of reflection. *International Journal of Science Education*, *42*(2), 271–289. https://doi.org/10.1080/09500693.2019.1710000

Ramirez-Arellano, A., Bory-Reyes, J., & Hernández-Simón, L. M. (2019). Emotions, motivation, cognitive–metacognitive strategies, and behavior as predictors of learning performance in blended learning. *Journal of Educational Computing Research*, *57*(2), 491–512. https://doi.org/10.1177/0735633117753935

Ritter, S., Joshi, A., Fancsali, S. E., & Nixon, T. (2013). Predicting standardized test scores from cognitive tutor interactions. In *Proceedings of the 6th International Conference on Educational Data Mining, EDM 2013*. International Educational Data Mining Society (IEDMS).

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Romero-Zaldivar, V.-A., Pardo, A., Burgos, D., & Delgado Kloos, C. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education*, *58*(4), 1058–1067. https://doi.org/10.1016/j.compedu.2011.12.003

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Delgado Kloos, C. (2018). Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators. *Expert Systems*, *35*(6), e12298. https://doi.org/10.1111/exsy.12298

Saqr, M., Fors, U., & Nouri, J. (2018). Using social network analysis to understand online problem-based learning and predict performance. *PloS One*, *13*(9), e0203590. https://doi.org/10.1371/journal.pone.0203590

Saqr, M., Nouri, J., Vartiainen, H., & Malmberg, J. (2020). What makes an online problem-based group successful? A learning analytics study using social network analysis. *BMC Medical Education*, *20*(1), 80. https://doi.org/10.1186/s12909-020-01997-7

Scheffel, M., Drachsler, H., de Kraker, J., Kreijns, K., Slootmaker, A., & Specht, M. (2017). Widget, widget on the wall, Am I performing well at all? *IEEE Transactions on Learning Technologies*, *10*(1), 42–52. https://doi.org/10.1109/TLT.2016.2622268

Schumacher, C., & Ifenthaler, D. (2021). Investigating prompts for supporting students' self-regulation – A remaining challenge for learning analytics approaches? *The Internet and Higher Education*, *49*, 100791. https://doi.org/10.1016/j.iheduc.2020.100791

Sharma, B., Nand, R., Naseem, M., & Reddy, E. V. (2020). Effectiveness of online presence in a blended higher learning environment in the Pacific. *Studies in Higher Education*, *45*(8), 1547–1565. https://doi.org/10.1080/03075079.2019.1602756

Smith, D. H., Hao, Q., Hundhausen, C. D., Jagodzinski, F., Myers-Dean, J., & Jaeger, K. (2021). Towards modeling student engagement with interactive computing textbooks: An empirical study. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 914–920). ACM Press.

Soffer, T., & Cohen, A. (2019). Students' engagement characteristics predict success and completion of online courses. *Journal of Computer Assisted Learning*, *35*(3), 378–389. https://doi.org/10.1111/jcal.12340

Stadler, M., Hofer, S., & Greiff, S. (2020). First among equals: Log data indicates ability differences despite equal scores. *Computers in Human Behavior*, *111*, 106442. https://doi.org/10.1016/j.chb.2020.106442

Strang, K. D. (2016). Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes? *Education and Information Technologies*, *22*(3), 917–937. https://doi.org/10.1007/s10639-016-9464-2

Summers, R. J., Higson, H. E., & Moores, E. (2020). Measures of engagement in the first three weeks of higher education predict subsequent activity and attainment in first year undergraduate students: A UK case study. *Assessment and Evaluation in Higher Education*, *46*(5), 821–836. https://doi.org/10.1080/02602938.2020.1822282

Svihla, V., Wester, M. J., & Linn, M. C. (2015). Distributed revisiting: An analytic for retention of coherent science learning. *Journal of Learning Analytics*, *2*(2), 75–101. https://doi.org/10.18608/jla.2015.22.7

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers & Education: Artificial Intelligence*, *3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Tacoma, S., Drijvers, P., & Jeuring, J. (2021). Combined inner and outer loop feedback in an intelligent tutoring system for statistics in higher education. *Journal of Computer Assisted Learning*, *37*(2), 319–332. https://doi.org/10.1111/jcal.12491

Tacoma, S., Geurts, C., Slof, B., Jeuring, J., & Drijvers, P. (2020). Enhancing learning with inspectable student models: Worth the effort? *Computers in Human Behavior*, *107*, 106276. https://doi.org/10.1016/j.chb.2020.106276

Tan, X., She, J., Chen, S., Ohno, S., & Kameda, H. (2021). Analysis of student learning behavior based on moodle log data. In *International Conference on Human System Interaction, HSI* (Vol. 2021, pp. 1–4). IEEE. https://doi.org/10.1109/HSI52170.2021.9538680

Tan, Y., Zhang, X., Luo, H., Sun, Y., & Xu, S. (2018). Learning profiles, behaviors and outcomes: Investigating international students' learning experience in an English MOOC. In *Proceedings - 2018 International Symposium on Educational Technology, ISET 2018* (pp. 214–218). IEEE. https://doi.org/10.1109/ISET.2018.00055

Tempelaar, D., Rienties, B., & Nguyen, Q. (2020). Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. *PLOS ONE*, *15*(6), e0233977. https://doi.org/10.1371/journal.pone.0233977

Theobald, M., Bellhäuser, H., & Imhof, M. (2018). Identifying individual differences using log-file analysis: Distributed learning as mediator between conscientiousness and exam grades. *Learning and Individual Differences*, *65*, 112–122. https://doi.org/10.1016/j.lindif.2018.05.019

Tian, H., Lai, S., & Wu, F. (2019). Does time play a role? Prediction of learning performance with time-use habits in online assignments. *Proceedings - International Joint Conference on Information, Media, and Engineering, IJCIME*, *2019*, 473–477. https://doi.org/10.1109/IJCIME49369.2019.00101

Ulfa, S., & Fatawi, I. (2021). Predicting factors that influence students' learning outcomes using learning analytics in online learning environment. *International Journal of Emerging Technologies in Learning*, *16*(1), 4–17. https://doi.org/10.3991/ijet.v16i01.16325

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wei, H., Peng, H., & Chou, C. (2015). Can more interactivity improve learning achievement in an online course? Effects of college students' perception and actual use of a course-management system on their learning achievement. *Computers & Education*, *83*, 10–21. https://doi.org/10.1016/j.compedu.2014.12.013

Widyahastuti, F., Riady, Y., & Fransiskus, D. (2017). Performance prediction as a new feature in e-learning. *The 8th International Conference on e-Learning*, 237–243.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5–13. https://doi.org/10.18608/jla.2015.22.2

Wu, Z., Zhao, B., & Wang, Y. (2021). Analysis of students' learning behavior under network learning environment. In *2021 IEEE 3rd International Conference on Computer Science and Educational Informatization, CSEI 2021* (pp. 46–50). IEEE. https://doi.org/10.1109/CSEI51395.2021.9477755

Xu, Z., Yuan, H., & Liu, Q. (2021). Student performance prediction based on blended learning. *IEEE Transactions on Education*, *64*(1), 66–73. https://doi.org/10.1109/TE.2020.3008751

Yamada, M., Okubo, F., Oi, M., Shimada, A., Kojima, K., & Ogata, H. (2016). Learning analytics in ubiquitous learning environments: Self-regulated learning perspective. In *ICCE 2016 - 24th International Conference on Computers in Education: Think Global Act Local - Main Conference Proceedings* (pp. 306–314). Asia-Pacific Society for Computers in Education.

Yang, D., Zargar, E., Adams, A. M., Day, S. L., & Connor, C. M. (2021). Using interactive e-book user log variables to track reading processes and predict digital learning outcomes. *Assessment for Effective Intervention*, *46*(4), 292–303. https://doi.org/10.1177/1534508420941935

Yoo, J., & Kim, J. (2014). Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education*, *24*(1), 8–32. https://doi.org/10.1007/s40593-013-0010-8

You, J. W. (2015). Examining the effect of academic procrastination on achievement using LMS data in E-learning. *Educational Technology & Society*, *18*(3), 64–74.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, *29*, 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003

Yu, T., & Jo, I. H. (2014). Educational technology approach toward learning analytics. In *Proceedings of the 4th International Conference on Learning Analytics & Knowledge (LAK 2014)* (pp. 269–270). ACM Press. https://doi.org/10.1145/2567574.2567594
Zacharis, N. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, *27*, 44–53. https://doi.org/10.1016/j.iheduc.2015.05.002
Zarrabi, F., & Bozorgian, H. (2020). EFL students' cognitive performance during argumentative essay writing: A log-file data analysis. *Computers and Composition*, *55*, 102546. https://doi.org/10.1016/j.compcom.2020.102546

How to cite this article: Wang, Q., & Mousavi, A. (2023). Which log variables significantly predict academic achievement? A systematic review and meta-analysis. *British Journal of Educational Technology*, *54*, 142–191. https://doi.org/10.1111/bjet.13282

APPENDIX A: SAMPLE SIZE AND CONTEXT INFORMATION IN THE REVIEWED STUDIES

Authors	*N*	Learning type	Learning theme
Bravo-Agapito et al. (2021)	802	Fully online	ICT and Sociology
Dewar et al. (2021)	1042	Not specified	Medicine
Galikyan et al. (2021)	633	Fully online	Linguistics
Huang et al. (2021)	62	Fully online	CS
Jost et al. (2021)	62	Blended	Psychology
Jovanović et al. (2021)	344	Blended	Medicine
Maier (2021)	620	Fully online	K-12
Mangaroska et al. (2021)	153	Blended	CS
Mills (2021)	265	Blended	K-12
Ober et al. (2021)	329	Blended	K-12
Schumacher and Ifenthaler (2021)	101	Fully online	Business
Smith et al. (2021)	80	Blended	CS
Tan et al. (2021)	172	Fully online	Language
Ulfa and Fatawi (2021)	53	Blended	CS
Wu et al. (2021)	Not specified	Fully online	CS
Xu et al. (2021)[1]	55	Blended	CS
Xu et al. (2021)[2]	72	Blended	CS
Yang et al. (2021)[1]	565	Fully online	K-12
Yang et al. (2021)[2]	426	Fully online	K-12
Chan et al. (2021)	98	Blended	Medicine
Han and Ellis (2020)	335	Blended	CS
Han et al. (2020)	53	Blended	Education
Li et al. (2020)[1]	238	Fully online	Chemistry
Li et al. (2020)[2]	238	Fully online	Chemistry
Papamitsiou and Economides (2021)	67	Blended	Business

Authors	N	Learning type	Learning theme
Pei et al. (2020)	125	Blended	K-12
Saqr et al. (2020)	598	Blended	Medicine
Sharma et al. (2020)	873	Blended	ICT
Stadler et al. (2020)	329	Fully online	K-12
Summers et al. (2020)	1602	Not specified	Multiple courses
Tacoma et al. (2021)	521	Blended	Statistics
Tacoma, et al. (2020)	599	Blended	Statistics
Tempelaar et al. (2020)[1]	14	Blended	Mathematics
Tempelaar et al. (2020)[2]	14	Blended	Statistics
Zarrabi and Bozorgian (2020)	72	Fully online	Language
Chen et al. (2019)	70	Blended	CS
Foung and Chen (2019)	7156	Blended	English
Gu and Xu (2019)	546	Blended	K-12
Jokhan et al. (2019)	1403	Blended	ICT
Jovanović et al. (2019)	486	Blended	Engineering
Koh et al. (2019)	263	Blended	K-12
Lee et al. (2019)	343	Fully online	K-12
Musabirov et al. (2019)	189	Blended	CS
Ramirez-Arellano et al. (2019)	137	Blended	Life sciences and Chemistry
Soffer and Cohen (2019)	646	Fully online	Humanities and Medicine
Tian et al. (2019)	67	Fully online	K-12
Chiu and Hew (2018)	1563	Fully online	Literature
Conijn et al. (2018)	199	Blended	CS
Jiang et al. (2018)	1985	Fully online	K-12
Li and Baker (2018)	71,457	Fully online	Mathematics
Li et al. (2018)	2454	Fully online	Language
Liu et al. (2018)	2697	Fully online	Multiple courses
Ruipérez-Valiente et al. (2018)	69	Fully online	Chemistry and Physics
Saqr et al. (2018)	215	Blended	Medicine
Tan et al. (2018)	87	Fully online	Language
Theobald et al. (2018)	424	Fully online	Education
Conijn et al. (2017)[1]	889	Blended	Mathematics
Conijn et al. (2017)[2]	1164	Blended	Mathematics
Conijn et al. (2017)[3]	742	Blended	Mathematics
Conijn et al. (2017)[4]	815	Blended	Mathematics
Conijn et al. (2017)[5]	587	Blended	Mathematics
Conijn et al. (2017)[6]	673	Blended	Mathematics
Conijn et al. (2017)[7]	279	Blended	Mathematics
Conijn et al. (2017)[8]	302	Blended	Physics
Conijn et al. (2017)[9]	620	Blended	Psychology
Conijn et al. (2017)[10]	234	Blended	Physics

(Continues)

Authors	N	Learning type	Learning theme
Conijn et al. (2017)[11]	227	Blended	Physics
Conijn et al. (2017)[12]	189	Blended	Physics
Conijn et al. (2017)[13]	189	Blended	Psychology
Conijn et al. (2017)[14]	61	Blended	Mathematics
Conijn et al. (2017)[15]	164	Blended	Mathematics
Conijn et al. (2017)[16]	198	Blended	Mathematics
Ellis et al. (2017)	291	Blended	Engineering
Grubišić et al. (2017)	156	Fully online	CS
Jo et al. (2017)	43	Blended	Public administration
Li et al. (2017)	2582	Blended	CS
Lin et al. (2017)	48	Fully online	K-12
Mwalumbwe and Mtebe (2017)[1]	111	Blended	Life sciences and Chemistry
Mwalumbwe and Mtebe (2017)[2]	60	Blended	Engineering
Pardo et al. (2017)	145	Blended	Engineering
Scheffel et al. (2017)	134	Fully online	Sustainable Development
Widyahastuti et al. (2017)	95	Fully online	Language
Choi et al. (2016)	21,171	Fully online	Multiple courses
de Marcos et al. (2016)	161	Blended	ICT
Goggins and Xing (2016)	24	Fully online	Education
Naumann and Salmerón (2016)	533	Fully online	K-12
Strang (2016)	228	Fully online	Business
Yamada et al. (2016)	93	Blended	CS
You (2016)	530	Fully online	Arts
Joksimović, Gašević, Kovanović, et al. (2015)	44	Fully online	CS
Joksimović, Gašević, Loughin, et al. (2015)	204	Fully online	CS
Jo et al. (2015)	200	Fully online	Business
Junco and Clem (2015)	233	Not specified	Multiple courses
Kennedy et al. (2015)	6635	Fully online	CS
Pardo et al. (2015)	149	Blended	Not specified
Svihla et al. (2015)	835	Fully online	K-12
Wei et al. (2015)	381	Fully online	Education
You (2015)	569	Fully online	Arts
Zacharis (2015)	134	Blended	CS
Agudo-Peregrina et al. (2014)	138	Fully online	ICT and business
Yoo and Kim (2014)	370	Fully online	CS
Yu and Jo (2014)	84	Blended	Public administration
Gijlers and de Jong (2013)	50	Fully online	K-12
Lin and Chiu (2013)	528	Blended	Multiple courses
Miller and Soh (2013)	134	Fully online	CS

Authors	N	Learning type	Learning theme
Ritter et al. (2013)	3224	Blended	K-12
Bernacki et al. (2012)	160	Fully online	Education
Romero-Zaldivar et al. (2012)	172	Blended	Engineering

APPENDIX B: THE INFORMATION OF THE DEPENDENT VARIABLE, LOG VARIABLES, AND REGRESSION COEFFICIENTS

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Bravo-Agapito et al. (2021)	Total score	Task factor	Generic	Elaborated	0.418
		Access factor	Generic	Elaborated	0.209
		Questionnaire factor	Course-specific	Elaborated	0.217
Dewar et al. (2021)	Exam score	Engagement score	Course-specific	Elaborated	0.280
Galikyan et al. (2021)	Total score	Number of different threads a learner contributed to	Course-specific	Basic	6.700
Huang et al. (2021)[1]	Exam score	Page synchronization ratio	Course-specific	Elaborated	0.320
Huang et al. (2021)[2]	Exam score	Page synchronization ratio	Course-specific	Elaborated	0.398
Jost et al. (2021)	Exam score	On-peak time	Generic	Basic	0.350
		Studying regularity	Generic	Elaborated	−0.310
Jovanović et al. (2021)	Total score	Total session length	Generic	Basic	4.077
		Entropy of daily counts of the forum_contribute learning actions	Generic	Elaborated	2.322
		Entropy of weekly counts of the lecture_viewed active days	Generic	Elaborated	2.777
		Entropy of weekly counts of the course_main_viewed active days	Generic	Elaborated	−2.703
Maier (2021)	Exam score	Time learners spent reading the rules and examples	Course-specific	Basic	0.010
		Number of clicks between rules and examples	Course-specific	Basic	−0.660
		Ratio between the number of failed trainings divided by the number of completed trainings	Course-specific	Elaborated	−0.410

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Mangaroska et al. (2021)	Total score	Level of complexity of a chosen coding exercise	Course-specific	Basic	0.650
		Time students spend navigating in mastery grids to monitor and reflect on their progress	Course-specific	Basic	0.420
		Number of submitted assignments	Generic	Basic	0.990
		Number of incorrect submissions	Generic	Basic	−0.320
		Number of incomplete assignments	Generic	Basic	−0.170
Mills (2021)	Total score	Combined (active and inactive) time logged in the program	Generic	Basic	0.110
		Comparison between participants mastered and learned topics	Course-specific	Elaborated	−6.670
		Quotient of students' total topics mastered to total topics practiced	Course-specific	Elaborated	44.710
Ober et al. (2021)	Total score	Number of assignments completed	Generic	Basic	0.760
		Number of clicks to the results page	Course-specific	Basic	0.860
Schumacher and Ifenthaler (2021)	Exam score	Number of views of handout	Generic	Basic	0.218
Smith et al. (2021)	Exam score	Interaction	Generic	Elaborated	0.540
Tan et al. (2021)	Total score	Times of late submission	Generic	Elaborated	7.358
		Times of on-time submission	Generic	Elaborated	7.253
		Completeness of course material	Generic	Elaborated	2.365
Ulfa and Fatawi (2021)	Exam score	Number of working on exercises	Course-specific	Basic	0.303
Wu et al. (2021)	Total score	Total number of tasks and videos watched	Course-specific	Basic	0.845
		Sum of learning chapters	Course-specific	Basic	0.265

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Xu et al. (2021)[1]	Total score	Study time	Generic	Basic	0.359
		Used time for MCQ	Course-specific	Basic	0.291
		Submission delay	Generic	Elaborated	0.419
Xu et al. (2021)[2]	Total score	Number of posts	Course-specific	Basic	0.082
Yang et al. (2021)[1]	Exam score	Percentage of questions answered correctly	Course-specific	Elaborated	0.050
		Number of attempts to answer questions	Course-specific	Basic	−0.570
Yang et al. (2021)[2]	Exam score	Percentage of questions answered correctly	Course-specific	Elaborated	0.050
		Number of implausible decisions students made while reading the wke-Book	Course-specific	Basic	−0.350
		Number of attempts to answer questions	Course-specific	Basic	−0.870
Chan et al. (2021)	Exam score	Quiz access	Course-specific	Basic	−0.349
Han and Ellis (2020)	Total score	Frequency of access to the learning resources	Generic	Basic	0.220
		Frequency of access to the study kit	Course-specific	Basic	0.130
		Frequency of correctly answered exercise sequences	Course-specific	Basic	0.610
		Frequency of incorrectly answered exercise sequences	Course-specific	Basic	−0.550
Han et al. (2020)	Exam score	Engagement heterogeneity	Course-specific	Elaborated	0.783
Li et al. (2020)[1]	Total score	Proportion of units accessed before the deadline	Generic	Elaborated	0.160
		Students studying a unit in advance of the deadlines	Generic	Elaborated	0.143
		Slope of a simple linear regression that regressed time on task in a given module on the module number (1)	Course-specific	Elaborated	0.245

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
		Slope of a simple linear regression that regressed time on task in a given module on the module number (2)	Course-specific	Elaborated	−0.115
Li et al. (2020)[2]	Exam score	Students studying a unit in advance of the deadlines	Generic	Elaborated	0.155
		Slope of a simple linear regression that regressed time on task in a given module on the module number (1)	Course-specific	Elaborated	0.206
		Slope of a simple linear regression that regressed time on task in a given module on the module number (2)	Course-specific	Elaborated	−0.080
Papamitsiou and Economides (2021)	Exam score	Time to answer correctly	Course-specific	Basic	0.023
		Time to answer wrongly	Course-specific	Basic	−0.025
Pei et al. (2020)	Exam score	Time spent in the reflecting steps	Course-specific	Basic	0.270
Saqr et al. (2020)	Total score	Tutor eigencentrality	Generic	Elaborated	−0.398
		User count	Generic	Basic	−0.354
		Centralization outdegree	Generic	Elaborated	−0.255
		Reciprocity	Generic	Elaborated	0.204
Sharma et al. (2020)	Total score	Total frequency	Generic	Basic	0.287
		Total duration	Generic	Basic	0.591
Stadler et al. (2020)	Exam score	Time-on-task	Generic	Basic	0.090
		Number of interactions	Generic	Basic	−0.090
Summers et al. (2020)	Total score	Number of lectures that the student attended	Generic	Basic	0.287
		Number of times the student accessed course materials	Generic	Basic	0.278
		Number of times the student viewed recorded lectures	Generic	Basic	0.089

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Tacoma et al. (2021)	Exam score	Time-on-task	Generic	Basic	0.190
Tacoma et al. (2020)	Exam score	Total time on task	Generic	Basic	0.280
		Student viewed student models for four or five homework sets.	Course-specific	Elaborated	4.000
		Student viewed student models of at most four homework sets.	Course-specific	Elaborated	1.900
		Homework-Practice-Other topic, students who made all three decisions at least once	Course-specific	Elaborated	−1.000
		Practice-Other topic, students who made the decisions Practice and Other topic at least once and never made the decision Homework	Course-specific	Elaborated	−1.320
		Other, students who always worked on another topic after viewing the student model	Course-specific	Elaborated	−0.950
Tempelaar et al. (2020)[1]	Exam score	Total number of clicks	Generic	Basic	0.080
		Proportion of exercises correctly solved	Course-specific	Elaborated	0.637
		Number of attempts to solve an exercise	Course-specific	Basic	−0.551
		Total number of hints asked for	Course-specific	Basic	−0.067
Tempelaar et al. (2020)[2]	Exam score	Total number of clicks	Generic	Basic	0.061
		Proportion of exercises correctly solved	Course-specific	Elaborated	0.864
		Number of attempts to solve an exercise	Course-specific	Basic	−0.522
		Time-on-task	Generic	Basic	−0.127
Zarrabi and Bozorgian (2020)	Exam score	Revision behavior	Course-specific	Basic	0.320
		Time-on-task	Generic	Basic	0.750
		Pause behavior	Course-specific	Basic	0.870
		Pausing strategy	Course-specific	Basic	0.680

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Chen et al. (2019)	Total score	Frequency of turning to the previous page	Course-specific	Basic	0.257
		Frequency of jumping to a bookmark	Course-specific	Basic	0.224
		Frequency of deleting markers	Course-specific	Basic	0.271
Foung and Chen (2019)	Total score	Total number of attempts at indiwork	Generic	Basic	0.139
		Days after the term starts	Generic	Basic	–0.078
Gu and Xu (2019)	Exam score	Number of diaries posted by each student	Course-specific	Basic	0.392
		Number of topics that each student initiated for discussion	Course-specific	Basic	0.284
		Number of stories posted by each student	Course-specific	Basic	0.159
Jokhan et al. (2019)	Total score	Weekly average logins	Generic	Basic	0.360
		Weekly average completion rates	Course-specific	Basic	0.524
Jovanović et al. (2019)	Exam score	Entropy of weekday session counts	Generic	Elaborated	2.580
		Entropy of weekly use of summative exercises	Course-specific	Elaborated	2.830
		Entropy of weekly use of course videos for the pre-class activities	Course-specific	Elaborated	3.300
		Entropy of weekly access to the course e-book	Course-specific	Elaborated	3.870
		Entropy of requests to see a solution for a formative MCQ	Course-specific	Elaborated	−2.380
		Frequency of change in the ‘pattern’ of learning resource use during pre-class activities	Course-specific	Basic	−2.770
		Number of weeks of high engagement with summative exercises	Course-specific	Basic	−7.270
Koh et al. (2019)	Exam score	Word count of posts	Course-specific	Basic	0.208
		Out-degree centrality	Generic	Elaborated	−0.202
		Out 2-step reach centrality	Generic	Elaborated	0.170
		Class arc reciprocity	Generic	Elaborated	0.335

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Lee et al. (2019)	Exam score	Number of argumentation tasks where students used hasbot to revise	Course-specific	Basic	0.260
Musabirov et al. (2019)	Total score	Number of lines of code written	Course-specific	Basic	−2.870
		Percent of additional assignment accomplished	Course-specific	Elaborated	12.660
		Percent of wrong tasks' submission	Course-specific	Elaborated	−29.610
Ramirez-Arellano et al. (2019)	Total score	Number of missed learning activities	Generic	Basic	−0.190
Soffer and Cohen (2019)	Exam score	Number of times the student entered the forums	Course-specific	Basic	0.100
		Number of times the student entered the course's homepage	Generic	Basic	0.220
		Number of times the student entered a page in the learning units, divided by the total number of pages	Generic	Elaborated	0.220
		Number of students' hits in the course website pages	Generic	Basic	−0.530
		Number of assignments the student submitted, divided by the number of assignments in the course	Generic	Elaborated	0.300
Tian et al. (2019)	Exam score	Number of assignments submission	Generic	Basic	23.436
		Extent to which the student postpones the assignment until the deadline	Generic	Elaborated	−15.246
		How quickly students complete homework	Course-specific	Elaborated	22.297
Chiu and Hew (2018)	Total score	Total views	Course-specific	Basic	0.065
		Total comments	Course-specific	Basic	0.062

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Conijn et al. (2018)	Exam score	Videos started	Course-specific	Basic	−0.430
		Quizzes finished	Course-specific	Basic	0.390
		Resources read	Course-specific	Basic	−0.190
		Unique resources read	Course-specific	Basic	0.320
		Peer assignments started	Course-specific	Basic	0.080
		Peer assignment finished	Course-specific	Basic	−0.670
Jiang et al. (2018)	Exam score	Frequency of opening the notepad window	Course-specific	Basic	0.230
		Total amount of time in minutes that notepad was open	Course-specific	Basic	0.160
		Number of words in note-taker's note	Course-specific	Basic	0.150
		Number of sentence segments in note-taker's note	Course-specific	Basic	0.210
		Frequency of note-taking actions	Course-specific	Basic	0.200
		Frequency of note-reaccessing actions	Course-specific	Basic	0.200
		Total amount of time (in minutes) spent on taking notes	Course-specific	Basic	0.140
		Total amount of time (in minutes) spent on note-reaccessing episodes	Course-specific	Basic	0.120
		Average duration (in minutes) of a note-taking action	Course-specific	Basic	−0.120
		Average duration (in minutes) of a note-reaccessing action	Course-specific	Basic	0.070
		Ratio of time spent on note-taking actions and total time on notepad	Course-specific	Elaborated	−0.120
		Ratio of time spent on note-reaccessing actions and total time on notepad	Course-specific	Elaborated	0.120

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Li and Baker (2018)	Total score	Quiz coverage	Course-specific	Elaborated	0.389
		Lecture coverage	Course-specific	Elaborated	0.546
Li et al. (2018)	Exam score	Number of completed quizzes	Course-specific	Basic	0.230
		Study irregularity	Generic	Elaborated	−0.158
		Total access time	Generic	Basic	0.104
		Pacing	Course-specific	Basic	0.116
Liu et al. (2018)	Total score	Stop video	Course-specific	Basic	0.264
		Pause video	Course-specific	Basic	0.143
		Seek video	Course-specific	Basic	0.316
Ruipérez-Valiente et al. (2018)	Exam score	The average number of attempts	Course-specific	Basic	0.187
		Percentage of correct exercises without use of hints and answering correctly at the first attempt	Course-specific	Elaborated	0.324
		The average number of minutes spent by the student each day	Generic	Basic	0.342
		Percentage of exercises and videos that were started but the student never completed or achieved proficiency in them	Course-specific	Elaborated	0.155
		Students' failure of correctly solving exercises, attempting too many times, or asking too many hints	Course-specific	Elaborated	−0.264
Saqr et al. (2018)	Total score	Tutor out-degree	Generic	Elaborated	−0.140
		Eigencentrality	Generic	Elaborated	0.130
		Density	Generic	Elaborated	0.220
		AV group clustering	Generic	Elaborated	0.680
Tan et al. (2018)	Exam score	Number of PDF file downloads or clicks	Generic	Basic	0.235
		Number of the learners view course page	Generic	Basic	0.275
		Number of the learners create posts	Course-specific	Basic	0.257

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Theobald et al. (2018)	Exam score	Number of weeks in which each student had accessed the LMS irrespective of the actual amount of time students spent online	Generic	Elaborated	−0.245
Conijn et al. (2017)[1]	Exam score	Total number of clicks	Generic	Basic	−0.200
		Number of course page views	Generic	Basic	0.310
		Irregularity of study interval	Generic	Elaborated	−0.370
		Largest period of inactivity	Generic	Basic	0.160
Conijn et al. (2017)[2]	Exam score	Total number of clicks	Generic	Basic	−0.390
		Number of online sessions	Generic	Basic	0.240
		Total time online	Generic	Basic	−0.190
		Number of course page views	Generic	Basic	0.330
		Irregularity of study interval	Generic	Elaborated	−0.320
		Largest period of inactivity	Generic	Basic	0.150
		Time until first activity	Generic	Basic	−0.100
		Average time per session	Generic	Basic	0.220
Conijn et al. (2017)[3]	Exam score	Total number of clicks	Generic	Basic	−0.360
		Number of course page views	Generic	Basic	0.350
		Largest period of inactivity	Generic	Basic	−0.250
Conijn et al. (2017)[4]	Exam score	Irregularity of study interval	Generic	Elaborated	−0.190
		Time until first activity	Generic	Basic	−0.270
Conijn et al. (2017)[5]	Exam score	Number of course page views	Generic	Basic	0.630
Conijn et al. (2017)[6]	Exam score	Total time online	Generic	Basic	0.250
		Irregularity of study time	Generic	Elaborated	0.260
Conijn et al. (2017)[7]	Exam score	Total number of clicks	Generic	Basic	−0.300
		Irregularity of study time	Generic	Elaborated	−0.570
		Largest period of inactivity	Generic	Basic	−0.560
		Time until first activity	Generic	Basic	−0.580
		Average time per session	Generic	Basic	0.350
Conijn et al. (2017)[8]	Exam score	Total number of clicks	Generic	Basic	1.300
		Number of course page views	Generic	Basic	−1.380
		Irregularity of study interval	Generic	Elaborated	−0.520
		Largest period of inactivity	Generic	Basic	0.560
		Time until first activity	Generic	Basic	−0.270

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Conijn et al. (2017)[9]	Exam score	Number of online sessions	Generic	Basic	0.550
		Total time online	Generic	Basic	0.340
		Number of course page views	Generic	Basic	−0.420
Conijn et al. (2017)[10]	Exam score	Total number of clicks	Generic	Basic	0.140
		Number of online sessions	Generic	Basic	0.440
		Total time online	Generic	Basic	−0.340
		Number of course page views	Generic	Basic	−2.280
		Irregularity of study time	Generic	Elaborated	0.170
		Irregularity of study interval	Generic	Elaborated	−0.220
		Largest period of inactivity	Generic	Basic	0.320
Conijn et al. (2017)[11]	Exam score	Total number of clicks	Generic	Basic	0.210
		Number of online sessions	Generic	Basic	0.240
		Total time online	Generic	Basic	−0.180
		Number of course page views	Generic	Basic	−0.520
		Largest period of inactivity	Generic	Basic	0.190
		Average time per session	Generic	Basic	0.130
Conijn et al. (2017)[12]	Exam score	Total number of clicks	Generic	Basic	1.910
		Total time online	Generic	Basic	0.760
Conijn et al. (2017)[13]	Exam score	Number of online sessions	Generic	Basic	0.600
		Irregularity of study time	Generic	Elaborated	−0.400
		Largest period of inactivity	Generic	Basic	0.330
Conijn et al. (2017)[14]	Exam score	Number of online sessions	Generic	Basic	0.460
		Total time online	Generic	Basic	−0.230
Conijn et al. (2017)[15]	Exam score	Irregularity of study interval	Generic	Elaborated	−0.530
		Largest period of inactivity	Generic	Basic	0.500
		Time until first activity	Generic	Basic	−0.120
		Average time per session	Generic	Basic	−0.180
Conijn et al. (2017)[16]	Exam score	Largest period of inactivity	Generic	Basic	0.720
Ellis et al. (2017)	Total score	Number of student views of any page of the course notes	Generic	Basic	0.880
		Number of multiple-choice questions answered	Course-specific	Basic	0.290
Grubišić et al. (2017)	Exam score	Total number of concepts	Course-specific	Basic	0.323
		Total score gained on tutor	Course-specific	Basic	0.410

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Jo et al. (2017)	Total score	Irregularity of access interval	Generic	Elaborated	−0.270
		In-degree centrality	Generic	Elaborated	0.330
		Out-degree centrality	Generic	Elaborated	0.570
Li et al. (2017)	Exam score	Number of lesson quizzes taken	Course-specific	Basic	1.112
		Number of completed lessons	Course-specific	Basic	−0.364
		Logon numbers	Generic	Basic	−0.044
		Rate of announcements read	Course-specific	Elaborated	0.029
		Average time spent on the platform of each logon	Generic	Basic	0.030
		Number of completed units	Course-specific	Basic	0.051
		Number of videos watched incompletely	Course-specific	Basic	0.037
Lin et al. (2017)	Exam score	Total learning time	Generic	Basic	0.297
		Inquiry simulation experiment and activity learning time	Course-specific	Basic	0.397
Mwalumbwe and Mtebe (2017)[1]	Exam score	Number of peer interactions	Generic	Basic	0.196
		Number of forum posts	Course-specific	Basic	0.771
Mwalumbwe and Mtebe (2017)[2]	Exam score	Number of forum posts	Course-specific	Basic	0.194
		Number of exercises performed	Course-specific	Basic	0.544
Pardo et al. (2017)	Total score	Frequency of resource view	Generic	Basic	0.850
		Frequency of students interact with the multiple-choice questions (question is answered correctly, question is answered incorrectly, and the student requests to see the answers)	Course-specific	Basic	0.310
Scheffel et al. (2017)	Total score	Number of posts	Course-specific	Basic	0.185
		Number of comments to posts	Course-specific	Basic	0.521
		Number of page views	Generic	Basic	−0.287
Widyahastuti et al. (2017)	Exam score	Online text submissions	Course-specific	Basic	4.514

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Choi et al. (2016)	Total score	Average number of discussion topics posted by a teacher	Course-specific	Basic	−0.177
		Average number of wiki topics posted by a teacher	Course-specific	Basic	−0.295
		Average number of assignments posted by a teacher	Course-specific	Basic	0.106
		Average number of wiki viewed by a student	Course-specific	Basic	0.141
		Average ratio of completed quiz by a student	Course-specific	Elaborated	0.164
		Average ratio of completed assignments by a student	Generic	Elaborated	0.177
		Average number of discussions participated by a teacher	Course-specific	Basic	0.160
de Marcos et al. (2016)	Total score	Closeness centrality	Generic	Elaborated	0.553
		Eccentricity	Generic	Elaborated	−0.762
		Eigenvector centrality	Generic	Elaborated	−0.965
Goggins and Xing (2016)	Total score	Number of posts a student wrote during this module	Course-specific	Basic	0.500
		Total frequency of reading action recorded by CAN during this time	Course-specific	Basic	0.810
		An inverse value of the total delay time for response divided by number of posts to create an average	Course-specific	Elaborated	0.480
		Total time student used for consecutive reading actions	Course-specific	Basic	0.680
Naumann and Salmerón (2016)	Exam score	Number of task-relevant pages visited	Course-specific	Basic	0.130
		Number of task-irrelevant pages visited	Course-specific	Basic	−0.020
Strang (2016)	Total score	Course logins	Generic	Basic	0.137
		Lesson reading activity identified by Moodle system logs	Course-specific	Basic	0.134
		Lesson quiz activity identified by Moodle system logs	Course-specific	Basic	0.378
		Lesson quiz scores identified by Moodle system logs	Course-specific	Basic	0.158

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Yamada et al. (2016)	Total score	Number of slide pages that learners read	Course-specific	Basic	0.469
You (2016)[1]	Total score	Whether the students consistently accessed the course material without delay and completed the assigned learning content	Generic	Elaborated	0.400
		Student's failure to submit assignments on time	Generic	Basic	−0.36
		Number of course logins	Generic	Basic	0.150
		Whether a student downloaded and read the course information	Generic	Basic	0.100
You (2016)[2]	Exam score	Whether the students consistently accessed the course material without delay and completed the assigned learning content	Generic	Elaborated	0.350
		Student's failure to submit assignments on time	Generic	Basic	−0.180
		Number of course logins	Generic	Basic	−0.180
		Whether a student downloaded and read the course information	Course-specific	Basic	0.110
Joksimović, Gašević, Kovanović, et al. (2015)	Total score	Continuing a thread	Course-specific	Basic	0.930
Joksimović, Gašević, Loughin, et al. (2015)	Total score	Number of the student-content interactions	Generic	Basic	−0.09
		Time spent on student-system interaction types	Generic	Basic	0.030
Jo et al. (2015)	Exam score	(Ir)regularity of login interval	Generic	Elaborated	−0.590
Junco and Clem (2015)	Total score	Number of days the student used their textbook	Generic	Basic	0.338
Kennedy et al. (2015)	Total score	Active days	Generic	Basic	0.204
		Assignment switches	Course-specific	Basic	0.314
Pardo et al. (2015)	Exam score	Number of annotations submitted by each student	Course-specific	Basic	0.230

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Svihla et al. (2015)	Exam score	Total minutes for unit	Generic	Basic	0.139
		Number of days visited static curriculum step	Course-specific	Basic	−0.172
		Number of days visited dynamic visualization step	Course-specific	Basic	0.180
Wei et al. (2015)	Exam score	Number of times that students logged into the online course	Generic	Basic	0.060
		Number of times that students spent on reading learning materials	Course-specific	Basic	0.090
		Number of postings on the discussion board	Course-specific	Basic	0.060
You (2015)[1]	Total score	Absence of weekly attendance	Generic	Elaborated	−0.460
		Late submission of assignments	Generic	Elaborated	−0.400
You (2015)[2]	Exam score	Absence of weekly attendance	Generic	Elaborated	−0.420
		Late submission of assignments	Generic	Elaborated	−0.210
Zacharis (2015)	Total score	Frequency of reading and posting messages via either discussion forum, chat or emails	Course-specific	Basic	0.481
		Content creation in classwiki and site blog	Course-specific	Basic	0.299
		Quiz attempt, quiz continue attempt, and quiz close attempt	Course-specific	Basic	0.152
		Uses of the available resources	Course-specific	Basic	0.124
Agudo-Peregrina et al. (2014)	Total score	Exchanges between the students enrolled in a course	Generic	Basic	0.209
		Participation level of teachers and the extent to which students perceive a teacher's proximity through online presence	Course-specific	Basic	0.508
		Completing and sending individual and group assignments, quizzes, questionnaires, or other similar tasks	Course-specific	Basic	0.563

(Continues)

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
		Interactions in active task	Course-specific	Basic	0.489
Yoo and Kim (2014)	Total score	Average posting time to deadline	Course-specific	Basic	0.200
Yu and Jo (2014)	Total score	Total studying time in LMS	Generic	Basic	0.238
		Interactions with peers	Generic	Basic	0.278
Gijlers and de Jong (2013)[1]	Exam score	Number of messages related to integration-oriented consensus building (social mode of collaboration)	Course-specific	Basic	0.179
		Number of messages related to orientation (inquiry-learning process)	Course-specific	Basic	−0.051
Gijlers and de Jong (2013)[2]	Exam score	Number of messages related to integration-oriented consensus building (social mode of collaboration)	Course-specific	Basic	0.349
Lin and Chiu (2013)	Total score	Total number of online sessions	Generic	Basic	0.414
		Total number of follow-up posts created	Course-specific	Basic	0.132
		Total number of posts read	Course-specific	Basic	−0.152
Miller and Soh (2013)	Exam score	Assessment total clicks	Generic	Basic	0.125
		Tutorial total seconds	Course-specific	Basic	0.002
		Tutorial average seconds on a page	Course-specific	Basic	0.043
		Tutorial minimum seconds on a page	Course-specific	Basic	0.265
		Tutorial minimum clicks on a page	Course-specific	Basic	7.595
		Exercise average seconds on a page	Course-specific	Basic	0.013
		Exercise minimum seconds on a page	Course-specific	Basic	0.017
		Exercise average entries	Course-specific	Basic	0.141
		Exercise total interval	Course-specific	Basic	−0.0001

Authors	The type of the dependent variable	Log variables	The type of log variables		Regression coefficients
Ritter et al. (2013)	Exam score	Average of the sum of the number of hints and the number of errors in each problem	Course-specific	Basic	−0.186
		Total number of sections attempted by the student	Course-specific	Basic	0.267
		Total number of skills within the sections that the student encountered	Course-specific	Basic	−0.206
Bernacki et al. (2012)	Exam score	Highlight made	Course-specific	Basic	0.049
		Click on checklist of learning goals	Course-specific	Basic	0.060
Romero-Zaldivar et al. (2012)	Total score	Time in minutes using the virtual machine	Course-specific	Basic	8.071