



Learning analytics in distance education: A systematic review study

Abdulkadir Palanci¹ · Rabia Meryem Yılmaz² · Zeynep Turan¹

Received: 25 July 2023 / Accepted: 20 April 2024
© The Author(s) 2024

Abstract

This study aims to reveal the main trends and findings of the studies examining the use of learning analytics in distance education. For this purpose, journal articles indexed in the SSCI index in the Web of Science database were reviewed, and a total of 400 journal articles were analysed within the scope of this study. The systematic review method was used as the research methodology. The first study on the subject was published in 2011, and the publications continued to increase over the years. It was found that the publications on the subject were primarily found in “Computers and Education” and “Education and Information Technologies” journals. It was observed that China, USA, and Spain were the leading countries where the related studies were conducted. The studies primarily used the quantitative method, and university students were included as the sample. In addition, within the scope of learning analytics, it was observed that the data were mainly analysed with “regression analysis”, “correlation analysis”, “special algorithms-models”, “ANOVA”, and “cluster analysis” methods. It was determined that the most preferred platforms in the studies were learning management systems and MOOCs, learning behaviours were mostly examined, and log data were mainly used in this process. It was observed that the variables tested in the studies mainly consisted of students’ behaviours on the platform, learning performances, communication processes, dropout behaviours and course designs. Furthermore, in the studies examined, the advantages of learning analytics in the context of distance education are mostly related to the possibilities of improving the teaching process, and as disadvantages, it is stated that learning analytics is not suitable for use in some situations, negatively affect students’ performances, have limited interaction with students and are an expensive investment.

Keywords Learning analytics · Distance education · E-learning · Open education · Systematic review

Extended author information available on the last page of the article

1 Introduction

Distance education has become one of the prominent education methods, especially with the COVID-19 outbreak, and educational institutions worldwide have urgently switched to distance education with the pandemic (Masalimova et al., 2022). Distance education, widely used today, is frequently criticised for its shortcomings, such as needing help to individualise the teaching process and meeting individual needs sufficiently (Meacham et al., 2020). Distance education has some disadvantages as well as advantages. Some of these are that the relationships in face-to-face education cannot be easily provided, students who do not have the habit of self-learning cannot be adequately supported, and they cannot effectively develop skills and attitudes in practice-based courses. In addition, despite the increasing demand for distance education, students generally tend to drop out at a high rate (Bağrıacık Yılmaz et al., 2022). Research on online environments shows that students' lack of engagement is associated with failure and dropout (Caspari-Sadeghi, 2022). Some students in open and distance education leave their education due to reasons arising from the school and the programme (Okur et al., 2019). Also, students' motivational attrition, satisfaction, individual perceptions, poorly designed courses, and lack of digital competence may cause them to drop out of distance education (Bawa, 2016).

For this reason, care should be taken to ensure that students are retained and social interaction opportunities are provided, student-student interaction should be guaranteed, and students in the risk group should be identified by taking into account the academic achievements of the students (Bağrıacık Yılmaz et al., 2022). In this direction, learning analytics that enables monitoring, measuring, and visualising online behaviours can be used to identify students in the risk group (Brito et al., 2019; Caspari-Sadeghi, 2022; Celik et al., 2022; Lacave et al., 2018; Queiroga et al., 2020). Analysing this data allows for valuable knowledge discoveries to improve learning and assist institutions in strengthening distance learning courses and reducing student dropout rates (Silva et al., 2022). Furthermore, there is an increasing interest in research on using learning analytics in distance education (Silva et al., 2022). Indeed, the application of learning analytics in an online learning environment is becoming increasingly prevalent among educational researchers as it can help make standardised and measurable decisions about student achievement (Kew & Tasir, 2022).

Learning analytics is a research discipline that uses data analysis to support learners and optimise learning processes and learning environments (Saqr, 2018). Learning analytics, which represents the research area, can be defined as measuring, collecting, analysing, and interpreting the data obtained about the context to make sense of it (Siemens et al., 2011; Siemens, 2013). Learning analytics, a subfield of technology-supported learning with its rapid development and rich outputs is closely related to various fields such as educational data mining, web analytics, statistics, and artificial intelligence (Chen et al., 2022). While some systems with embedded learning analytics tools inform educators about how to intervene in students' education, some systems provide students with information about their progress (Jones, 2019). Learning analytics provides information on analysing learner data and learning processes and aims to develop evidence-based learning systems. Learning analytics also facilitates personalised learning, enabling students to have a more effective learning experience

(Caspari-Sadeghi, 2022). It also guides instructors in grading students' performance and keeping track of their diaries, attendance, and pace (Smith et al., 2012). Learning analytics encourages trainers and educational institutions to adapt to learners' needs and ability levels. Learning analytics has rapidly become popular in recent years due to the advantages of data-driven decision-making in teaching and learning (Du et al., 2021). In this context, learning analytics can support and encourage applications to evaluate students' progress, motivation, attitude, and satisfaction (Mangaroska & Giannakos, 2019). Considering all these, the possibilities offered by learning analytics provide new opportunities for learning (Greller & Drachsler, 2012). In addition, learning analytics, a data-driven approach, has great potential to provide an objective picture of the learning process by revealing the relationships between learning processes and achievement variables (Caspari-Sadeghi, 2022).

One of the purposes of using learning analytics is to understand and improve learning activities in higher education (Corrin et al., 2020). It is essential to understand the decisions made in using learning analytics applications in education for guidance purposes and the pedagogical and ethical implications of these decisions (Siemens, 2019). In particular, several ethical challenges and questions affect higher education optimisation (Long et al., 2011; Slade & Prinsloo, 2013). Ethical principles for learning analytics include data privacy, ownership and control, anonymity, transparency, do no harm, benefit, governance and security, consent, and openness (Corrin et al., 2019).

The number of studies using learning analytics in distance education is relatively high. Still, considering the increasing interest in these studies, a limited number of literature reviews have been published (Hantooobi et al., 2021). In this context, although it is seen that various systematic review studies on the use of learning analytics in educational environments have been conducted in the literature, only Kilis and Gülbahar (2016) have been found in the context of distance education. In the systematic review conducted by Kilis and Gülbahar (2016), the relevant studies in Eric and ScienceDirect databases were examined, and the findings of 25 studies on the subject were presented. Our study differs from this one in that it is a more comprehensive review of databases and analyses the studies in detail. Therefore, more studies are needed to investigate the benefits of learning analytics in distance education, to get detailed information about data collection processes, and to get information about the actual results (Kew & Tasir, 2022; Larrabee Sønderlund et al., 2019). This study is critical in providing guiding findings for future researchers and practitioners. This study aims to examine studies examining the use of learning analytics in distance education and to present a holistic picture of the literature. In line with this purpose, the following research questions were sought to be answered within the scope of the study:

1. What are the methodological characteristics of studies examining learning analytics in distance education?
 - a. How is their distribution according to years?
 - b. How is their distribution according to the journals in which they were published?
 - c. How is their distribution according to the countries where they were conducted?

- d. How is their distribution according to research methods?
 - e. How is their distribution according to sample groups?
 - f. How is their distribution according to sample sizes?
 - g. How is their distribution according to research topics?
 - h. How is their distribution according to the data analysis methods used?
 - i. How is their distribution according to the types of data used?
 - j. What is their distribution according to the data sources used?
 - k. What is their distribution according to the learning environments used?
 - l. What is their distribution according to the variables tested?
2. How does learning analytics affect student learning in distance education?
 3. What are the advantages and challenges of learning analytics in distance education?

2 Method

This study aims to examine scientific studies on learning analytics and descriptive content analysis using a systematic review method to identify the current situation. The systematic review is a method that requires careful analysis of the consistency of research findings (Moher et al., 2009; Tranfield et al., 2003). PRISMA guidelines were used to support careful planning, review, consistent execution of the process, research integrity, and transparency of reviews before starting the research (Moher et al., 2015).

2.1 Data collection

A Web of Science database was used to access the studies in this systematic review. Studies on using learning analytics in distance education were searched among SSCI-indexed journal articles with 12 research queries. The search was completed on 27 November 2023. In the screening process, “Distance Education AND Learning Analytics”, “Distance Learning AND Learning Analytics”, “Distance Teaching AND Learning Analytics”, “Open Learning AND Learning Analytics”, “Open Teaching AND Learning Analytics”, “Open Education AND Learning Analytics”, “Online Learning AND Learning Analytics”, “Online Teaching AND Learning Analytics”, “Online Education AND Learning Analytics”, “e-learning AND Learning Analytics”, “Electronic Learning AND Learning Analytics” and “Electronic Teaching AND Learning Analytics” queries were used. No filtering was made except for the SSCI index and search queries during the data collection phase.

2.2 Data analysis

In this study, the studies accessed with the specified keywords were analysed by two researchers. The other researcher checked the analyses to increase the reliability of the study. A form was created in Microsoft Excel for data analysis of each study. The

form includes sections that answer the research questions. These sections consist of the name of the study, type of the study, year of publication, journal in which the study was published, country in which the study was conducted, method of the study, size of the sample, subject of the study, data analysis method, tools used, sources of data obtained in the study, educational level of the sample, type of learning environment, variables tested, evidence of learning analytics to support learning, advantages/ findings and challenges. The studies included in the systematic review were meticulously read, and the prepared form sections were filled in separately. The data in the filled forms were transformed into codes, categories, graphs, and tables. The data obtained were presented descriptively.

Using the Web of Science database, as a result of the predetermined search queries, the full text of 4227 studies was accessed, as shown in Fig. 1. Two studies were excluded because they were single-page questionnaires or questionnaire answers. After excluding duplicates, 1823 studies were analysed according to their titles and abstracts. It was determined that 11 of these studies were not in English, 175 were not related to learning analytics, and 1055 were not associated with distance education. The full texts of the remaining 582 studies were analysed for relevance. Regarding

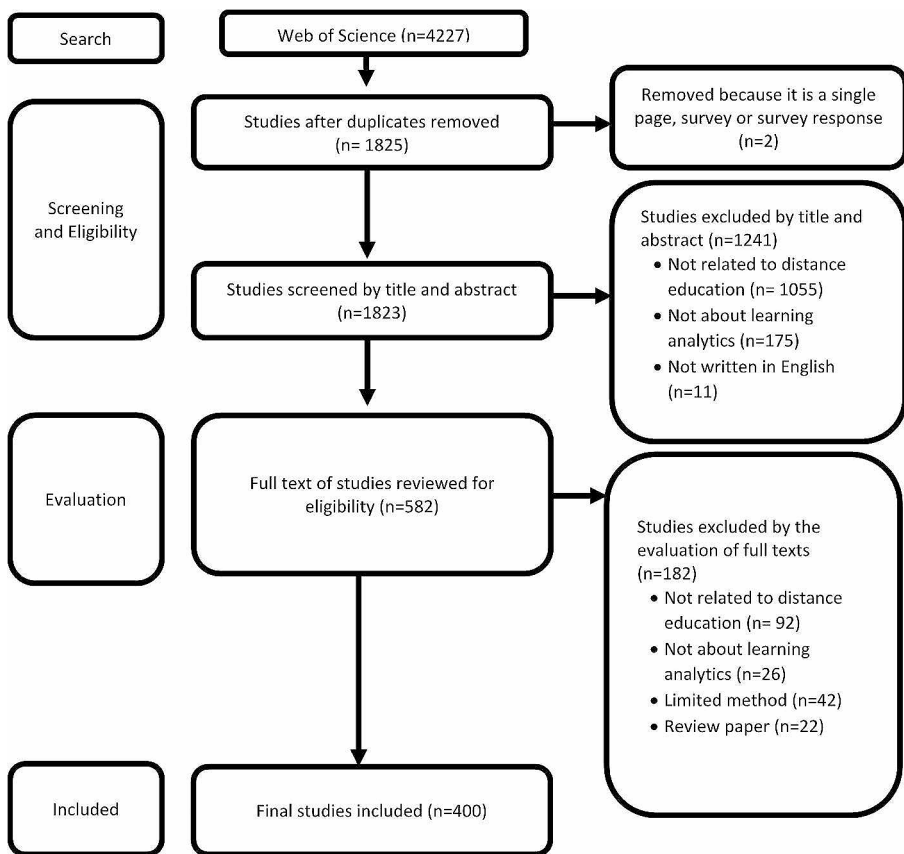


Fig. 1 PRISMA flow diagram

appropriateness, 26 studies were not included in the analyses because they were unrelated to learning analytics, 42 had limited methods, and 92 were unrelated to distance education. In addition, 22 review studies were not included in the analysis because they contained limited findings. As a result of all analyses, 400 articles were included in the systematic review.

3 Findings

3.1 Distribution of the studies based on the years

The distribution of the studies included in the analyses within the scope of the study by years was examined and presented in Fig. 2. According to Fig. 2, the studies accessed within the scope of the study first started in 2011, and it was determined that the number of published studies increased after 2019. It was determined that the number of studies examined reached the highest in 2021 ($n=74$) and 2022 ($n=69$), respectively.

3.2 Distribution of the studies in journals published

The distribution of the studies examined within the scope of the study according to the journals in which they were published was analysed and presented in Table 1. When Table 1 was examined, it was found that most studies were published in “Computers & Education” ($n=33$) and “Education and Information Technologies” ($n=26$).

3.3 Distribution of the countries where the studies conducted

Table 2 shows the distribution of the analysed studies according to the countries. According to Table 2, most studies were conducted in China ($n=69$), the United States of America ($n=62$), and Spain ($n=58$).

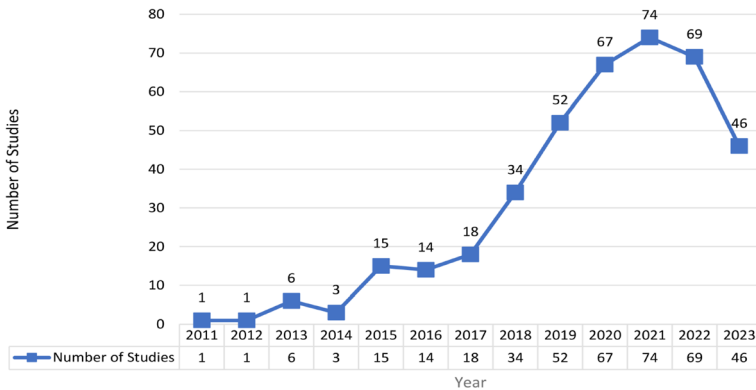


Fig. 2 Distribution of the studies based on the years

Table 1 Distribution of studies according to published journals

Journal Name	f	Journal Name	f
Computers & Education	33	AERA Open	1
Education and Information Technologies	26	Asia Pacific Education	1
Sustainability	22	Business Process Management Journal	1
Interactive Learning Environments	21	Computational Intelligence and Neuroscience	1
Computers in Human Behavior	20	Decision Support Systems	1
IEEE Transactions on Learning Technologies	19	Educational Psychology	1
International Review of Research in Open and Distributed Learning	17	Higher Education	1
Applied Sciences	16	IEEE Transactions on Emerging Topics in Computing	1
Journal of Computer Assisted Learning	16	IEEE Transactions on Engineering Management	1
British Journal of Educational Technology	13	Interacting with Computers	1
Education Technology Research & Development	13	International Journal of Human-Computer Interaction	1
The Internet and Higher Education	10	International Journal of Intelligent Systems	1
International Journal of Educational Technology in Higher Education	10	International Journal of Technology and Design Education	1
Educational Technology & Society	8	Journal of Medical Internet Research	1
<i>Frontiers in Psychology</i>	8	Journal of the Association for Information Science and Technology	1
Journal of Educational Computing Research	8	Journal of the Australian Library and Information Association	1
Computer Assisted Language Learning	7	Educational Sciences: Theory & Practice	1
IEEE Access	6	Language Learning & Technology	1
Distance Education	6	Learning and Instruction	1
BMC Medical Education	6	Life Sciences Education	1
Assessment & Evaluation in Higher Education	5	Nurse Education in Practice	1
Australasian Journal of Educational Technology	5	Teaching and Learning in Medicine	1
Journal of Computing in Higher Education	5	Teaching and Teacher Education	1
Internet and Higher Education	4	Telematics and Informatics	1
PLOS ONE	4	User Modeling and User-Adapted Interaction	1
Behaviour & Information Technology	4	Journal of Sport and Health Science	1
Future Generation Computer Systems	4	Information Systems Research	1
British Educational Research Association	4	BMC Public Health	1
Computer Applications in Engineering Education	3	Journal of Pacific Rim Psychology	1
Journal of Computing in Higher Education	3	Education+Training	1
Journal of Universal Computer Science	3	Sensors	1
American Behavioral Scientist	2	Academic Medicine	1
Digital Object Identifier	2	ReCALL	1
International Forum of Educational Technology & Society	2	Technology, Pedagogy and Education	1
Learning and Individual Differences	2	International Journal of Educational Research	1
Higher Education Research & Development	2	Teaching in Higher Education	1
Nutrients	2	Universal Access in the Information Society	1
International Journal of Computer-Supported Collaborative Learning	2	Journal of Science Education and Technology	1

Table 1 (continued)

Journal Name	f	Journal Name	f
Physical Review Physics Education Research	2	International Journal of Environmental Research and Public Health	1
Scientific Reports	2	Systems	1
Studies in Higher Education	2	Social Science Computer Review	1
Medical Teacher	2	Data Technologies and Applications	1
Journal of Research on Technology in Education	2	Education for Chemical Engineers	1
Accounting & Finance	1		
Behavioral Sciences	1		
Journal of Management Information Systems	1		
Active Learning in Higher Education	1		

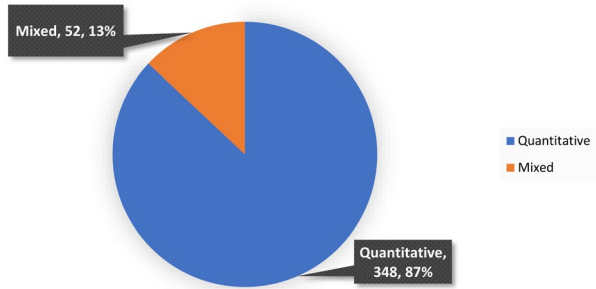
Table 2 Distribution of the countries where the studies were conducted

Country	f	Country	f
China	69	Greece	2
United States of America	62	Mauritius	2
Spain	58	Portugal	2
Australia	34	Singapore	2
United Kingdom	19	South Korea	2
Korea	13	Algeria	1
Canada	12	Austria	1
Israel	11	Belgium	1
Netherlands	10	Estonia	1
Turkey	10	France	1
Germany	10	Guatemala	1
Saudi Arabia	9	Serbia	1
England	7	Slovakia	1
Brazil	5	South Africa	1
India	5	Fiji	1
Czech Republic	4	Albania	1
Japan	4	Ireland	1
New Zealand	4	Thailand	1
Finland	4	Denmark	1
Pakistan	4	Malaysia	1
Chile	3	Cyprus	1
Italy	3	Nigeria	1
Sweden	3	Luxembourg	1
Colombia	2	Romania	1
Norway	2	Oman	1
Egypt	2	Iceland	1

3.4 Distribution of the methods used in the studies

Figure 3 shows the distribution of the analysed studies according to their methods. When Fig. 3 is examined, it is seen that quantitative (87%) and mixed (13%) research methods are primarily used in the analysed studies.

Fig. 3 Distribution of methods used in the studies



3.5 Distribution of the studies according to the sample group

Figure 4 shows the distribution of studies using learning analytics in distance education according to the sample group. As shown in Fig. 4, it was determined that university students ($n=245$) were mostly preferred as the sample group in the related studies. In 18 studies, the sample group was not specified.

3.6 Distribution of studies according to sample sizes

The distribution of the studies according to sample sizes is shown in Fig. 5. As shown in Fig. 5, it was determined that the most frequently preferred sample size was between 0 and 499 ($n=189$). In 41 studies, it is seen that the sample size is not specified.

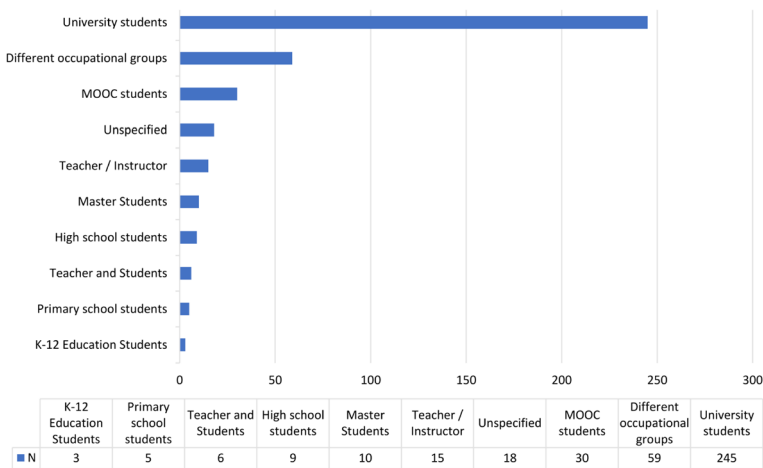


Fig. 4 Distribution of the studies based on the sample group

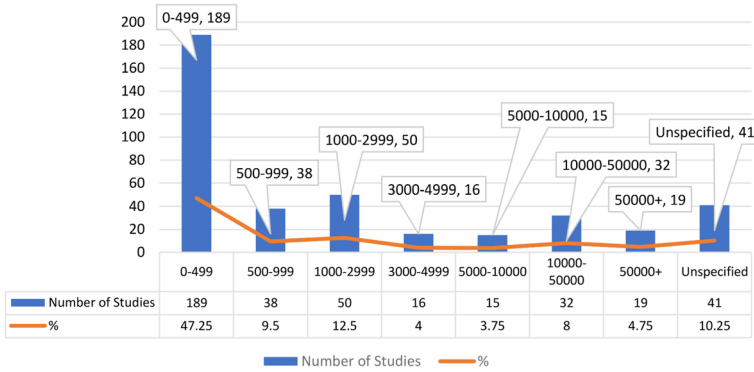


Fig. 5 Distribution of studies according to sample sizes

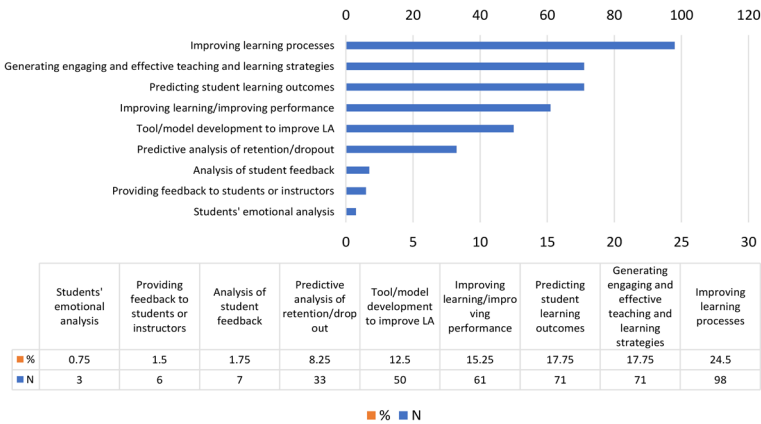


Fig. 6 Distribution of studies by research topic

3.7 Distribution of studies by research topic

The distribution of studies according to research topics is shown in Fig. 6. When Fig. 6 is analysed, it is determined that most studies were conducted on improving learning processes ($n=98$) and creating interesting and effective learning and teaching strategies ($n=71$).

3.8 Distribution of studies by data analysis methods

The data analysis methods used in the studies are shown in Table 3. When Table 3 is examined, it is seen that the most commonly used data analysis methods are regression analysis ($n=81$), correlation analysis ($n=67$), special algorithms-models ($n=56$), ANOVA ($n=36$), and cluster analysis ($n=33$).

Table 3 Data analysis methods used in the studies

Data analytics methods	f	Data analytics methods	f
Regression analysis	81	Exploratory analyses	2
Correlation analysis	67	A Bayesian network analysis	2
Special algorithms-models	56	Hidden Markov model	2
ANOVA	36	Qualitative analysis	2
Cluster analysis	33	Visualisation methods	2
Logistic regression	28	A spatio-temporal analysis	1
t-test	28	Behavior-sentiment topic mixture	1
Descriptive analysis	24	Benchmarking analysis	1
Social network analysis	23	Computer information systems	1
Decision trees	20	CHAID tree	1
Support vector machines	19	Data visualisation	1
Random forest	16	Explanatory analyses	1
Unspecified	15	ELECTRE TRI method	1
Coding-content analysis	14	Group comparison tests	1
Mann Whitney U test	11	General bayesian network	1
Machine learning methods	12	Generalised linear model	1
ANCOVA	10	Heuristic analysis	1
Structural equation modeling	10	Inferential statistical analysis	1
Chi-square test	9	Linear discriminant analysis	1
Thematic data analysis	8	Lag sequential analysis	1
Text mining	8	LOESS regression	1
Data mining	6	Multivariate analysis	1
Classification algorithms	6	Multi-platform learning analytics	1
Deep learning	6	Means testing	1
Process mining	6	Multiplex network analysis	1
Artificial neural network	5	Probability analysis	1
Factor analysis	5	Path analysis	1
MANOVA	4	Survival analysis	1
K-medoids clustering	4	Two-step cluster analysis	1
Epistemic network analysis	4	Tukey	1
K-Nearest Neighbour	4	Univariate and Bivariate analysis	1
Network analysis	3	Variance inflation factor	1
Partial least squares path modelling	3	Web analytic	1
Structural equation model	3	Sentimental analysis	1
Deep neural network	2		

3.9 Data types used in studies

The data from the analysed studies were coded and presented in 4 categories: learning behaviour data, learning emotional data, learning network data, and learning level data (Wu et al., 2015). Figure 7 shows the data types used in the analysed studies. When Fig. 7 is examined, it is determined that learning behaviour data (51%) col-

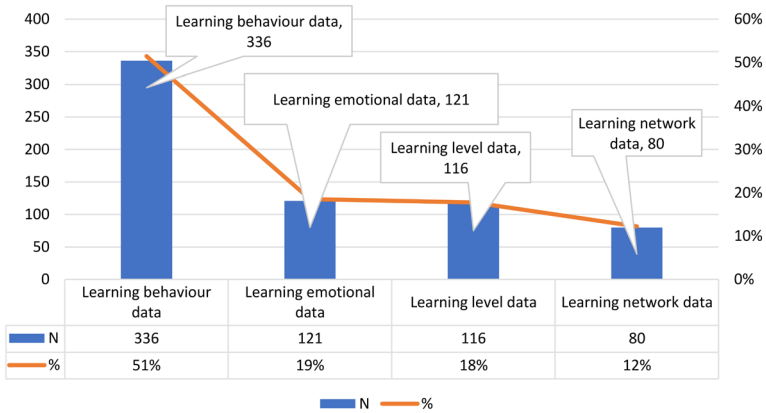


Fig. 7 Data types used in studies

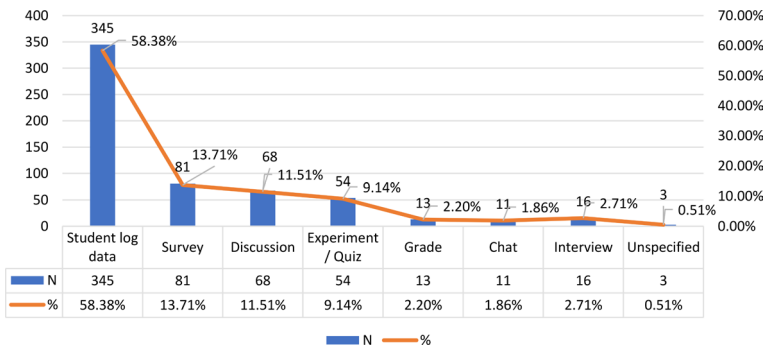


Fig. 8 Data sources used in studies

lected with the learning platform, including learners and learning resources, are frequently used.

3.10 Data sources used in studies

The data sources used in the analysed studies are given in Fig. 8. When Fig. 8 is analysed, it is seen that the most commonly used data source is student log data (58.38%).

3.11 Learning environments types used in studies

The types of learning environments used in the analysed studies are given in Fig. 9. When Fig. 9 is analysed, it is determined that the most commonly used data learning environments are LMS (48%) and MOOC (21%).

Fig. 9 Learning environments types

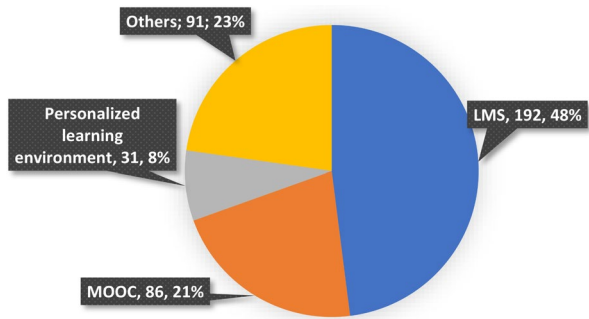


Table 4 Variables tested

Tested variables	f	Tested variables	f
Platform behaviour of students' data/log	165	Students' feelings	14
Learning performance of students	137	Students' motivation	14
Student's learning outcomes	47	Students' satisfaction	13
Course design	37	Student's attitudes	13
Student communication variables	34	Learner characteristics	11
Students' dropout behaviors	20	Students' confidence	7
Students' thoughts	17	Student attention	2
Ease of use	14	Course content	6

3.12 Tested variables in studies

The variables tested in the analysed studies are given in Table 4. When Table 4 is examined, it is determined that student data or daily platform behaviours ($n=165$) and students' learning performances ($n=137$) are the most tested variables.

3.13 Evidence of learning analytics to support learning in studies

The evidence of learning analytics supporting learning in the analysed studies is given in Fig. 10. When Fig. 10 is analysed, it is seen that the findings that learning analytics support learning (52%) are in the majority.

3.14 Mentioned advantages of using learning analytics

As a result of the use of learning analytics in distance education, the advantages stated in the analysed studies are given in Table 5. When Table 5 is analysed, it is seen that the findings that learning analytics in distance education improves learning outcomes ($n=145$), enables the examine students' activities ($n=134$), explains students' learning habits ($n=61$), provides to develop teaching strategy ($n=58$), increases students' participation ($n=44$), provides deep information about students' learning ($n=44$), enables to predict students' performance ($n=38$) come to the fore among the stated advantages.

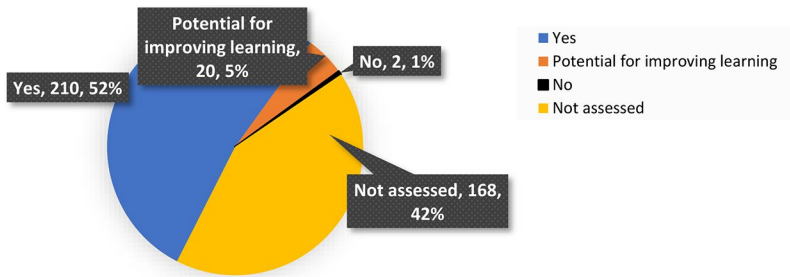


Fig. 10 Learning analytics evidence to support learning

Table 5 Advantages of using learning analytics in distance education

Advantages	f	Reference
Improves learning outcomes.	145	(Petrovic-Dziedz, 2019)
Allows to analyse the activities of students.	134	(Rizvi et al., 2020)
Describes the learning habits of students.	61	(Choi et al., 2018)
Provides to develop teaching strategy.	58	(Chen et al., 2020)
Increases student participation.	44	(Boulton et al., 2018)
Provides deep knowledge about students' learning.	44	(Yasmin, 2013)
It enables the prediction of student performance.	38	(Hernández-García et al., 2015)
Identifies students who are about to drop out.	34	(Burgos, 2019)
It ensures effective interventions.	32	(Pereira et al., 2020)
It helps students to adapt.	29	(Rienties et al., 2018a, b)
It ensures feedback.	28	(Choi & Cho, 2020)
Provides feedback to improve the learning design.	21	(Yang & Chen, 2020)
Provides assessment service.	16	(Akçapınar et al., 2019)
Provides personalised learning.	14	(Choi & McClenen, 2020)
Enriches learning environments.	13	(Yen et al., 2015)
Improves educational decision-making.	9	(Xia, 2020)
Provides modeling of student success.	8	(Moreno-Marcos et al., 2019)
It enables the improvement of instructor performance.	7	(Herodotou et al., 2019)
Increases self-awareness.	2	(Walsh, 2020)

Table 6 Challenges of using learning analytics in distance education in distance education

Challenges	f	Reference
Learning analytics is inadequate or unusable in some situations	18	(Yen et al., 2015)
Negative/ineffective performance for students	16	(Theobald et al., 2018)
Students had limited interaction	12	(Saqr et al., 2020)
Limited data	5	(Yin et al., 2019)
An expensive/costly investment	4	(Delgaty, 2013)
Students needed additional support	2	(Rienties et al., 2018a, b)

3.15 Mentioned challenges of using learning analytics

Within the scope of this systematic review, the difficulties encountered due to the use of learning analytics in distance education are given in Table 6. When Table 6 is analysed, it can be said that the challenges of using learning analytics in distance education are that learning analytics is inadequate or impractical in some cases ($n=18$), it shows negative or ineffective performance for students ($n=16$), students have limited interaction ($n=12$), it contains limited data ($n=5$), learning analytics is an expensive investment ($n=4$), and students need additional support ($n=2$).

4 Discussion

This study analysed the research trends and main findings of the articles on learning analytics in distance education in the SSCI-indexed journals in the Web of Science database. This systematic review showed increased articles on learning analytics in distance education after 2019. After 2019, the increasing number of studies can be explained as a result of the transition of many educational institutions to distance education due to the global pandemic. In addition, it was observed that there was a slight decrease in the number of studies on the subject after 2021. This may indicate that studies in the field of learning analytics have reached a certain level of saturation. In addition, with the prominence of machine learning algorithms and artificial intelligence concepts in recent years, research on the subject may have been gathered under these headings. In this context, it can be said that future studies on learning analytics will be more related to machine learning algorithms. It was determined that the studies on this subject were primarily published in “Computers & Education” and “Education and Information Technologies” journals. The fact that these journals are the leading journals in their fields, with high impact factors and prestigious journals, may explain the high number of publications on learning analytics, a new topic in these journals. In addition, it was determined that most publications on this subject were made in China, the USA, and Spain. However, it was determined that quantitative and mixed research designs were frequently used in the analysed studies. This is an expected result since learning analytics involves the analysis of big data. In the studies on the use of learning analytics in distance education, it has been observed that the sample group consists mainly of university students, and the number of samples is between 0 and 499. This situation can be explained by university students frequently

using distance education systems (MOOCs, LMS, etc.). Further, one of the primary purposes of using learning analytics is to understand and improve learning activities in higher education (Corrin et al., 2020). For this purpose, university students may have been focused frequently.

The research results showed that the most researched topic was improving learning processes and creating engaging and effective learning and teaching strategies. At this point, the advantages of learning analytics in distance education can be listed as enhancing learning processes, engaging and effective learning, creating learning strategies, predicting learning outcomes, predicting students' dropout or attendance, emotional analysis of students, and providing feedback to students or instructors. The results support that researchers frequently use learning analytics, the quality of learning environments is improved by analysing students' data in distance education environments, and meaningful data about learners can be obtained (Kew & Tasir, 2022). In this direction, the opportunities provided by learning analytics should be recognised, and learning analytics tools and techniques should be used and utilised to improve teaching and learning (Clow, 2013; Drugova et al., 2023).

The studies within this systematic review revealed that analysis methods that examine the relationships between variables, such as regression, correlation, ANOVA and cluster analysis, were primarily used to analyse the data obtained from learning analytics. Learning analytics is based on performing relational analyses to reveal how variables affect each other when analysing big data (Sghir et al., 2023). Therefore, this result is expected. Similar to the findings obtained from the study of Ifenthaler and Yau (2020), it was determined that the most commonly used data source in the studies was student log records. Data on distance education processes carried out through existing digital systems are obtained from system records. The fact that the most used learning environments are LMS and MOOCs supports these results. The variables tested were generally students' behaviours on the platform, learning performances, learning outcomes, communication status, dropout behaviours, students' opinions, and ease of use of the platforms. This finding is similar to the results obtained from the study of Sghir et al. (2023). It has been observed that learning behaviour data is frequently used through learning platforms in the studies examined. This finding is similar to the findings of Kew and Tasir (2022). Therefore, learning analytics tools can be developed and used to make healthier decisions about the teaching process (Kitto et al., 2017; Pantazos & Vatrappu, 2016).

In the studies examined, the findings that learning analytics support learning are predominant. This finding coincides with the results of Larrabee Sønderlund et al. (2019)'s review of intervention studies examining the effectiveness of learning analytics. Indeed, with the development of learning analytics, it has become easier to identify students' problems and intervene promptly using these tools in learning environments (Wong & Li, 2020). The use of learning analytics in distance education has been found to provide many advantages. Some of the advantages of using learning analytics in distance education are improving students' learning outcomes, enabling students to examine their learning activities, revealing learning habits, developing teaching strategies, allowing the students to participate in the course, receiving feedback from students, and identifying students who are about to drop the course.

In addition, learning analytics has the disadvantages of being inadequate or impractical in some cases, performing negatively for students, causing limited student interaction, and being an expensive investment. There are several ethical challenges and questions for learning analytics, especially ethical principles such as data privacy, ownership and control, anonymity, transparency, do no harm, benefit, governance and security, consent, and openness, which affect higher education optimisation (Corrin et al., 2019; Long et al., 2011; Slade & Prinsloo, 2013). The findings on the disadvantages of learning analytics are in parallel with the results of Avella et al. (2016). Despite all these, results show that learning analytics interventions have the potential to expand the scope of application further (Wong & Li, 2020). It can be said that learning analytics has an excellent potential to advance the innovation of personalised environments in particular (Dessi et al., 2019). Also, learning analytics can be useful with its techniques developed to collect and analyse large amounts of data, which seemed impossible in the early stages of distance education history (Gelan et al., 2018). Therefore, educational institutions may be advised to use learning analytics to identify students who are underperforming or at risk of absenteeism (Larrabee Sønderlund et al., 2019).

5 Conclusion, implications, and limitations

As a result, in this study, 400 journal articles regarding research trends and main findings from the Web of Science database were analysed. This research aimed to contribute by providing a general perspective on the use of learning analytics in distance education. The results showed that most studies were published in 2021, most were published in *Computers & Education* and *Education and Information Technologies* journals, and China, the USA, and Spain were leading in this field. Methodologically, it was concluded that quantitative and mixed methods were mainly used, university students and people from different professions were specifically studied, data primarily were obtained from log data, and regression analysis and correlation analysis were frequently preferred.

It was observed that the studies were conducted to improve learning processes and to create engaging/effective learning and teaching strategies, that the focus was mainly on student behaviour data, and that LMS and MOOCs were frequently used. It is seen that the most studied variables are student data or daily platform behaviours and students' teaching-learning performances. It has been determined that learning analytics in distance education improves learning outcomes, supports learning, and enables the analysis of students' activities. However, it has also been found that learning analytics is sometimes inadequate or impractical and performs negatively for students. In the light of these results, recommendations are as follows:

- Considering the rapid spread of distance education processes in recent years and their active use in lifelong learning processes, increasing the number of studies in this field is recommended.
- Researchers working on learning analytics in distance education can benefit primarily from *Computers & Education* and *Education and Information Technologies*.

gies journals, which have the highest number of publications.

- Researchers who want to collaborate internationally and will work on using learning analytics in distance education should contact countries such as China, the USA, and Spain, which are pioneers in this field.
- Since quantitative methods are frequently used in the studies to be conducted in this field, it is suggested that mixed-method studies should be included more in future studies.
- In the studies conducted, university students were frequently included as a sample. In future studies, MOOC students, primary and high school students, and teachers can also be included.
- For future studies, predictive analysis of retention/dropout, student feedback, and emotional analysis variables, which are relatively less focused in the current studies, can be focused on.
- It is recommended that machine learning and deep learning algorithms, which have been popular in recent years, should be included in data analysis processes for future research using more extensive data.
- In the existing studies, learning network data and learning level data types are included at a limited level. Future studies can be designed for these data types.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability I confirm I have included a data availability statement in my main manuscript file: My manuscript has no associated data.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16. <https://doi.org/10.1186/s41239-019-0172-z>.
- Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, 20(2), 13–29.

- Bağrıacıoğlu Yılmaz, A., & Karataş, S. (2022). Why do open and distance education students drop out? Views from various stakeholders. *International Journal of Educational Technology in Higher Education*, 19(1), 28. <https://doi.org/10.1186/s41239-022-00333-x>.
- Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—A literature review. *SAGE Open*, 6(1), 1–11. <https://doi.org/10.1177/2158244015621777>.
- Boulton, C. A., Kent, C., & Williams, H. T. P. (2018). Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university. *Computers & Education*, 126, 129–142. <https://doi.org/10.1016/j.compedu.2018.06.031>.
- Brito, M., Medeiros, F., & Bezerra, E. P. (2019). An infographics-based tool for monitoring dropout risk on distance learning in higher education In 2019, the 18th International Conference on Information Technology Based Higher Education and Training (ITHET), 1–7. IEEE. <https://doi.org/10.1109/ithet46829.2019.8937361>.
- Burgos, D. (2019). Background similarities as a way to predict students' behaviour. *Sustainability*, 11(24), 6883. <https://doi.org/10.3390/su11246883>.
- Caspari-Sadeghi, S. (2022). Applying learning analytics in online environments: Measuring learners' engagement unobtrusively. *Frontiers in Education*, 7, 840947. <https://doi.org/10.3389/educ.2022.840947>.
- Celik, I., Gedrimiene, E., Silvola, A., & Muukkonen, H. (2022). Response of learning analytics to the online education challenges during pandemic: Opportunities and key examples in higher education. *Policy Futures in Education*, 21(4), 387–404. <https://doi.org/10.1177/14782103221078401>.
- Chen, Z., Xu, M., Garrido, G., & Guthrie, M. W. (2020). Relationship between students' online learning behavior and course performance: What contextual information matters? *Physical Review Physics Education Research*, 16(1), 1–15. <https://doi.org/10.1103/physrevphyseducres.16.010138>.
- Chen, X., Zou, D., & Xie, H. (2022). A decade of learning analytics: Structural topic modeling based bibliometric analysis. *Education and Information Technologies*, 27(8), 10517–10561. <https://doi.org/10.1007/s10639-022-11046-z>.
- Choi, Y., & Cho, Y. I. (2020). Learning analytics using social network analysis and bayesian network analysis in sustainable computer-based formative assessment system. *Sustainability*, 12(19), 7950. <https://doi.org/10.3390/su12197950>.
- Choi, Y., & McClenen, C. (2020). Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Applied Sciences*, 10(22), 8196. <https://doi.org/10.3390/app10228196>.
- Choi, S. P. M., Lam, S. S., Li, K. C., & Wong, B. T. M. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Educational Technology & Society*, 21(2), 273–290.
- Crow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6), 683–695. <https://doi.org/10.1080/13562517.2013.827653>.
- Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper. <https://melbourne-cshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics>.
- Corrin, L., Scheffel, M., & Gašević, D. (2020). Learning analytics: Pathways to impact. *Australasian Journal of Educational Technology*, 36(6), 1–6. <https://doi.org/10.14742/ajet.6853>.
- Delgaty, L. (2013). A critical examination of the time and workload involved in the design and delivery of an e-module in postgraduate clinical education. *Medical Teacher*, 35(5), e1173–e1180. <https://doi.org/10.3109/0142159x.2012.737963>.
- Dessi, D., Fenu, G., Marras, M., & Reforgiato Recupero, D. (2019). Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections. *Computers in Human Behavior*, 92, 468–477. <https://doi.org/10.1016/j.chb.2018.03.004>.
- Drugova, E., Zhuravleva, I., Zakharova, U., & Latipov, A. (2023). Learning analytics driven improvements in learning design in higher education: A systematic literature review. *Journal of Computer Assisted Learning*, 1–15. <https://doi.org/10.1111/jcal.12894>.
- Du, X., Yang, J., Shelton, B. E., Hung, J. L., & Zhang, M. (2021). A systematic meta-review and analysis of learning analytics research. *Behaviour & Information Technology*, 40(1), 49–62. <https://doi.org/10.1080/0144929x.2019.1669712>.

- Gelan, A., Fastré, G., Verjans, M., Martin, N., Janssenswillen, G., Creemers, M., & Thomas, M. (2018). Affordances and limitations of learning analytics for computer-assisted language learning: A case study of the VITAL project. *Computer Assisted Language Learning*, 31(3), 294–319. <https://doi.org/10.1080/09588221.2017.1418382>.
- Greller, W., & Drachler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, 15(3), 42–57.
- Hantoobi, S., Wahdan, A., Al-Emran, M., & Shaalan, K. (2021). A review of learning Analytics studies. In M. Al-Emran, & K. Shaalan (Eds.), *Recent advances in Technology Acceptance models and theories* (pp. 119–134). Springer International Publishing.
- Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior*, 47, 68–80. <https://doi.org/10.1016/j.chb.2014.10.038>.
- Herodotou, C., Rienties, B., Boroooa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational Technology Research and Development*, 67(5), 1273–1306. <https://doi.org/10.1007/s11423-019-09685-0>.
- 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, 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>.
- Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education* pp. 16, 1–22. <https://doi.org/10.1186/s41239-019-0155-0>.
- Kew, S. N., & Tasir, Z. (2022). Learning analytics in online learning environment: A systematic review on the focuses and the types of student-related analytics data. *Technology Knowledge and Learning*, 27(2), 405–427. <https://doi.org/10.1007/s10758-021-09541-2>.
- Kilis, S., & Gülbahar, Y. (2016). *Learning analytics in distance education: A systematic literature review* Paper presented at the 9th European Distance and E-learning Network (EDEN) Research Workshop, Oldenburg, Germany.
- Kitto, K., Lupton, M., Daavis, K., & Waters, Z. (2017). Designing for student-facing learning analytics. *Australasian Journal of Educational Technology*, 33(5), 152–168. <https://doi.org/10.14742/ajet.3607>.
- Lacave, C., Molina, A. I., & Cruz-Lemus, J. A. (2018). Learning analytics to identify dropout factors of Computer Science studies through bayesian networks. *Behaviour & Information Technology*, 37(10–11), 993–1007. <https://doi.org/10.1080/0144929x.2018.1485053>.
- Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618. <https://doi.org/10.1111/bjet.12720>.
- Long, P., Siemens, G., Conole, G., & Gasevic, D. (2011). *Proceedings of the 1st international conference on learning analytics and knowledge (LAK11)* Paper presented at the International Conference on Learning Analytics and Knowledge (LAK11), Banff.
- Mangaroska, K., & Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, 12(4), 516–534. <https://doi.org/10.1109/tlt.2018.2868673>.
- Masalimova, A. R., Khvatova, M. A., Chikileva, L. S., Zvyagintseva, E. P., Stepanova, V. V., & Melnik, M. V. (2022). Distance Learning in Higher Education during Covid-19. *Frontiers in Education*, 7. <https://doi.org/10.3389/educ.2022.822958>.
- Meacham, S., Pech, V., & Nauck, D. (2020). AdaptiveVLE: An Integrated Framework for Personalized Online Education using MPS JetBrains Domain-Specific modeling Environment. *Ieee Access: Practical Innovations, Open Solutions*, p. 8, 184621–184632. <https://doi.org/10.1109/access.2020.3029888>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.1136/bmj.b2535>.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1–9. <https://doi.org/10.1186/2046-4053-4-1>.
- Moreno-Marcos, P. M., De Laet, T., Muñoz-Merino, P. J., Van Soom, C., Broos, T., Verbert, K., & Delgado Kloos, C. (2019). Generalizing predictive models of admission test success based on online interactions. *Sustainability*, 11(18), 4940. <https://doi.org/10.3390/su11184940>.

- Okur, M. R., Başı, D. F., & Güneş, E. (2019). Examination of dropout causes in open and distance learning. *Yükseköğretim ve Bilim Dergisi*, 9(2), 225–235. <https://doi.org/10.5961/jhes.2019.324>.
- Pantazos, K., & Vatraru, R. (2016). *Enhancing the Professional Vision of Teachers: A Physiological Study of Teaching Analytics Dashboards of Students' Repertory Grid Exercises in Business Education* Paper presented at the 2016 49th Hawaii International Conference on System Sciences (HICSS), 41–50. <https://doi.org/10.1109/HICSS.2016.14>.
- Pereira, F., De Oliveira, E. H. T., Oliveira, D., Cristea, A. I., Carvalho, L. S., Fonseca, S. C., Toda, A. M., & Isotani, S. (2020). Using learning analytics in the Amazonas: Understanding students' behaviour in introductory programming. *British Journal of Educational Technology*, 51(4), 955–972. <https://doi.org/10.1111/bjet.12953>.
- Petrovic-Dziedz, M. (2019). Gamifying online tests to promote retrieval-based learning. *The International Review of Research in Open and Distributed Learning*, 20(2), 24–43. <https://doi.org/10.19173/irrodl.v20i2.3812>.
- Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., & Cechinel, C. (2020). A learning analytics approach to identifying students at risk of dropout: A case study with a technical distance education course. *Applied Sciences*, 10(11), 3998. <https://doi.org/10.3390/app10113998>.
- Rienties, B., Herodotou, C., Olney, T., Schencks, M., & Boroowa, A. (2018a). Making sense of learning analytics dashboards: A technology acceptance perspective of 95 teachers. *The International Review of Research in Open and Distributed Learning*, 19(5). <https://doi.org/10.19173/irrodl.v19i5.3493>.
- Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2020a). Investigating variation in learning processes in a FutureLearn MOOC. *Journal of Computing in Higher Education*, 32(1), 162–181. <https://doi.org/10.1007/s12528-019-09231-0>.
- Rienties, B., Lewis, T., McFarlane, R., Nguyen, Q., & Toetnel, L. (2018b). Analytics in online and offline language learning environments: The role of learning design to understand student online engagement. *Computer Assisted Language Learning*, 31(3), 273–293. <https://doi.org/10.1080/09588221.2017.1401548>.
- Saqr, M. (2018). *Using learning analytics to understand and support collaborative learning* (Doctoral dissertation, Department of Computer and Systems Sciences, Stockholm University).
- Saqr, M., Viberg, O., & Vartiainen, H. (2020). Capturing the participation and social dimensions of computer-supported collaborative learning through social network analysis: Which method and measures matter? *International Journal of Computer-Supported Collaborative Learning*, 15(2), 227–248. <https://doi.org/10.1007/s11412-020-09322-6>.
- Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 28, 8299–8333. <https://doi.org/10.1007/s10639-022-11536-0>.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>.
- Siemens, G. (2019). Learning analytics and open, flexible, and distance learning. *Distance Education*, 40(3), 414–418. <https://doi.org/10.1080/01587919.2019.1656153>.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., & Verbert, K. (2011). & R. S. J. D. Baker. Open Learning Analytics: an integrated & modularized platform. <https://solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf>.
- Silva, L. M., Dias, L. P. S., Rigo, S. J., Anjos, J. C. S., Geyer, C. F. R., Leithardt, V. R. Q., & Barbosa, J. L. V. (2022). Learning analytics and collaborative groups of learners in distance education: A systematic mapping study. *Informatics in Education*, 21(1), 113–146. <https://doi.org/10.15388/infedu.2022.05>.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>.
- Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, 16(3), 51–61.
- 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>.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14, 207–222.

- Walsh, C., Mital, A., Ratcliff, M., Yap, A., & Jamaledine, Z. (2020). A public-private partnership to transform online education through high levels of academic student support. *Australasian Journal of Educational Technology*, 36(5), 30–45. <https://doi.org/10.14742/ajet.6107>.
- Wong, B. T. M., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, 7(1), 7–28. <https://doi.org/10.1007/s40692-019-00143-7>.
- Wu, F., Huang, L., & Zou, R. (2015). The design of intervention model and strategy based on the behavior data of learners: A learning analytics perspective. In S. K. S. Cheung, L. Kwok, H. Yang, J. Fong, & R. Kwan (Eds.), *International Conference on Hybrid Learning* (pp. 294–301). Springer International Publishing. https://doi.org/10.1007/978-3-319-20621-9_24.
- Xia, X. (2020). Learning behavior mining and decision recommendation based on association rules in interactive learning environment. *Interactive Learning Environments*, 31(2), 593–608. <https://doi.org/10.1080/10494820.2020.1799028>.
- Yang, T., & Chen, S. Y. (2020). Investigating students' online learning behavior with a learning analytic approach: Field dependence/independence vs. holism/serialism. *Interactive Learning Environments*, 31(2), 1041–1059. <https://doi.org/10.1080/10494820.2020.1817759>.
- Yasmin. (2013). Application of the classification tree model in predicting learner dropout behaviour in open and distance learning. *Distance Education*, 34(2), 218–231. <https://doi.org/10.1080/01587919.2013.793642>.
- Yen, C. H., Chen, I. C., Lai, S. C., & Chuang, Y. R. (2015). An analytics-based approach to managing cognitive load by using log data of learning management systems and footprints of social media. *Educational Technology & Society*, 18(4), 141–158.
- Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human factor approach. *International Journal of Human-Computer Interaction*, 35(4–5), 313–322. <https://doi.org/10.1080/10447318.2018.1543077>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Abdulkadir Palanci¹  · Rabia Meryem Yılmaz²  · Zeynep Turan¹ 

✉ Zeynep Turan
zeynepatauniv@hotmail.com; zeynepturan@atauni.edu.tr

Abdulkadir Palanci
abdulkadirpalanci.29@gmail.com

Rabia Meryem Yılmaz
rabia.kufrevi@gmail.com; rkufrevi@atauni.edu.tr

¹ Department of Computer Education & Instructional Technology, Ataturk University, Erzurum 25240, Turkey

² Department of Software Engineering, Engineering Faculty, Ataturk University, Erzurum 25240, Turkey