Artificial Intelligence in educational assessment: A bibliometric review of current applications and future perspectives

Fezile OZDAMLI[1,2], Erinc ERCAG[3]

[1] Department of Innovation and Knowledge Management,
Near East University, Nicosia

[2] Computer Information Systems Research and Technology Center,
Near East University, Nicosia

[3] Computer Education and Instructional Technology Department,
Kyrenia University, Kyrenia

fezile.ozdamli@neu.edu.tr, erinc.ercag@kyrenia.edu.tr

Abstract: *This study aimed to examine the studies on the use of artificial intelligence in the educational assessment process using the bibliometric analysis method. Scopus and Web of Science databases were preferred in the study and 733 documents were analyzed. The majority of the documents were published as journal articles and conference papers. In addition, English was preferred in almost all of them. It was determined that the studies on the use of artificial intelligence applications in the Educational Assessment process were carried out in the USA, China and England. The most frequently used keywords in the studies by the researchers were determined as "Artificial Intelligence", "Students", "Learning systems", "human", "teaching", "machine learning", "education", "humans".*

Keywords: Educational Assessment, Artificial Intelligence, AI, Formative Assessment.

1. Introduction

Artificial intelligence is a field of science that aims to create systems that can make decisions and perform tasks performed by humans faster and without the need for human intervention (Fetzer, 1990). Artificial intelligence applications can revolutionize many areas, such as disease diagnosis in health, financial forecasts and decision-making in economics, and autonomous vehicles in transportation. Artificial intelligence systems maximize performance by using natural language processing, expert systems, and deep learning algorithms, a machine learning sub-branch.

Big data is analyzed through algorithms, allowing artificial intelligence to recognize patterns and discover insights beyond human cognition. With artificial intelligence applications, ways to access information have changed. Applications

https://doi.org/10.58503/icvl-v20y202534

that process big data, identify complex patterns, and offer personalized solutions have become an excellent resource for every individual who learns.

As in many sectors, artificial intelligence applications have been widely used in the education sector for years. In fact, in 2019, the Artificial Intelligence and Education Council Conference was held, where many country ministers, United Nations representatives and academics participated, and decisions were made by discussing the use of artificial intelligence applications in education. At the conference held in Beijing, it was decided that the Sustainable Development Goals should be adhered to, especially the principle of quality education, and that artificial intelligence applications should be respectful of human rights and always human-centered. Accordingly, it was suggested that artificial intelligence applications should be used in a format that supports teachers. It was suggested that regulations should be developed in a way that data will be used transparently and securely (Krotz & Schelhowe, 2020). Artificial intelligence applications that will be developed in accordance with these principles will help learners personalize their learning processes while providing effective and interactive learning environments that will reduce the workload of educators. Despite the many opportunities that artificial intelligence applications provide, concerns about how they can be integrated into education continue (Sbaffi & Zhao, 2022; Gustilo, Ong & Lapinid, 2024). Applications such as ChatGPT have also begun to be used in education because they produce fluent texts (Baskara & Mukarto, 2023; Gustilo, Ong & Lapinid, 2024). Their use for content generation purposes is seen, from planning lesson plans and activities to student evaluation (Khasawneh, Ismail & Hussen, 2024).

In their studies, researchers mention that well-designed assessment processes and tools should be used to determine whether learners have learned a subject. They also note that traditional assessment practices such as multiple-choice tests and short-answer questions are widely used and cause some problems. The most important issue is that the design process is time-consuming (Swiecki et al., 2022) . Similarly, common exam stress can distract learners from the learning process, and traditional exam processes are time-consuming for both teachers and learners (Luckin, 2017).

Many researchers mention that using artificial intelligence applications for evaluation purposes in the learning process is one of the most important opportunities (Cope, Kalantzis & Searsmith, 2021; Swiecki et al., 2022; Hopfenbeck et al., 2023; Zhai & Nehm, 2023). Researchers today indicate that formative assessment motivates students more (Leenknecht et al., 2021; Kültür & Kutlu, 2021). AI applications are designed to partially or completely automate traditional assessment practices. Assessments can be designed with AI applications and even be conducted and scored by the application. Researchers have designed AI-powered assessments to provide teachers, learners, and families with regular feedback on how and how much students are learning and how much they are progressing toward their learning goals (Luckin, 2017).

Researchers have shown that artificial intelligence applications help evaluate the participation of learners in the specified lesson, while also helping to determine successful and relatively weak students in the educational institution according to their levels. Individual curriculums can be designed for the specified groups through artificial intelligence applications to increase students' performance. These applications not only increase learner success; they can also prevent school dropouts by helping to attract more students to the learning process (Dhara et al., 2022).

Many studies have examined the effects of using artificial intelligence applications in formative assessment processes. Researchers support teachers in formative assessment practices such as providing timely feedback to students and designing curriculum according to results (Zhai, He & Krajcik, 2022). In addition to many opportunities, it also causes concerns in the education world. Therefore, this study aims to determine the current research trends in artificial intelligence applications used for formative assessment. For the reasons mentioned above, a bibliometric study was conducted to determine the research topics and important trends in the field. This study, which can be a guide for future research, will help understand the effects of the use of artificial intelligence applications in educational assessment.

2. Methodology

Bibliometric analysis was conducted to systematically evaluate studies investigating the use of artificial intelligence applications in the educational assessment process. Large-volume data sets can be analyzed with bibliometric analysis, which is used in many studies (Ellegaard & Wallin, 2015; Donthu et al., 2021). In the educational assessment process, which researchers conducted the most studies on the use of artificial intelligence applications, which studies received the most citations in the studies on the subject, which countries gave the most importance to the subject, and the most frequently used keywords for determining the conceptual structure were determined.

2.1 Data collection

The data needed for the bibliometric analysis method was obtained from the SCOPUS and Web of Science (WoS) Core Collection databases. SCOPUS and WoS, which are the information and technology providers of the scientific research community, were selected because they are databases containing many high-impact scientific resources, including open-access journals (https://webofscience.help.clarivate.com/Content/home.htm, https://www.elsevier.com/products/scopus). In the research, the topic option, which searches for article title, abstract, keyword and author keywords in the Web of Science database, and the Article Title and abstract, Keywords in Scopus were preferred, and the keywords ("artificial intelligence" OR "AI") AND ("educational assessment" OR "learning evaluation" OR "academic assessment"

OR "learning assessment" OR "digital assessment" OR "formative assessment") were searched. All years, languages and all document types were included in the search. The query on Scopus yielded 633 publications, and the query on WoS yielded 261 publications. The data collection process ended on December 28, 2024. Publications indexed in both databases were identified and duplicates were deleted. As a result, the study was conducted with 733 publications.

2.2 Data analysis

In this study, VOSviewer software, which has multidimensional analysis, datasets, mapping and visualization features, was used (VOSviewer, 2020). VOSviewer is used as a free software and can easily create bibliometric maps consisting of big data (van Eck & Waltman, 2010; VOSviewer, 2020). The distribution of the publications included in the study according to publication years, countries, document types, and languages was calculated using percentage values.

2.3 Limitations of the study

In terms of method, the study was limited to the bibliometric analysis method and the data were analyzed with citation and co-occurrence analysis techniques. The search index of the research was limited to the keywords "Artificial Intelligence, AI, educational assessment, learning evaluation, academic assessment, learning assessment, digital assessment, formative assessment". Another limitation of the study is that only Scopus and WoS databases were used.

3. Findings

3.1 Distribution of studies according to publication years

The findings obtained from the analysis showed that 733 documents were in the databases examined "regarding the use of artificial intelligence in the educational assessment process" between 1996 and 2025. Figure 1 presents the distribution of the documents by year.

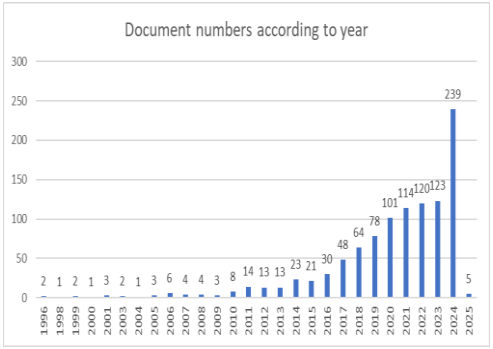

Figure 1. Publications over years

As seen in Figure 1, the first studies on using artificial intelligence applications in the educational assessment process were published in 1996 (n=2). The years when intensive research on the subject started were after 2016 (n=30). The most studies on the subject were in 2024 (n=239). This finding shows us that this subject is a trending research topic today.

3.2 Distribution of research according to document types

Table 1 presents the numbers and percentages of studies conducted on the use of artificial intelligence applications in the educational assessment process in the WoS and Scopus databases according to document types.

Table 1. Frequency and percentages of documents

Document Type	Frequency	Percentage
Articles	365	49.80
Conference paper	306	41.75
Book Chapter	58	7.91
Book	4	0.55

As seen in Table 1, 49.80 % of the research on using artificial intelligence in the educational assessment process was published as a journal article. The subject was also included 41.75 % as a conference paper.

3.2 Distribution of research according to the languages of the document

When the languages in which the documents were published are examined, almost all (98.80%) were published in English, as seen in Table 2.

Table 2 Distribution of documents according to languages

Document Language	Frequency	Percentage
English	722	98.8
Spanish	5	0.75
Persian	1	0.15
German	1	0.15
Chinese	1	0.15

3.3 Co-occurrence of all keywords

In the studies conducted on the use of artificial intelligence applications in the educational assessment process, a co-occurence analysis was performed to determine the most frequently used keywords. In order to determine the trend keywords related to the subject, a minimum number of co-occurence of a keyword was determined as 5. For each of the 169 keywords repeated 5 times among 3834 keywords, the total link strength found with other keywords was

calculated. The keywords with the largest total link strength were selected and a visual was created. The strongest expressions in terms of total link strength were determined as “artificial intelligence”, “students”, and “learning systems” as a result of the analyses performed with 169 observation units that were seen at least five times and had relationships, a total of 5 clusters, 3651 links and 9397 total link strengths were determined.

Table 3 Most popular keywords and total link strength

Keyword	Occurrences	Total Link Strength
Artificial intelligence	287	1749
Students	121	867
Learning systems	92	645
Human	52	577
Teaching	70	522
Machine learning	67	502
Education	63	486
Humans	37	437
E-learning	62	431
Formative assessment	82	411

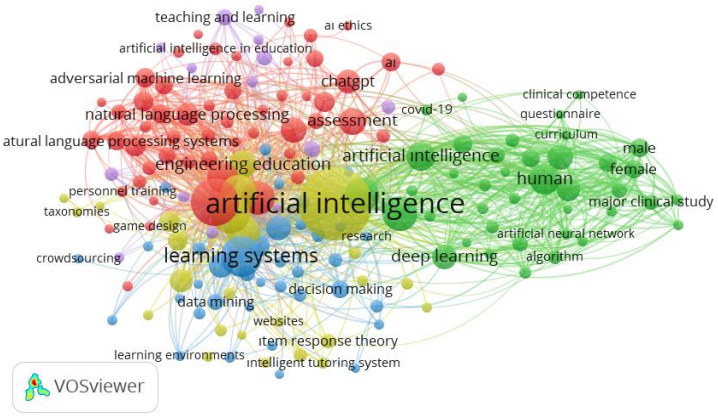

Figure 2. Most used keywords

As seen in Figure 2, 5 different clusters were formed in the mapping of the research conducted on the use of artificial intelligence applications in the educational assessment process. It shows that the subject of “Artificial Intelligence” is the main theme. As seen in the green cluster, human-centered assessment themes are taken as the basis in the research conducted on the use of

artificial intelligence in educational assessment processes. Another striking point in the visual is that although the keyword "Generative Artificial Intelligence" is not repeated as much as the others, the total link strength is 57. This finding shows us that these generative AI technologies are a rising research area in the educational assessment process.

3.4 The most cited articles about educational assessments with AI

The citations to publications show us how much other researchers benefit from that publication. In short, we can say that publications with high citation numbers lead the field. Citation-document analysis was performed with 203 publications that received at least 5 citations in the network mapping analysis. Only 13 documents meet the thresholds. This finding shows us that the studies in this field are not carried out with much collaboration. Figure 3 shows the network visualization between 13 documents. 6 clusters and 18 connections were determined in the mapping.

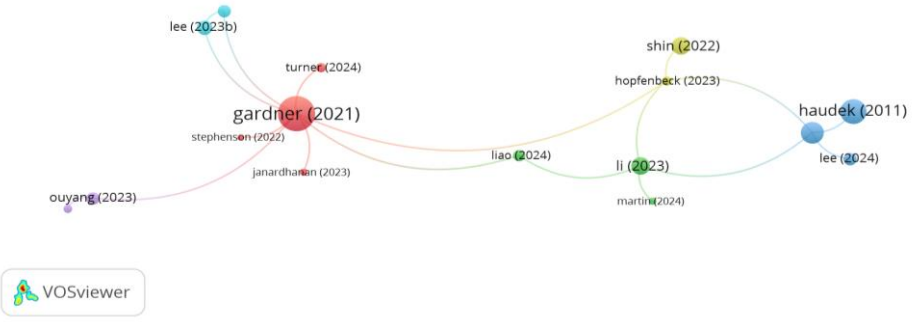

Figure 3 Most cited authors

The 6 most cited studies in the research conducted on the use of artificial intelligence applications in the educational assessment process are presented in the table.

Table 4. The most cited articles

Rank	References	Citation
1	"Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. Sustainability, 15(7), 5614"	216
2	"González-Calatayud, V., Prendes-Espinosa, P. & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. Applied Sciences (Switzerland), 11(12), 5467."	153
3	"Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment:'Breakthrough? Or buncombe and ballyhoo?'. Journal of Computer Assisted	90

	Learning, 37(5), 1207-1216”	
4	“Elbanna, S., & Armstrong, L. (2024). Exploring the integration of ChatGPT in education: adapting for the future. Management & Sustainability: An Arab Review, 3(1), 16-29.”	79
5	“Haudek, K. C., Kaplan, J. J., Knight, J., Long, T., Merrill, J., Munn, A., ... & Urban-Lurain, M. (2011). Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. CBE—Life Sciences Education, 10(2), 149-155.”	46
6	“Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. Journal of Research in Science Teaching, 60(6), 1390-1398.”	33

When the most cited studies are examined, it is seen that the study titled “Chatbots in education and research: A critical examination of ethical implications and solutions.” is in the first place with 216 citations. This study focuses on the ethical problems that will arise from the use of artificial intelligence applications in the evaluation process. This study, which attracted the attention of researchers, was published in the journal Sustainability.

3.5 The most productive countries

In the educational assessment process, a citation-countries analysis was conducted to determine the most productive countries for the use of artificial intelligence applications. When mapping the data, it was determined as a criterion that countries should have at least 5 publications and 15 citations. Although 33 countries met these criteria, 16 countries with links were determined in the mapping. According to this finding, the country mapping according to the number of publications and citations is presented in Figure 4.

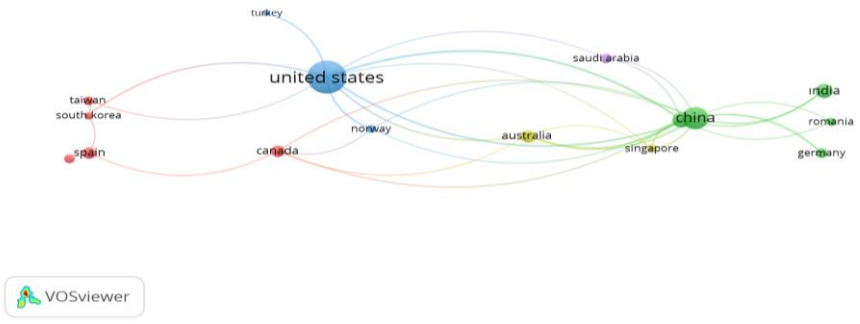

Figure 4. Documents by countries

5 clusters, 33 links and 57 total link strengths were determined. The country with the most citations and publications was determined as US (179 articles, 2132 citations). Then China (85 articles, 484 citations), United Kingdom

(40 articles, 509 citations), Australia (26 articles, 369 citations). The top three countries in terms of total link strength are United Kingdom (22), United State (20) and China (19).

4. Discussion and conclusions

This study aims to examine 762 published studies on the use of artificial intelligence applications in the educational assessment process scanned in the Web of Science and Scopus databases. Since thirty-two documents in the dataset were scanned in both databases, duplicates were deleted and the analyses were performed with 763 documents. While scanning the Web of Science, the keywords "artifical intelligence", "AI", "educational assessment", "learning evaluation", "academic assessment", learning assessment", "digital assessment" and "formative assessment" were preferred in the Subject field and while scanning Scopus, the keywords "artifical intelligence", "AI", "educational assessment", "learning evaluation", "academic assessment", learning assessment", "digital assessment" and "formative assessment" were searched with appropriate conjunctions. The first studies on the use of artificial intelligence applications in the educational assessment process were carried out in 1996. Since 2013, the number of studies has started to increase every year. The increase in deep learning and machine learning algorithms after 2010 may have encouraged their use in education.

According to the results obtained from the analyses carried out in order to determine trending topics in research on the use of artificial intelligence applications in the educational assessment process; The topics of "formative assessment", "machine learning", "e-learning", "learning systems", "human", "chatgpt", and "higher education" have been intensively researched. The increasing use of online learning environments, especially after COVID-19, has caused changes in evaluation processes in education.

When the most cited publications were examined, it was determined that all six publications were journal articles. This result shows us that peer-reviewed journals scanned in WOS and Scopus increase visibility and contribute to science.

When we look at the distribution of publications by country, it was revealed that the publishers were the US, the UK, and China; almost all of the documents were published in English. In the future, researchers conducting research on the use of artificial intelligence in the educational assessment process can plan their research by considering the results obtained from this study. In particular, researchers can understand which studies on the subject should be read first from this study.

The limitation of the study is that the analyses were conducted only on publications obtained from keywords in the WoS and Scopus databases. In future studies, studies can be conducted in internationally respected databases such as IEEE and Ebsco. In addition, according to the results of this study, systematic

analysis can be performed to obtain more detailed information in future studies on the use of artificial intelligence applications in the educational assessment process.

REFERENCES

Baskara, R. & Mukarto, M. (2023) Exploring the Implications of ChatGPT for Language Learning in Higher Education. *Indonesian Journal of English Language Teaching and Applied Linguistics*. 7(2), 343–358. doi: 10.21093/ijeltal.v7i2.1387.

Cope, B., Kalantzis, M. & Searsmith, D. (2021) Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*. 53(12), 1229–1245. doi: 10.1080/00131857.2020.1728732.

Dhara, S., Chatterjee, S., Chaudhuri, R., Goswami, A. & Ghosh, S.K. (2022) Artificial Intelligence in Assessment of Students' Performance. In: *Artificial Intelligence in Higher Education*. CRC Press. pp. 153–167.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W.M. (2021) How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*. 133, 285–296. doi: 10.1016/j.jbusres.2021.04.070.

Van Eck, N.J. & Waltman, L. (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 84(2), 523–538. doi: 10.1007/s11192-009-0146-3.

Ellegaard, O. & Wallin, J.A. (2015) The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*. 105(3), 1809–1831. doi: 10.1007/s11192-015-1645-z.

Fetzer, J.H. (1990) What is Artificial Intelligence?. In: Artificial Intelligence: Its Scope and Limits. *Studies in Cognitive Systems*, vol 4. Springer, Dordrecht. doi: 10.1007/978-94-009-1900-6_1.

Gustilo, L., Ong, E. & Lapinid, M.R. (2024) Algorithmically-driven writing and academic integrity: exploring educators' practices, perceptions, and policies in AI era. *International Journal for Educational Integrity*. 20(1), 3. doi: 10.1007/s40979-024-00153-8.

Hopfenbeck, T. N., Zhang, Z., Sun, S.Z ., Robertson, P. & McGrane, J. A. (2023) Challenges and opportunities for classroom-based formative assessment and AI: a perspective article. *Frontiers in Education*. 8. doi:10.3389/feduc.2023.1270700.

Khasawneh, M.A.S., Ismail, S.M. & Hussen, N. (2024) The blue sky of AI-assisted language assessment: autonomy, academic buoyancy, psychological

well-being, and academic success are involved. *Language Testing in Asia*. 14(1), 47. doi: 10.1186/s40468-024-00318-9.

Krotz, F. & Schelhowe, H. (2020) Beijing consensus on artificial intelligence and education. *merz| medien+ erziehung*. 64(5), 29. https://www.merz-zeitschrift.de/article/download/1945/1782. [Accessed 4th February 2024].

Kültür, Z.Y. & Kutlu, O.M. (2021) *The effect of formative assessment on high school students' mathematics achievement and attitudes*. Çukurova University. https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=3ljk0nsj4lAD6l5jDfUTPw&no=eHH-V-2tHPsdD-zHknjiVw&utm_source. [Accessed 4th February 2024].

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R. & Loyens, S. (2021) Formative assessment as practice: the role of students' motivation. *Assessment & Evaluation in Higher Education*. 46(2), 236–255. doi: 10.1080/02602938.2020.1765228.

Luckin, R. (2017) Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*. 1(3), 0028. doi: 10.1038/s41562-016-0028.

Sbaffi, L. & Zhao, X. (2022) Evaluating a pedagogical approach to promoting academic integrity in higher education: An online induction program. *Frontiers in Psychology*. 13. doi: 10.3389/fpsyg.2022.1009305.

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 and Education: Artificial Intelligence*. 3, 100075. doi: 10.1016/j.caeai.2022.100075.

VOSviewer (2020) *Visualizing scientific landscapes*. https://www.vosviewer.com.

Zhai, X., He, P. & Krajcik, J. (2022) Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*. 59(10), 1765–1794. doi: 10.1002/tea.21773.

Zhai, X. & Nehm, R.H. (2023) AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*. 60(6), 1390–1398. doi: 10.1002/tea.21885.