A SCOPING REVIEW OF RELATED JOURNAL ARTICLES ON ARTIFICIAL INTELLIGENCE CLASSROOM TESTING AND EVALUATION

[1]Goodness Oluwamayokun Opateye, [2] Opateye, Johnson Ayodele, [3] Abubakar Bello,
[1] *Al-Muhibbah Open University*
2 Bicados Street, Kubwa FO1, Abuja
gopateye@gmail.com, +2348066855893

[2]*Department of Educational Foundations*
National Open University of Nigeria
91 Cadastral Zone, University Village, Off Nnamdi Azikiwe Expressway
Jabi, Abuja.
jopateye@noun.edu.ng; +234-8065810979;
[3]*Africa Centre of Excellence on Technology Enhanced Learning*
National Open University of Nigeria
91 Cadastral Zone, University Village, Off Nnamdi Azikiwe Expressway
Jabi, Abuja.
aabello@noun.edu.ng; +234-8060600690

Corresponding Author: Opateye, Johnson Ayodele, *jopateye@noun.edu.ng; +234-8065810979*

Abstract

Artificial Intelligence (AI) has been used to automate processes in the commerce, healthcare, management, finance, and education sectors. Using relevant machine learning codes, AI has impacted curriculum development, instructional delivery, learning, and the use of AI-based instructional aids. This study used a content analytical approach to carry out a scoping review of related articles on artificial intelligence classroom testing and evaluation (AICTE). The study employed a descriptive survey research design. The population consisted of 55 articles published on general education extracted from the Scopus database. Twenty research articles related to artificial intelligence classroom testing and evaluation were extracted to be included in the review. Five research questions were answered. A review information format was developed by the researcher to extract data from the selected journal articles on current focus, research design, locations, analysis, and future focus of artificial intelligence classroom testing and evaluation. Descriptive statistics were used to analyze the data collected. Results showed a very low focus on classroom testing and evaluation research, and most authors used literature reviews and surveys. Most of the articles published on artificial intelligence for testing and evaluation originated in the global north, compared with those from the global south, and the areas of research focused on were general assessment, evaluation, and feedback. Researchers in the global south should carry out more research on the use of artificial intelligence to enhance classroom testing and evaluation, other areas to be researched into our examination administration, question bank, item generation, and results processing, and also apply various research designs to undertake research in these areas. It is therefore concluded that there were not enough articles that focus on Artificial Intelligence classroom testing and evaluation from the global south.

Keywords: Scoping review, Journal articles, Artificial intelligence, Classroom testing, Evaluation and Location

Introduction

The use of technology is a new dimension in all human endeavors, and it is drastically influencing the way humans think, behave, and relate. The capturing of data using computers has been the foundational application of technology, but in recent times its waves have rolled into all facets of nations' economy, health, and education. The newest development is the application of artificial intelligence (AI) with data coding, mining, and cloud computing coming to the fore. Artificial Intelligence has been used for operating processes in commerce, health, management, finance, and agriculture in various dimensions. Utilizing relevant machine learning codes, AI has affected curriculum development, instructional delivery, learning, and the use of instructional aids. Its application in assessment, testing, and evaluation had not been embraced as expected due to the tendency to compromise examination integrity. AI is spreading very fast in other nations of the world, aside from Africa, which is just embracing its adaptability to increase continental awareness, and it can be used to better the black race. Olawale, Aladesanmi & Adediran (2016) enunciated several challenges facing the adoption and implementation of AI in Africa. These challenges include limited digital Infrastructure; many African countries lack the digital infrastructure to support AI development. For example, the estimation of internet penetration is 28% as of 2019. Another challenge is that the lack of quality data is substantial for AI development in Africa. African data environments are still in the early stages of the African data revolution. Many African countries require data gathering mechanisms and data governance structures with better data quality. These issues revolve around the regulatory frameworks that impact the creation and application of AI technology required in Africa. Ethical and legal considerations relating to safety and transparency, informed consent to use data, algorithmic fairness and biases, and data privacy in AI is another problem. The cost of AI technologies, both hardware and software, is still high in many African countries, making it challenging to use AI technologies fully, and many countries in the region suffer from a shortage of AI experts.

Despite the enormous opportunities that AI might afford to support teaching and learning, new ethical implications and risks come with the development of AI applications in higher education. For example, in times of budget cuts, it might be tempting for administrators to replace teaching with profitable automated AI solutions. Faculty members, teaching assistants, student counselors, and administrative staff may fear that intelligent tutors, expert systems, and chatbots will take their jobs. AI has the potential to advance the capabilities of learning analytics, but on the other hand, such systems require huge amounts of data, including confidential information about students and faculty, which raises serious issues of privacy and data protection. (Zawacki-Richter, Marín, Bond & Gouverneur, 2019).

AI has penetrated the field of education to a small extent. Intelligent tutoring systems (ITS) can be utilized to enhance one-to-one personal tutoring. Tutors can make decisions about the learning path of an individual learner and the content to select, give cognitive scaffolding and assistance, to engage the learner in dialogue.

IT has enormous possibilities, especially in large-scale distance teaching institutions, which undertake modules with thousands of students, where human face-to-face tutoring is impossible. Different models are usually joined together in ITS. These include the student model (e.g. information about the student's knowledge level, cognitive ability, learning motivation, and learning styles); the teacher model (e.g. analysis of the current state of students, selecting teaching strategies and methods, providing help and guidance), the domain model for knowledge representation of both students and teachers and the diagnosis model that deals with the evaluation of errors and defects based on domain model (Huang & Chen, 2016). Learning with AI has been the focus of the AI in education (AIED) academic research field since at least the 1980s. The International Journal of Artificial Intelligence in Education was first published in 1989, while the International AI in Education Society (IAIED) was established in 1993. However, the origins of AIED are in the 1930s, when Artificial Intelligence and education saw the development of "teaching machines" and their twin promises of personalised learning and saving teacher time were emphasized by Watters (2021). Holmes, Persson, Chounta, Wasson, and Dimitrova (2022) pointed out that learning with AI might be further grouped into learner-supporting AI, teacher-supporting AI, and institution-supporting AI. Learning with AI involves the use of AI-driven tools in teaching and learning. It includes: the use of AI to support learners directly, involving tools such as those known as intelligent tutoring systems, dialogue-based tutoring systems, exploratory learning environments, automatic writing evaluation, learning network orchestrators, and chatbots. Other involvements are AI to support learners with disabilities; the use of AI to support administrative systems (such as recruitment, timetabling, and learning management); and the use of AI to boost teachers' delivery directly.

A scoping review is a type of research synthesis that aims to map out the existing literature on a particular topic or research question, identify gaps in the literature, and highlight areas for future research. Studies have shown that AI applications can perform assessment and evaluation jobs at very high accuracy and efficiency levels. However, due to the need to calibrate and train the systems (supervised machine learning), they are more applicable to courses or programs with large student numbers. Articles focusing on assessment and evaluation applications of AI at the teaching and learning level were classified into four sub-categories: automated grading, feedback, evaluation of student understanding, engagement and academic integrity, and evaluation of teaching. Some articles reported on student-facing tools that evaluate student understanding of concepts (Zhu, Marquez, & Yoo, 2015) and Gelegenis (2015), while Hussain et al. (2018) used machine learning algorithms to evaluate student engagement in social science and involved final results, assessment scores and the number of clicks that students make in the virtual learning environment. Munn, Micah, Peters, Tufanaru, McArthur & Aromataris (2018) pointed out that, scoping reviews are perfect instruments to ascertain the scope or coverage of a body of literature on a given topic and give a clear indication of the volume of literature and studies available as well as an overview (broad

or detailed) of its focus. Scoping reviews are useful for examining emerging evidence when it is still unclear what other, more specific questions can be posed and valuably addressed by a more precise systematic review (Armstrong, Hall, Doyle & Waters, 2011). They can give indications on the types of evidence that address and inform practice in the field and the way the research has been conducted. The general purpose of undertaking scoping reviews is to identify and map the available evidence. Arskey and O'Malley (2005) described a structure for scoping reviews, provided four specific reasons why a scoping review may be conducted. The scoping review may be conducted for the following purposes:

- To identify the types of available evidence in a given field
- To clarify key concepts/ definitions in the literature
- To examine how research is conducted on a certain topic or field
- To identify key characteristics or factors related to a concept
- As a precursor to a systematic review
- To identify and analyse knowledge gaps (Munn, et al (2018).

Scoping reviews do not aim to produce a critically appraised and synthesized result/answer to a particular question, but rather aim to provide an overview or map of the evidence. Due to this, an assessment of methodological limitations or risk of bias of the evidence included within a scoping review is generally not performed unless there is a specific requirement due to the nature of the scoping review aim (Peters, Godfrey, Khalil, McInerney, Parker & Soares, 2015). According to Steinberg, Greenfield, Mancher, Wolman, and Graham (2011) assertion, systematic reviews can be broadly defined as a type of research synthesis that are conducted by review groups with specialized skills, who set out to identify and retrieve international evidence that is relevant to a particular question or questions and to appraise and synthesize the results of this search to inform practice, policy and in some cases, further research.

Articles that used automated grading systems emanated from a range of specialisations, for example, Biology, Medicine, Business Studies, and English as a Second Language. The benefits of using algorithms that find patterns in text responses, however, have been found to lead to encouraging more revisions by students (Ma & Slater, 2015 and to move away from merely measuring student knowledge and abilities by multiple-choice tests (Nehm, Ha, & Mayfield, 2012). On the evaluation of student understanding, engagement, and academic integrity, Jain, Gurupur, Schroeder, and Faulkenberry (2014) stated that student-facing tools should evaluate student understanding of concepts. The evaluation of teaching and data mining algorithms to evaluate lecturer performance through course evaluations (Agaoglu (2016); Ahmad & Rashid (2016) using four different classification techniques, where many questions in the evaluation questionnaire were irrelevant. The application of an algorithm to evaluate the impact of teaching methods in a differential equations class found that online homework with immediate feedback was more effective than clickers.

Without any doubt, assessment is a vital part of learning during the teaching process. Assessment is carried out to ascertain what learners have gained in the course of study. This produces information that is required for several purposes: for the grading of students, for individual attention required by any student, for creating merit, and so on. Another strength of assessing students and their learning is that it tells how to mold the learning methodologies and instructional materials. Several forms of tests can be conducted in a classroom. Such tests could be multiple choice, essay, matching, fill-in-the-blank, or even oral. The core reason behind taking these tests is that they give an insight into the performance of every student in the classroom. Evaluation is a result of tests and assessments that are carried out in the classroom. Evaluation could be formative or summative. It is done to decide on the outcomes of learning that have taken place. Researching classroom testing and evaluation results into the publications of reports in relevant journals to disseminate the findings.

The Problem

Artificial intelligence in education is used in education for different purposes. It has been integrated into intelligent tutoring, automated grading systems, and instructional technologies. AI supports students' personalization and collaboration, learning scheduling and adaptive feedback, prediction of learners' attrition, students' profiles, and monitoring of students' progress. Despite these applications, the use of artificial intelligence for measurement, testing, and evaluation has not been so focused on. Most studies in AI focused on general education and instructional delivery with little attention to classroom testing and evaluation. Also, the compromise of assessment integrity posed a lot of challenges to using AI to conduct examinations and other evaluation processes. Scoping reviews have not been sufficiently carried out on how AI impacts classroom testing and evaluation. Therefore, the study carried out a scoping review of related journal articles on artificial intelligence classroom testing and evaluation in the areas of focus, research designs, geographical location, and use of analytical tools.

Research Questions

1. What is the general focus of most research articles published on artificial intelligence classroom testing and evaluation?
2. What are the areas of future research available for artificial intelligence-enabled classroom testing and evaluation?
3. What research designs do authors use as a framework in AI classroom testing and evaluation studies?
4. What are the geographical locations of the authors of selected journal articles on artificial intelligence classroom testing and evaluation?
5. In what ways were various statistical and computational tools used to analyze data gathered in the artificial intelligence classroom testing and evaluation-related journal articles?

Methodology

This research is based on the use of a scoping review to ascertain key elements of the journal articles on artificial intelligence classroom assessment, testing, and evaluation. The process was undertaken based on the prescribed criteria of inclusion and extraction of research articles that focused on these areas. For a research paper to be used, it must relate artificial intelligence to classroom assessment, testing, and evaluation. The year of consideration is peer-reviewed articles published between 2000 and 2023. Research articles in conference proceedings, book chapters, or grey literature, or those articles not published in journals, are excluded. The articles are written in the English language. The articles were empirical or theoretical.

Based on these criteria, the Google search engine was applied to get a database of Scopus published articles on the web especially thus:

- Artificial Intelligence and Assessment Journal Articles (AIAJA) 2000 - 2023
- Artificial intelligence and Classroom testing Journal Articles (AICTJA) 2000 - 2023
- Artificial Intelligence and Evaluation Journal Articles (AIEJA) 2000 – 2023
- Artificial Intelligence and Measurement Journal Articles (AIMJA) 2000 - 2023

Fifty-five articles were identified in the first instance from the database.

The flowchart for the selection of articles is shown in Figure 1 below.

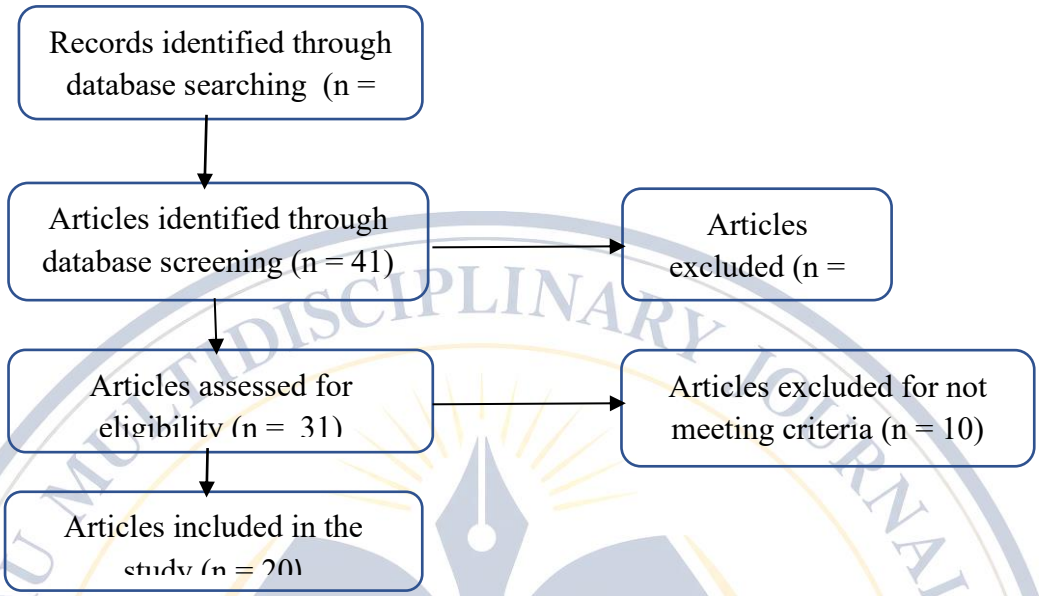

Figure 1: Flowchart of articles' selection process

After the selection of 20 articles based on the criteria, data required to answer the research questions were extracted into the Data Extraction Format (DEF) that had nine columns of article number, title of paper, author's name (AN), author's area of specialisation (AS), area of focus (AF), research design (RD), name of journal (NJ), data analysis tool (DAT), and authors' geographical location (AGL).

Table 1: Data Extraction Format (DEF)

Article Number	Title of Articles	Authors' Name	Academic area of Specialisation	Areas of Focus	Research Designs	Name of Journals	Data Analysis Tool	Authors' Geographical Location

The information was extracted by examining the title, abstract, keywords and body of each article.

Each of the articles were scrutinized to identify the authors' name, area of specialization, area of focus (AF), research design (RD), name of journals (NJ), data analysis tool (DAT) and author's geographical location AGL) and filled into DEF as shown on Table 2

Table 2: Data Extraction from articles

Article Number	Articles	AN	AS	AF	RD	NJ	DAT	AGL
01	An AI – Application – Oriented in In-class Teaching Evaluation Model by Statistical Modelling and Essemble	Jungi *et al* (2021)	Technology	Teaching and evaluation	Input and Output Evaluation model	Sensors 21(1) / Online	Statistical modelling	Italy
02	Artificial Intelligence in Assessment of Students' Performance	Dhara *et al* (2022)	Technology (Data Mining)	Assessment of students' performance	Literature Review	Taylor & Francis (Chapter in Book) Artificial Intelligence in Higher Education	Theoretical analysis	Singapore
03	Implementation of Artificial Intelligence assessment in Engineering Laboratory Education	Samarakou *et al* (2014)	Energy Technology/Computer Sciences	Practical Assessment	Survey	Conference proceedings	Frequency and percentages	Greece
04	Assessment in the age of artificial intelligence	Swiecki *et al* (2022)	Educational Psychology	Assessment	Literature review	ELSEVIER	Theoretical analysis	Australia
05	Artificial intelligence in educational assessment: Breakthrough? Or buncombe and ballyhoo?	Gardner *et al* (2021)	Assessment and Information Technology	Educational assessment, Adaptive testing, automated essay scoring	Literature review	Journal of Computer Assisted Learning WILEY 2021	Content analysis	United Kingdom
06	Machine learning based feedback on textual student answers in large courses	Bernius *et al* (2022)	Computer Science Education and Automated assessment	Automatic assessment grading, Assessment feedback, Assessment support	Modeling and evaluative	Computer and Education : Artificial Intelligence 3, (2022), 100081 Elsevier	Frequency and percentages	Germany
07	Artificial intelligence for assessment and feedback to enhance student success in	Hooda *et al* (2022)	English Language and Curriculum development	Assessment analytics, feedback	Explorative and comparative study	Mathematical problems in engineering volume 2022	Comparative analysis with Random forest and	India

	higher education						Decision trees	
08	Machine learning algorithm to predict students' performance: A systematic review	Sandra *et al* (2021)	Psychology, computer science	Students' performance	Literature review	TEM Journal 10(4), 2021	Algorithms Logistic regression and decision tree	Indonesia
09	Evaluating artificial intelligence in education for next generation	Joshi *et al* (2022)	Technology and engineering	Evaluation	Literature review	Journal of physics, Conference series 1714, 012039	Frequency and percentages drawn on Bar charts	India
10	AI-Driven Assessment of students: Current and Research Trends	Sanchez-Prieto *et al* (2020)	Social education, e-learning and mobile learning	Students' assessment	Literature review and Mapping	Learning and collaboration technologies, designs and experience	Frequency counts and charts	Spain
11	Research on educational evaluation in the era of artificial intelligence	Wang *et al* (	Language	Educational Evaluation and assessment	Literature review	International Journal of Frontiers in Sociology, 2 (3) 2021	Theoretical	China
12	Artificial intelligence for student assessment: A systematic review	Gonzalez-Calatayud et al (2021)	Mathematics, computer sciences	Student assessment, grading	Systematic review, meta-analysis	Applied sciences, 11, 2021	Frequency counts	Spain
13	Artificial Intelligence-Based student Assessment and recommendations system for e-learning in big data	Bagunaid *et al* (2022)	Computer science, Information technology	Student assessment	AISAR system,	Sustainability, 14, 10551, (2022)	Recurrent neutral network (RNN) and Density-based spatial clustering algorithm (DBSCAN)	Australia
14	A domain question answering algorithm based on the contrastive language pre-	Zang Zuoxing *et al* (2023)		Questions and answers	CLIP Mechanism design	Journal of computer and education 11(5)	Intelligence question-answering algorithm	China

	training mechanism							
15	A systematic review of artificial intelligence impact assessment	Stahl, Bernd Carsten *et al* (2023)	Computer science	Impact Assessment	Qualitative	Artificial intelligence review	Qualitative analysis	UK
16	Artificial Intelligence: Transforming the future of feedback in education	Wonguorachan Tarid *et al* (2023)	Educational psychology	Educational assessment, Feedback	Qualitative Literature review	Journal of applied testing technology, 23(1), 2022	Qualitative analysis	Canada
17	Rise of the machines? The evolving role of AI technologies in high-stakes assessment	Richardson Mary *et al* (2021)	Education (Assessment)	High stakes assessment	Literature review	London Review of education, 19(1), 2021	Theoretical paper	London
18	Validity arguments meet artificial intelligence in innovative educational assessments.: A discussion and look forward	Dorsey David W. (2022)	Quantitative psychology	Educational assessment	Review of literature	Journal of educational measurement 59(3), 2022	Theoretical paper	USA
19	Optimising implementation of Artificial Intelligence-based automated scoring: An Evidence centre approach for designing assessments for AI-Based scoring	Ercikan Kadriye *et al* (2022)	Language	Assessment scoring	Literature review	Journal of educational measurement 60(1), 2022	Content analysis	Canada
20	Educational assessment using intelligence systems	Shute, Valerie J. *et al* (2008)	Educational psychology	Educational assessment	Literature review	ETS Research Report series, Volume 2008 (2)	Theoretical	USA

Tallies were used to obtain the frequency of occurrence of variables on which the research questions were focused. Frequency counts, percentages, and charts were used to analyse the data extracted.

Results and Discussions

Research Question One: What is the general focus of research articles published on artificial intelligence classroom testing and evaluation?

Figure 2: Frequency of Research AI Classroom and Evaluation Articles' Areas of Focus

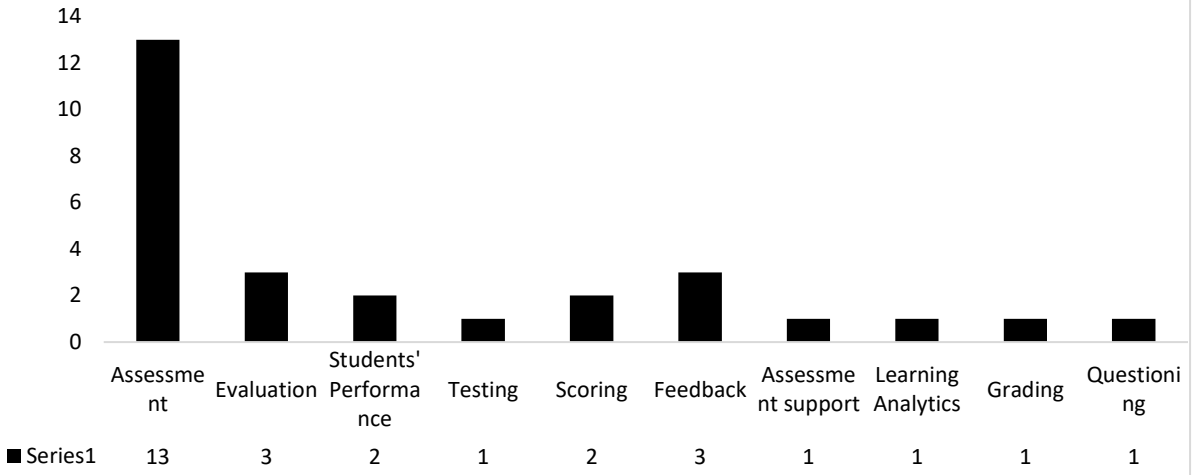

Figure 1shows that the assessment fixture was as identified 13 times in the reviewed articles on AI classroom testing and evaluation, 3 of the articles centred on evaluation, 2 examined AI and students' performance, 1 article focused on AI, and 1 article focused on classroom testing and AI. Also, 2 focused on scoring, 3 examined AI and assessment feedback as assessment support, learning analytics, grading, and questioning had 1 article each focused on them. It is therefore deduced that most of the articles reviewed focused on general assessment, followed by evaluation and assessment feedback with AI. The limited area of focus of research on AI empowered research had not been directed towards relevant aspects of classroom testing and evaluation. This is due to the fact that application of AI testing and evaluation is still being seen as avenue to compromise assessment integrity.

Research Question Two: What are the areas of future researches available for artificial intelligence empowered classroom testing and evaluation?

Based on the results shown on Figure 2, areas where current AI classroom testing and evaluation researches focused on were enunciated. Other areas of testing and evaluation articles which future research need to focus on with artificial intelligence are item development, item analysis, assessment reporting, test administration, electronic examination, assessment record keeping, scores interpretation and measurement of affective and psychomotor behaviours of learners.

Research Question Three: What research designs do authors use as a framework in AI classroom testing and evaluation studies?

Table 3: Frequency of research articles' research design

Articles' Research Design	Frequency
Survey	4
Literature review (Theoretical papers)	9
Modelling	1
Explorative and Comparative	1
Mapping	1
Meta-analysis	3
Qualitative	1
Total	20

Figure 3 : Frequency of Research Designs in Testing and Evaluation AI Related Articles

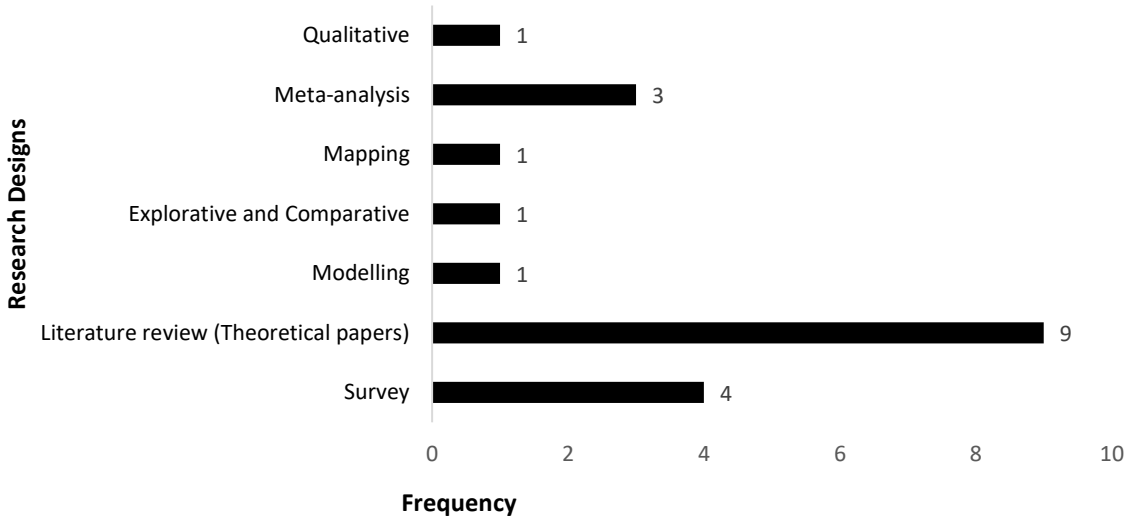

Table 3 shows the extent to which authors of artificial intelligence classroom testing and evaluation journal articles utilised various research designs as frameworks for the research articles. It is discovered from the table that 1 article used a qualitative research design, mapping, explorative /comparative, and modelling. The analysis of research designs showed that literature reviews dominated the field, with nine articles using this approach, compared to four articles that utilized a survey research design. Therefore, it could be deduced that articles related to artificial intelligence classroom testing and evaluation employed literature reviews mostly, and this was followed by a survey research design. Most of the papers were theoretical in

nature due to the fact that application of artificial intelligence to education generally and to testing and evaluation specifically is just evolving, and researchers found it more convenient to review existing literature to draw out issues related to these areas. Research designs that are specifically related to computer science are to be employed and merged with those that are generally used in educational research.

Research Question Four: What are the geographical locations of authors of selected journal articles on artificial intelligence classroom testing and evaluation?

Table 4: Frequency of geographical locations of journal article authors

Articles Authors' Country	Frequency (%)
Italy	1(5.0)
Singapore	1(5.0)
Greece	1(5.0)
Australia	2(10,0)
United Kingdom	3(15.0)
Germany	1(5.0)
India	2(10.0)
Indonesia	1(5.0)
Spain	2(10.0)
China	2(10.0)
Canada	2(10.0)
America	2(10.0)

Table 4 shows the country affiliations of authors. One article each, representing 5.0% of the sample, originated from Italy, Singapore, Greece, and Germany. Two articles each, representing 10.0% of the sample, originated from Australia, India, Spain, China, Canada, and the United States. The highest number of articles, three (15.00%), originated from the United Kingdom. This indicates that most articles on AI classroom testing and evaluation came from Europe and Asia. Notably, no African research article was found on how AI could facilitate effective classroom testing and evaluation. This finding still show low research works on artificial intelligence classroom testing and evaluation in the Africa as compared to Europe and Asia, where technology is advanced and AIs almost been applied to all human endeavours.

Research Question 5: How were various statistical and computational tools used to analyse data gathered in the Artificial Intelligence classroom testing and evaluation related journal articles?

Table 5: Frequency and percentages of statistical and computational tools in artificial intelligence, classroom testing, and evaluation-related journal articles

Data Analysis Tool	Frequency (%)
Modelling	1(5.0)
Frequency and Percentages	8(40.0)

Content analysis	7(35.0)
Random forest and Decision Tree	1(5.0)
Logistic regression	1(5.0)
Recurrent Neutral Network	1(5.0)
Questions and Answer Algorithm	1(5.0)

Table 5 reveals that 1(5.0%) of the artificial intelligence classroom testing and evaluation related journal articles employed modelling analysis, while 8(40.0%) used frequency and percentage to analyse the data. Content analysis was undertaken on 7(35.0%) of the articles as 1(5.0%) used random effect and decision tree. Logistic regression, recurrent neutral network and questions and answers algorithm had 1(5.0%) of the articles each as their analysis tools. It is depicted that most of the artificial intelligence classroom testing and evaluation-related journal articles utilised frequencies and percentages, followed by content analysis. The major works that have been carried out on AICTE used frequency and percentages more because most of the published articles itemise the key words that are countable for easy transcription to percentages. AICTE basically is computer-based, and it is advantageous to use algorithmic analytical models for analysing the data captured for each of the studies.

Limitations

Though this scoping review was undertaken as rigorously as possible, each review is narrowed by its search strategy. In order to obtain the articles, the search was limited to the Scopus journals database, which was undertaken from 2001 to 2023. The language of the publications of the articles was only English. Future research could consider using a larger number of databases, publication types, and publication languages in order to widen the scope of the review.

Conclusion

In this paper, we have explored a scoping review of related journal articles on artificial intelligence classroom testing and evaluation. Most of the papers reviewed focused on general assessment, followed by evaluation and assessment feedback with AI. Other areas of testing and evaluation articles that future research needs to focus on with artificial intelligence are item development, item analysis, assessment reporting, test administration, electronic examination, assessment record keeping, and score interpretation. Articles related to artificial intelligence classroom testing and evaluation employed literature reviews mostly, and this was followed by a survey research design. The majority of the articles published on AI classroom testing and evaluation came from Europe and Asia, and few African research articles were found on how AI could facilitate effective classroom testing and evaluation.

Recommendations

Given the findings of this study, the following were recommended:

1. Test and measurement experts should develop an interest in the employment of artificial intelligence tools for test item generation, examination administration, scoring, and item analysis.
2. Scoping reviews on knowledge and application should be carried out in all aspects of test and measurement using technology.
3. Researchers from the global south should focus more on studies that would bring greater focus on the use of artificial intelligence for assessment and evaluation.
4. Investigators should employ more robust analytical and statistical tools to analyse artificial intelligence –based test and evaluation data.
5. Research designs that take cognizance of the nature of artificial intelligence and classroom test and evaluation should be applied to have a solid foundation on which the methodology is based.

References

Agaoglu, M.N. (2016). Can (should) theories of crowding be unified? *Journal of Vision,* 16 (10), 1-22.

Arksey, H. & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*. 8(1), 19–32.

Armstrong R., Hall B.J., Doyle J. & Waters E. (2011). Scoping the scope of a Cochrane review. *Journal of Public Health*, 33(1),147–168.

Banguard, W. Chilamkurti, N. & Veeraraghavan, P. (2022). AISAR: Artificial intelligence-based students' assessment and recommendation system for e-learning in big data. *Sustainability,* 14, 10551

Bernius, J.P., Krusche, S. & Bruegge, B. (2022). Machine learning, based on feedback on textual students' answers in large courses. *Computers and Education: Artificial Intelligence,* 3, 10081.

Ceilk, I., Dindar, M., Muukkonen, H. & Jarvela, S. (2022). The promises of and challenges of Artificial intelligence for teachers: a systematic review. *Technical Trends*. 66, 616-630.

Chatterjee, S. & Chaudhuri, R. (2022). Artificial intelligence in the assessment of students' performance. *Artificial Intelligence in Higher Education,* Taylor & Francis. 1-17

Dhara, S., Chatterjee, S. & Chaudhuri, R. (2022). Artificial intelligence in the assessment of students' performance. *Artificial Intelligence in Higher Education,* Taylor & Francis. 1-17

Gardner, J., O'leary, M. & Yuan, L. (2020). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo? *Journal of Computer-assisted Learning,* 37(5), 1207-1216.

Guo, J., Bai, L., Yu, Z., Zhao, Z. & Vvan, B. (2021). An AI-Application-oriented in-class teaching evaluation model by using statistical modeling and ensemble learning. *Sensors (Basel),* 21(1), 241.

Holmes, W., Persson, J., Chounta, I., Wasson, B. & Dimitrova, V. (2022). Artificial intelligence and education. France: The Council of Europe Publication Department.

Hooda, M., Rana, C., Dahiya, O., Bizwan, A. & Hossain, S. (2022). *Artificial intelligence for assessment and feedback to enhance student success in higher education. Mathematical Problems in Engineering. Article ID 5215722, 1234-1237.*

Huang, J., & Chen, Z. (2016). The research and design of web-based intelligent tutoring system. *International Journal Of Multimedia and Ubiquitous Engineering*, *11*(6), 337 - 348. https://doi.org/10.14257/ijmue.2016.11.6.30.

Hussain, S.T., Lei, S., Akram, T., Haider, M.J., Hussain, S.H. and Ali, M. (2018). Kurt Lewin's process model for organizational Change: The role of leadership and employee involvement: A critical review. *Journal of Innovation and Knowledge*, 3, 123-127.

Jain, G.P., Gurupur, V.P., Schroeder, J. & Faulkenberry, E.D. (2014) Artificial Intelligence-Based Student learning evaluation: A Concept map-based approach for analyzing a student's understanding of a topic. *IEEE Transactions on Learning Technologies*, 7(3), 267-279.

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *American Journal of Distance Education*, *9*(2), 7–25. https://doi.org/10.1080/08923649509526885.

Joshi, S., Rambola, R.K., & Churi, P. (2020). Evaluating artificial intelligence in education for next generation. *Journal of Physics: Conference Series,* 1714, 012039.

Ma, H. & Tammy, S. (2015). Using the Developmental Path of Cause to Bridge the Gap between AWE Scores and Writing Teachers' Evaluations. *Writing & Pedagogy*, 7, 623-634.

Munn, Z., Micah D. J., Peters, C. S., Tufanaru, C, McArthur, A. & Aromataris, E. (2018). Systematic review or scoping review? *BMC Medical Research Methodology*, 18(143), 1-7. https://doi.org/10.1186/s12874-018-0611-x

Nehm, R.H., Ha, M. & Mayfield, E. (2012). Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations. *Journal of Science Education Technology.* DOI 10.1007/s10956-011-9300-9.

Olaf Zawacki-Richter, Victoria I. Marín, Melissa Bond & Franziska Gouverneur (2019). *International Journal of Educational Technology in Higher Education*, 16, 547-572.

Olawale, T., Aladesanmi, O. & Adediran, A. (2016). AI Readiness in Africa. *CSEA participated in the Policy Research Methodology and Analysis Training at the West Africa Civil Society Institute (WACSI), Accra Ghana on September 12-15, 2016.*

Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthccare,* 13(3), 141–156.

Samarakou, M., Fylladitakis, E., Prentakis, P. & Athineos, S. (2014). Implementation of artificial intelligence assessment in engineering laboratory education. *International Journal of e-Learning,* 5(4), 2567-2578.

Sanchez-Prieto, J., Cruz-Benito, J. Theron, R. & Gracvia-Penalvo, F.J. (2020). AI-driven assessment of students: Current uses and research trends. In P. Zaphiris and A. Ioannous (Eds.), *Learning and Collaboration Technologies, Design, Experiences.* Springer, 292-302.

Steinberg, E. Greenfield S, Mancher M, Wolman DM, Graham R. (2011). *Clinical practice guidelines we can trust.* Institute of Medicine. Washington, DC: National Academies Press; 2011.

Sweicki, Z., Khosravi, H. , Milligan, S., Selwyn, N. & Gasevic, D. (2022). Assessment in the age of Artificial intelligence. *Computers and Education: Artificial Intelligence,* 3, 100075.

Tarik A. Rashid,T.A. & Ahmad, H.A. (2016). Lecturer performance system using neural network with Particle Swarm Optimization. *Computer applications in engineering education*, 24(4), 629-638

Wang, L. & Wang, J. (2020). Research on educational evaluation in the era of artificial intelligence. *International Journal of Frontiers in Sociology,* 2(3), 40-49.

Watters, A. (2021). *Teaching machines: The history of personalized learning*. USA: MIT Press

Zhu, W. & Marquez, A. & Yoo, J. (2015). *Development of engineering economics career mobile app.* Paper presented at the 2015 ASEE Annual Conference & Exposition, Seattle, Washington. 10.18260/p.23855.