

Exploring the potential of artificial intelligence tools in educational measurement and assessment

Valentine Joseph Owan^{1,2*} , Kinsgley Bekom Abang¹ , Delight Omoji Idika¹ , Eugene Onor Etta³ , Bassey Asuquo Bassey¹ 

¹ Department of Educational Foundations, University of Calabar, Calabar, Cross River State, NIGERIA

² Ultimate Research Network, Calabar, Cross River State, NIGERIA

³ Department of Public Administration, University of Calabar, Calabar, Cross River State, NIGERIA

Received 08 May 2023 • Accepted 14 June 2023

Abstract

Artificial intelligence (AI) is transforming various industries, and education is no exception. Rapid advancements in AI technology have become essential for educators and educational assessment professionals to enhance teaching and learning experiences. AI-powered educational assessment tools provide numerous benefits, including improving the accuracy and efficiency of assessments, generating personalized feedback for students, and enabling teachers to adapt their teaching strategies to meet the unique needs of each student. Therefore, AI has the potential to revolutionize the way education is delivered and assessed, ultimately leading to better educational outcomes for students. This paper explores the various applications of AI tools in educational measurement and assessment. Specifically, it discusses the integration of large language AI models in classroom assessment, in specific areas such as test purpose determination and specification, developing, test blueprint, test item generation/development, preparation of test instructions, item assembly/selection, test administration, test scoring, interpretation of test results, test analysis/appraisal, and reporting. It analyses the role of teachers in AI-based assessment and the challenges of using AI-powered tools in educational assessment. Finally, the paper presents strategies to address these challenges and enhance the effectiveness of AI in educational assessment. In conclusion, using AI in educational assessment has benefits and limitations. As such, educators, policymakers, and stakeholders must work together to develop strategies that maximize the benefits of AI in educational assessment while mitigating the associated risks. The application of AI in educational assessment can ultimately transform education, improve learning outcomes, and equip students with the skills needed to succeed in the 21st century.

Keywords: ChatGPT, educational assessment, reliability, skills, technology, validity

INTRODUCTION

Artificial intelligence (AI) is a rapidly evolving field of technology that involves the development of intelligent machines that can perform tasks that typically require human intelligence, such as understanding natural language, recognizing patterns, and making decisions based on data. AI is the ability of machines to adapt to new and emerging situations, problem-solve, answer questions, create plans, and perform other

intelligent functions typically associated with human beings. AI refers to the field of computer science that involves creating computer programs capable of imitating intelligent behavior and ideally enhancing human-like abilities (Naqvi, 2020). AI, a swiftly expanding discipline, encompasses the development of intelligent robots capable of emulating human thought processes and actions, finding utility in diverse areas such as medical diagnosis, self-driving cars, and education (Wardat et al., 2023). AI-powered tools and

© 2023 by the authors; licensee Modestum. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

✉ owanvalentine@gmail.com (*Correspondence) ✉ bekomabang@gmail.com ✉ delightoidika@yahoo.com

✉ eugeneetta@gmail.com ✉ babssey67@gmail.com

Contribution to the literature

- The article expands the understanding of how AI can transform the assessment process in education by examining the use of AI in different stages of test development, such as test purpose determination, test item generation, test administration, scoring, interpretation of test results, and reporting.
- The article contributes to the literature by emphasizing the collaborative relationship between AI and teachers in enhancing educational assessment. Specifically, it sheds light on the changing role of teachers in the context of AI-powered assessment, which can be useful for teachers to adapt their teaching strategies and cater for individual student needs.
- The article adds depth to the existing literature on the complexities involved in integrating AI into education by discussing potential limitations and risks related to data privacy, bias, and the need for human oversight. The work presents practical strategies to address the challenges and optimize the use of AI in educational assessment. These strategies offer valuable insights for educators, policymakers, and stakeholders seeking to maximize the benefits of AI while mitigating the associated risks.

applications are now being used in many industries (Suh & Ahn, 2022), including education, to enhance the quality of services provided to students and teachers. AI tools such as Bing and ChatGPT have been referred to as objects individuals can think with, especially in the teaching-learning situation for learners to enhance their ability to think critically and reflectively, foster creativity, acquire problem-solving skills, and grasp concepts effectively (Vasconcelos et al., 2023).

The integration of AI in teaching effectively realized learner-centered learning (Huang, 2018). AI-powered tools and applications improve educational measurement, including testing, assessment, and evaluation. These tools can provide educators with valuable insights into student performance, learning outcomes, and instructional effectiveness. For example, AI-powered assessment tools can analyze student responses to assignments and provide personalized feedback to help students identify areas of strengths and weaknesses (Nazaretsky et al., 2022).

These tools can also provide teachers with insights into the effectiveness of their instruction and identify areas, where they may need to adjust their teaching strategies. In addition, AI-powered tools can help automate many aspects of the assessment process, saving time and reducing the burden on teachers. For example, AI-powered grading tools can analyze students' essays and provide feedback on grammar, structure, and content, reducing teachers' time grading assignments (Huang et al., 2023b).

AI-powered tools can also help identify students at risk of falling behind or benefit from additional assistance or remediation (Delgado et al., 2020). These tools can analyze students' data, such as test scores and attendance records, and identify patterns that may indicate a need for intervention. This can help teachers to provide targeted support to students who need it most. Therefore, the development of AI-powered tools and applications has revolutionized the field of education by providing educators with valuable insights into student performance, learning outcomes, and

instructional effectiveness. As this technology continues to evolve, it has the potential to transform education by providing personalized, data-driven instruction to students and enabling teachers to optimize their teaching strategies to improve student outcomes. Nevertheless, the use of AI in teaching and learning has its limitations. For example, there are worries that students might practice copying and pasting text from sources without undergoing critical analysis, neglecting to attribute the work to the original sources, resulting in plagiarism (Halaweh, 2023).

Moreover, the concern regarding user privacy is another significant matter (Elliott & Soifer, 2022; Hu & Min, 2023; Saura et al., 2022). Furthermore, concerns have been raised regarding plagiarism detection in content produced by ChatGPT, as well as the challenge of distinguishing between factual and fictional text generated (Chatterjee & Dethlefs, 2023; King & ChatGPT, 2023).

Instructors are increasingly concerned that students may utilize ChatGPT to complete their written assignments, as plagiarism detection tools can generate reports within seconds without being detected (Halaweh, 2023). Other issues border on detecting text written by AI tools using other AI tools. More specifically, questions have been raised about the accuracy with which AI tools can discriminate between text written by humans and one written by AI tools. This is because there are currently no methods available to distinguish between content generated by such a tool and those written by humans (Tovani-Palone, 2023).

What if we wrote the text, but an AI detector says it is machine (AI) generated? Proponents have also argued that AI has come to stay and have warned instructors to seek ways of guiding students on the ethical use of AI tools to maximize their benefits (e.g., Halaweh, 2023; Javaid et al., 2023; Mohamed et al., 2022; Rudolph et al., 2023).

Following the footprints of these scholars, the current paper discusses the penetration of AI and its usefulness specifically in educational measurement, a sub-

discipline in the field of education concerned with classroom assessment. This paper also discusses some matters arising from integrating AI in educational measurement, assessment and evaluation.

Applications of Artificial Intelligence Tools in Educational Measurement and Assessment

AI has been transforming education in recent years. This increasing significance of AI has garnered the interest of numerous scholars actively exploring diverse methods to incorporate various AI tools within the classroom environment (Halaweh, 2023; Mena-Guacas et al., 2023; Papapicco, 2020). With the increasing availability of data and the growing sophistication of machine learning algorithms, AI has the potential to revolutionize the way we learn, teach, and assess student progress. Several benefits of using AI in education have been advanced. For example, Adiguzel et al. (2023) presented some benefits of using AI for administrators, teachers and learners. AI has been found to play a crucial role in motivating students (Lin et al., 2021; Xia et al., 2022), raising their engagement levels (Huang et al., 2023a; Nazari et al., 2021), learning interest (Hou et al., 2022), learners' interaction (Karsenti, 2019), anxiety reduction (Hawes & Arya, 2023; Ren, 2020), prediction of students' future outcomes (Kumar, 2019; Luo et al., 2022) and academic performance (Khan et al., 2021). This section discusses how AI tools can be used in educational measurement and assessment. The following are specific ways AI can be applied in educational assessment.

1. **Personalized learning:** Tools and systems based on AI have the potential to customize learning experiences, boost the productivity of teachers, and heighten student engagement (Mena-Guacas et al., 2023). AI can help create personalized learning plans for individual students based on their learning progress, strengths, and weaknesses. AI can identify students' learning needs by analyzing data from multiple sources such as assessments, homework, and quizzes and provide targeted feedback. Besides preparing lesson contents and learning experiences, AI tools can also be used for individualized instruction (Adiguzel et al., 2023). This can help students learn at their own pace and focus on areas, where they need more support. AI-powered adaptive learning software such as DreamBox and Knewton use data analysis to create personalized learning plans for students based on their strengths and weaknesses. For example, students can complete a pre-assessment test that generates a personalized learning plan with specific goals and recommendations.
2. **Intelligent tutoring systems (ITS):** AI-powered tutoring systems can provide personalized support and feedback to students. These systems

can adapt to students' learning styles and provide tailored instruction and support, helping students to improve their learning outcomes. These systems can help students stay motivated and engaged with their learning by providing immediate feedback. Carnegie Learning's AI-powered mathematics tutoring system provides immediate feedback and customized learning paths based on students' strengths and weaknesses. The system adapts to each student's learning pace and provides interactive instruction and practice problems. Educators can use ITS to provide individualized instruction, monitor student progress, and identify areas, where students need additional support. Examples include ALEKS, Carnegie Learning, and Knewton.

3. **Automated grading:** AI can help automate the grading process, saving time for teachers and providing students with immediate feedback on their assignments. AI can provide feedback on grammar, spelling, and syntax by analyzing essays, reports, and other written assignments. By using automated grading systems, teachers can focus more on essential tasks such as lesson planning and supporting students, resulting in significant time savings (Adiguzel et al., 2023). This can help students to improve their writing skills and reduce the workload for teachers. For example, Turnitin's AI-powered software uses natural language processing (NLP) to analyze essays and provide grammar, spelling, and syntax feedback. The software can also detect plagiarism, helping teachers to grade assignments more efficiently and accurately.
4. **Predictive analytics:** AI can analyze student attendance, engagement, and performance data to predict future outcomes. This information can be used to identify students who may need additional support, enabling teachers to provide targeted interventions. For instance, University of South Florida uses predictive analytics to identify at-risk students who may need additional support. The university's Student Success Center uses data analysis to monitor students' progress and provide targeted interventions.
5. **Natural language processing:** AI-powered NLP tools can help students learn languages and improve their writing skills by providing grammar, spelling, and punctuation feedback. These tools can also help students to develop their critical thinking skills by analyzing and evaluating arguments and evidence. These tools also enable educators to analyze and interpret natural language data, such as student essays, discussions, and social media posts, to gain insights into student learning and engagement.

For example, Grammarly's NLP-powered software provides real-time feedback on writing, including suggestions for grammar, punctuation, and sentence structure. This can help students to improve their writing skills and produce higher-quality written work. Other examples include Google Cloud Natural Language, IBM Watson, and Microsoft Azure Cognitive Services.

6. **Intelligent content:** AI can help create and curate learning materials tailored to individual student's needs and learning styles. By analyzing student behavior data, AI can identify students' learning preferences and create customized learning materials that are engaging and relevant. For instance, EdTech startup Smart Sparrow uses AI to create interactive, adaptive learning materials customized to individual student's needs. For example, a biology course might use AI to generate different lab scenarios based on students' learning progress and interests.
7. **Virtual assistants:** AI-powered virtual assistants can help students with administrative tasks such as scheduling, reminders, and task management. These tools help students stay organized and focused, enabling them to manage their time and complete their coursework more efficiently. For instance, AI-powered virtual assistants such as Brainly provide students with immediate support and answers to their academic questions. The platform uses machine learning to provide personalized support and connect students with tutors.
8. **Automated transcription and translation:** AI can help transcribe and translate lectures and other educational materials, making them accessible to a wider range of students. This can help students who may have difficulty understanding the language used in their coursework or may have hearing impairments. It can also help make educational content more inclusive and accessible. For instance, Otter.ai uses AI-powered speech recognition to transcribe real-time lectures and other educational materials. The software can also translate spoken content into multiple languages, making educational content more accessible to students who may speak different languages. Other examples include Amazon Transcribe, Dragon Naturally Speaking, and Google Voice.
9. **Learning management systems (LMS):** These are platforms that allow educators to create, deliver, and manage learning materials, assignments, assessments, and evaluations for students. Educators can use LMS platforms to administer online assessments, monitor student progress, and provide feedback on student performance.

Examples include Blackboard, Canvas, and Moodle.

10. **Automated essay scoring (AES) software:** AES software uses NLP algorithms to evaluate and grade essays and written assignments. Educators can use AES software to provide students with immediate feedback on their writing, save time on grading, and ensure consistent and objective evaluation. Examples include e-rater, Grammarly, and Turnitin.
11. **Learning analytics tools:** Learning analytics tools use data mining and AI algorithms to analyze student learning data and provide insights into student performance, engagement, and learning outcomes. Educators can use learning analytics tools to monitor student progress, identify at-risk students, and make data-informed decisions to improve student learning outcomes. Examples include Learning Analytics and Knowledge (LAK) and Open Learning Analytics (OLA).
12. **Computer-based testing (CBT) platforms:** CBT platforms allow educators to administer online assessments, including multiple-choice, true/false, and essay questions. Educators can use CBT platforms to assess student knowledge, save time on grading, and provide students with immediate feedback (Bassey et al., 2020). Examples include ExamSoft, JAMB CBT, UNICAL Postgraduate e-exams, ProProfs, and Questionmark.
13. **Gamification tools:** Gamification tools use game-based elements to motivate and engage students in learning activities and assessments. Educators can use gamification tools to increase student engagement, promote student learning, and provide students with immediate feedback on their performance. Examples include Classcraft, Kahoot!, and Quizlet.
14. **Virtual reality (VR) and augmented reality (AR) tools:** VR and AR tools use immersive technologies to provide students with interactive, experiential learning experiences. Educators can use VR and AR tools to engage students in hands-on learning activities, provide real-world experiences, and enhance student learning outcomes. Examples include Google Expeditions, Merge Cube, and Nearpod VR.
15. **Formative assessment tools:** Formative assessment tools enable educators to monitor students' learning in real-time, provide feedback, and adjust instruction based on their performance. Examples include Mentimeter, Nearpod, and Socrative.
16. **Online polling tools:** Online polling tools allow educators to gather student feedback on specific

- topics or questions. Examples include Google Forms, Kahoot!, and Poll Everywhere.
17. **Interactive whiteboards:** Interactive whiteboards enable educators to present and annotate digital content, engage students in interactive activities, and collaborate with students in real-time. Examples include Google Jamboard, Promethean ActivPanel, and SMART Board.
18. **Video conferencing tools:** Video conferencing tools allow educators to facilitate virtual classroom sessions, provide remote instruction, and connect with students in remote locations. Examples include Google Meet, Microsoft Teams, and Zoom.
19. **Digital portfolios:** Digital portfolios enable students to collect and showcase their work, reflect on their learning, and receive feedback from educators and peers. Examples include Google Sites, Seesaw, and WordPress.
20. **Data visualization tools:** Data visualization tools enable educators to analyze and present data in visual formats, such as graphs and charts, to gain insights into student performance and learning outcomes. Examples include Google Data Studio, Infogram, and Tableau.
21. **Social media platforms:** Social media platforms enable educators to connect with students, share learning resources, and promote student engagement in learning activities. Examples include Facebook, Instagram, and Twitter.
22. **Plagiarism detection:** AI-based plagiarism detection software uses NLP algorithms to analyze student work and detect instances of plagiarism. Examples include Copyscape, Grammarly, and Turnitin.
23. **Classroom response systems:** Classroom response systems allow educators to pose questions and receive real-time student feedback using electronic devices. Examples include iClicker, Poll Everywhere, and Top Hat.
24. **Digital assessment tools:** Digital assessment tools provide educators with the ability to assess student knowledge and skills using a variety of question types, including multiple-choice, short answer, and essay. Examples include Edulastic, ExamView, and Google Forms.

LARGE LANGUAGE ARTIFICIAL INTELLIGENCE MODELS AND CLASSROOM ASSESSMENT

Large language models (LLMs) are a specialized form of AI that are purpose-built for comprehending, generating, and manipulating human language. By leveraging NLP principles and machine learning principles, LLMs are designed to process and interpret

vast quantities of text data (Dergaa et al., 2023). The “large” in their name refers to the massive datasets they are trained on and numerous parameters they possess, which enable them to grasp the subtle nuances and intricacies of human language. LLMs can generate human-like text and are designed to understand and generate text in a contextually relevant and coherent way.

LLMs are part of the family of generative models (Ingraham et al., 2019), which means they can generate new text based on the patterns and structures learned from the data used to train them. LLMs have many applications, including NLP, conversational AI, text generation, machine translation, sentiment analysis, and content creation. They can be used in various industries, such as healthcare, finance, customer service, marketing, and entertainment, to automate tasks, provide insights, and improve user experiences. One of the key features of LLMs is their ability to understand and conversationally generate text (Kasneci et al., 2023). They can engage in interactive and dynamic conversations with users, respond to queries, provide information, and generate relevant and coherent responses. LLMs are trained to understand context, tone, and style, making them capable of generating text that closely mimics human-like conversation. The development of LLMs can be considered among the greatest scientific advancements or breakthroughs in AI.

Examples of LLMs include BARD AI (Google), BERT (Google), Chat GPT (OpenAI), DistillBERT (Hugging Face), ELECTRA (Google), MarianMT (Microsoft Translator), Megatron (NVIDIA), RoBERTa (Facebook), T5 (Google/DeepMind), UniLM (Microsoft Research), and XLNet (Carnegie Mellon University/Google). LLMs can play a significant role in various stages of educational assessment process, such as test planning, item generation, preparation of test instruction, item assembly/selection, test administration, test scoring, test analysis, interpretation, appraisal, reporting, and follow-up. Further elaborations on how LLMs can serve useful purposes in educational measurement are documented.

Test Purpose Determination/Specification

Test purpose determination is the foremost step in the test development cycle. It involves identifying relevant educational issues that need to be addressed or key areas that require producing new knowledge or modifying existing ones. To Joshua (2012), some of the main purposes of testing revolve around evaluating teachers' effectiveness and students' motivation, judging students' learning proficiency, their acquisition of essential skills and knowledge, diagnosing students' learning difficulties, ranking students' learning achievement, and measuring their growth over time. Since purpose of a test is built from the course content or subject (Joshua, 2012), LLMs can be useful in the test specification of purpose by analyzing the course content

and identifying key topics or concepts that need to be assessed. For instance, an LLM can analyze a large amount of text data related to a specific course or subject, identify main themes and concepts, and suggest appropriate test items that accurately measure students' understanding of those concepts.

Moreover, LLMs can be used to create adaptive tests that adjust the difficulty level of questions based on students' responses. This can ensure that students are challenged appropriately and that the test accurately measures their knowledge and skills. LLMs can also generate test items that align with specific learning objectives and outcomes. For example, an LLM can analyze the course content, identify the key skills or knowledge students are expected to acquire, and generate test items that align with those objectives. LLMs can also help determine the most appropriate testing method to address these purposes and generate test items that align with each purpose. For example, LLMs can analyze large amounts of text data related to students' learning difficulties and suggest test items that can diagnose those difficulties.

Similarly, LLMs can analyze student growth data over time and suggest test items that accurately measure that growth. Furthermore, LLMs can help to ensure that test items are valid, reliable, and relevant to the educational issues being addressed. By analyzing large amounts of data related to student learning and educational issues, LLMs can suggest appropriate test items that accurately measure students' knowledge and skills. Additionally, LLMs can help to ensure that test items are fair and unbiased, which is essential in ensuring that the test results accurately reflect students' knowledge and skills.

Developing Test Blueprint

The test blueprint, also known as the table of specification, is a two-dimensional table relating the levels of instructional objectives of Bloom's taxonomy in the cognitive domain with carefully outlined content areas. Usually, the levels of instructional objectives in the cognitive domain are horizontally organized in ascending order of complexity at the top column. In contrast, the content areas are vertically organized in the leftmost row of the table. The test blueprint specifies the number of items required in each cell formed by the intersection or cross-tabulation of each cognitive instructional objective level and each area of the course contents. A test blueprint is essential to ensure that a test accurately measures the intended learning outcomes. It helps ensure that the test covers all critical content areas and that the questions appropriately align with the instructional objectives. LLMs can be extremely useful in the test specification of purpose for creating test blueprints or tables of specification. They can assist in developing test blueprints by analyzing the content

areas and the levels of instructional objectives in the cognitive domain. They can help educators determine which instructional objectives are the most critical for a particular test and ensure that the test covers all necessary content areas. LLMs can aid in developing a test blueprint by electronically analyzing relevant texts to identify the key concepts, skills, and knowledge areas that need to be assessed and help determine the appropriate weightage or distribution of these items in the test. This can help ensure the test is aligned with the objectives and intended construct.

Test Item Generation/Development

Test item development involves translating the course contents into test items or questions that will stimulate the learners and elicit the required behavior specified in the instructional objectives of the course (Joshua, 2012). Test items can be broadly classified into two categories:

- (1) objective items (highly structured items that have a clear and specific correct answer, often in the form of multiple-choice, true/false, or matching questions) and
- (2) essay items (open-ended question that requires the test-taker to provide a written response that demonstrates their understanding of a topic, their ability to articulate ideas clearly and coherently, and often their ability to analyze, synthesize, and evaluate information).

During the test item generation phase, a large pool of items is expected to be gathered from relevant sources, more than the number of items required for the test. The initial item pools can be reviewed with the support of domain experts or peers to identify relevant, clear, specific and unambiguous test items for selection. Those that do not meet the criteria for selection can either be strengthened or dropped.

LLMs can assist in creating relevant, clear, specific, and unambiguous test items. These models can analyze course content and other relevant sources to generate a large pool of potential test items. Domain experts or peers can then review the generated items to identify suitable items for selection. LLMs can also assist in developing objective test items such as multiple-choice, true/false, or matching questions. These highly structured items have a clear and specific correct answer that can be generated using LLMs. Training LLMs on relevant texts can generate items that assess specific skills or knowledge areas. LLMs can also generate distractors or incorrect options for multiple-choice questions, ensuring they are plausible but incorrect. This can help in the creation of a diverse and balanced item pool. These models can analyze the course content and generate options that align with the instructional objectives. Moreover, LLMs can also be useful in developing essay test items. These open-ended

questions require the test-taker to provide a written response demonstrating their understanding of a topic, their ability to articulate ideas clearly and coherently, and often their ability to analyze, synthesize, and evaluate information. LLMs can assist in generating essay prompts that align with the instructional objectives and are relevant to the course content.

In an experiment, a study found that questions generated by AI were satisfactorily clear, acceptable and favorable to students, and relevant to the subject matter (Nasution, 2023).

Preparation of Test Instruction

For objectivity in testing, simple, specific and clear test instructions must be developed to guide the test administrator and the respondents. According to Joshua (2012), the instructions for the testing procedures should explain why they are necessary. They should also contain information about how to organize the testing environment, distribute and collect test materials, manage time, procedures for recording answers and deal with anticipated and unforeseen inquiries. For the test takers, the instructions should include the test's purpose, the time allowed for the test, the basis for answering, expected ethical behaviors (dos and do nots), and discipline to be accorded for any breach of such rules.

LLMs can be extremely useful in the preparation of test instructions. These models are designed to understand and process natural language, which makes them ideal for tasks that require human-like language understanding and processing. Regarding testing instructions, LLMs can help ensure that the instructions are unambiguous. They can also help identify confusion or misunderstanding using specific words or phrasing. Additionally, LLMs can suggest alternative phrasing or wording that may be clearer or more easily understood. Moreover, LLMs can help ensure the instructions are culturally appropriate and sensitive to different audiences. They can identify potentially offensive or insensitive language and suggest more appropriate alternatives. LLMs can also help with the localization of test instructions. For example, if the test is being translated into a different language, an LLM can help ensure that the translated instructions accurately convey the intended meaning of the original instructions.

Item Assembly/Selection

LLMs can be very useful in the test assembly process. Test assembly involves selecting and organizing test items or questions to create a test that accurately measures a specific construct or skill. By analyzing relevant texts, LLMs can assist in identifying the most relevant and appropriate items from the item pool based on the test blueprint and objectives. This can help ensure that the test is well-balanced, covers the intended

construct, and is appropriate for the target population. They can assist in this process by analyzing the content of test items and identifying potential issues with wording, phrasing, or cultural biases. For example, an LLM can identify items with difficult vocabulary or syntax that may be confusing to test-takers or items that use colloquial language that may not be familiar to all test-takers. LLMs can analyze the coherence and consistency of test items to ensure that they assess the intended construct or skill fairly and validly. They can identify potential redundancies or overlaps in the content of test items or identify items irrelevant to the intended construct or skill. Furthermore, they can assist with translating test items into different languages, ensuring that the translated items accurately convey the intended meaning of the original items. LLMs can also assist in selecting and organizing test items by using NLP techniques to identify relationships between items. For example, an LLM can analyze the content of items and group them based on shared themes or concepts. All of these can be done in just a matter of seconds.

Test Administration

According to Joshua (2012), all students must be allowed to manifest the desired behavior being measured during test administration. During the test administration, the examiners must announce the test in advance, telling the examinees what, when, where, and how the test will be administered. The examiners must also assure the examinees that the test conditions will be satisfactory. There is also a need to minimize cheating using diverse approaches, such as adjusting the sitting arrangements for physically taken tests (Owan et al., 2023) and proctoring electronically taken tests (Owan, 2020; Owan et al., 2019). LLMs can be useful in various ways during test administration to ensure that all students have a fair chance to manifest the desired behavior being measured and to minimize cheating. LLMs can generate clear and accessible test instructions that students can understand easily. This can help ensure that all students have an equal opportunity to demonstrate their knowledge and skills, regardless of their language proficiency or other factors affecting their ability to understand the instructions. LLMs can monitor test-taking behavior during the test administration to detect unusual patterns that might suggest cheating or other misconduct. LLMs can be trained on data from previous test administrations to identify common cheating patterns and help examiners identify potential misconduct cases. LLMs can be used to support remote proctoring for electronically taken tests. This can include facial recognition technology to verify the test-takers' identity, eye-tracking technology to detect unusual eye movements, and keystroke analysis to detect unusual typing patterns. Furthermore, LLMs can help ensure the security of tests by providing features such as password

protection, encryption, and monitoring tools to prevent cheating and unauthorized access to test content.

LLMs can be used in test administration to facilitate the delivery and management of assessments. For instance, LLMs can be employed in CBT environments to deliver test items, record student responses, and monitor test progress. LLMs can also accommodate students with disabilities, such as text-to-speech or speech-to-text capabilities. Using LLMs in test administration can improve the testing experience's quality and increase test results' accuracy and reliability. One of the most significant benefits of LLMs in test administration is their ability to provide immediate feedback to test-takers. This can help students or respondents see their progress and performance in real-time, increasing motivation and engagement. Immediate feedback can also help to identify areas, where the test-taker needs to improve and provide opportunities for remediation. LLMs can also assist in delivering tests to many test-takers simultaneously, which can be particularly useful in high-stakes testing situations such as college entrance exams or professional certifications. Additionally, LLMs can provide a more engaging and interactive testing experience. They can be programmed to provide multimedia content, such as images, videos, and audio recordings, enhancing the learning experience and helping test-takers better understand complex concepts.

Test Scoring

Test scoring refers to evaluating and assigning a numerical score or grade to a test or assessment taken by a student or group of students. The score is typically based on the number of correct answers the student(s) gave on the test questions. However, it may also consider other factors such as partial credit, essay responses, or subjective evaluations by the test scorer or teacher. Test scoring is a common practice in education used to measure student knowledge, understanding, and skill levels and provide feedback and guidance for further learning.

LLMs can be useful for test scoring in several ways. They can be used to automate the process of grading and scoring tests, which can save a significant amount of time and effort for teachers and instructors. This is particularly useful in cases, where large-scale tests need to be graded quickly, such as in standardized testing or online assessments. LLMs can provide more accurate and consistent scoring of tests than human graders, as they are not subject to biases or errors arising from fatigue, distraction, or personal preferences. They can also be programmed to recognize and account for common mistakes or misconceptions made by students, which can help to identify areas, where further teaching or support may be needed. LLMs can provide more detailed feedback and analysis of test scores than

traditional scoring methods. They can be programmed to provide explanations or examples of correct answers and highlight areas, where a student may need to improve or focus more attention. This can guide further learning and development and provide more personalized and targeted support for individual students. Training LLMs on a large corpus of text can be fine-tuned to provide scores for open-ended questions or essays based on various criteria, such as content, language use, and organization. Automated scoring with LLMs can provide quick and consistent results, particularly for large-scale assessments.

Interpretation of Test Results

Interpreting test results means analyzing and making sense of the scores or outcomes obtained from a test or assessment. The interpretation of test results is an important part of the testing process. It allows for meaningful conclusions to be drawn about a student's knowledge, skills, and abilities and can inform decisions about teaching, learning, and further assessment. The interpretation of test results typically involves comparing the scores obtained by an individual or group of individuals to established norms or standards, such as the scores of other students in the same grade or subject area, or to pre-determined criteria for proficiency or mastery. This comparison can help identify areas of strength and weakness and provide insight into the overall level of knowledge or achievement of the test taker(s). In addition to comparing scores to established norms or standards, interpreting test results may involve examining patterns or trends in the scores over time or comparing scores on different tests or subtests within a larger assessment. This can help identify areas, where further learning or intervention may be needed and track progress or improvement.

LLMs can aid in interpreting assessment results by providing insights into the meaning and implications of the data. LLMs can be used to analyze and interpret test scores, performance levels, and other assessment outcomes in the context of the test objectives, standards, and criteria. This can help educators and policymakers make informed decisions about the performance of individuals or groups of test takers. Apart from their ability to automate the interpretation of a large volume of test results in record time, LLMs can provide a more detailed and nuanced analysis of test results than traditional methods. They can be programmed to recognize patterns or trends in the scores that may not be immediately apparent to human graders or analysts. They can also identify relationships between test items or subtests relevant to further teaching or assessment. LLMs can provide personalized feedback and recommendations based on individual test results, which can help to guide further learning and development. For example, an LLM could analyze a student's test results and provide recommendations for

specific areas, where they may need to focus more, or practice based on their strengths and weaknesses.

Test Analysis/Appraisal

Test analysis refers to the process of examining the results of a test in order to extract meaningful information about the performance of the test-takers. Test analysis aims to identify areas of strength and weakness and provide insights into the test-takers overall level of knowledge or achievement. The test analysis process typically involves gathering the scores and other relevant information from the test and organizing it to allow easy analysis. Computing descriptive statistics such as mean, standard deviation, and frequency distributions to summarize the test scores and provide an overview of the distribution of scores. Examining item-level performance by analyzing the performance of individual test items to identify items that were particularly difficult or easy for test-takers and identify items that may have been ambiguous or unclear. Identifying patterns or trends in the test scores across different test-taker subgroups (e.g., gender, ethnicity, or age) and over time (e.g., comparing scores from different test administrations). Based on the findings of the test analysis, conclusions can be drawn about the performance of the test-takers, as well as recommendations for further teaching or assessment be made. Various stakeholders, including test developers, educators, and policymakers, may conduct test appraisals. The process typically involves using established criteria or standards to evaluate the test, such as those outlined by American Psychological Association or National Council on Measurement in Education. The results of a test appraisal may be used to inform decisions about test selection, interpretation, and use, as well as to guide improvements in test development and administration processes.

LLMs have the potential to be valuable tools in test appraisals. LLMs can be employed to analyze assessment data, including item statistics, item difficulty, discrimination indices, and other performance metrics. LLMs can help identify patterns, trends, and anomalies in the data and provide insights into the overall performance of test takers and the quality of test items. LLMs can be useful in providing detailed feedback to students, highlighting areas where they need improvement or providing explanations for correct answers. The analytics generated by LLMs can provide insights into student strengths and weaknesses, highlight areas where additional instruction may be needed, and help teachers and administrators make informed decisions about instruction and resource allocation. LLMs can also identify potential sources of confusion or misunderstanding in test questions, such as multiple word or phrase meanings. This information can be used to revise questions or provide additional clarification to ensure that all students have a fair and

accurate understanding of what is being asked. LLMs can also be used in item analysis. Item analysis involves analyzing the performance of individual test items to identify areas, where they may be flawed or ineffective. LLMs can use student responses to specific items to highlight areas, where the item may be too difficult, too easy, or poorly worded. LLMs can also identify patterns in student responses to specific types of items, such as multiple-choice or essay questions. This information can inform decisions about the design and format of future assessments and ensure that assessments are as effective and fair as possible.

Reporting

Reporting refers to communicating the results of a test or assessment to relevant stakeholders, such as teachers, students, parents, and administrators. It typically involves summarizing the test results clearly and concisely and providing information about how they can inform instruction and decision-making. In educational testing, reporting often involves providing scores or grades that indicate how well students performed on the test. For example, a test may be scored on a scale from 0-100, with scores above a certain threshold indicating proficiency in a particular skill or subject area. Reporting may also include information about how students performed on different test items, or their performance compared to other students in their class or school. Reporting may also involve providing feedback to students and teachers about areas, where students performed well and where they may need additional support or instruction. This feedback can inform instruction and help students improve their performance on future assessments. In addition to communicating the results of a test, reporting may also include information about the validity and reliability of the test. This information can be used to evaluate the quality of the test and ensure that it provides accurate and useful information about student performance.

LLMs can assist in the generation of test reports. By analyzing the assessment data, LLMs can generate comprehensive reports summarizing the test results, including descriptive statistics, performance profiles, and graphical representations. LLMs can also generate interpretive reports that provide insights and recommendations based on the test results. These reports can be used by educators, policymakers, and other stakeholders for decision-making and planning purposes.

ROLE OF TEACHERS IN ARTIFICIAL INTELLIGENCE-BASED ASSESSMENT

AI-based assessments have been gaining popularity in recent years due to their ability to provide a quick and efficient evaluation of student learning. However, the role of teachers in AI-based assessments is still critical

and cannot be undermined. According to Dillenbourg (2016), the shift from traditional education to digital education does not necessarily mean that teachers will become obsolete in the future. Rather than debating whether AI will replace teachers, Hrastinski et al. (2019) suggest that it would be more sensible to recognize AI's benefits and how these benefits could transform their role in the classroom. Therefore, in educational assessment, teachers still have a critical role in ensuring that AI is appropriately used for measurement and evaluation purposes. Some of these roles include:

1. **Designing assessment:** Teachers are responsible for designing the assessment and setting the learning objectives. They clearly understand the curriculum and learning outcomes and can design assessments that align with these objectives.
2. **Providing context:** Teachers can provide context to the assessment questions, making them more relevant and meaningful to students. Doing so makes students more likely to engage with the assessment and understand its purpose.
3. **Interpreting results:** While AI-based assessments provide immediate feedback, the teacher's role is to interpret the results and provide actionable feedback to students. Teachers can use their professional judgment to evaluate student responses and provide personalized feedback based on their strengths and weaknesses.
4. **Continuous improvement:** Teachers can use the results from AI-based assessments to improve their teaching practices. By analyzing the data, teachers can identify areas of weakness and adapt their teaching strategies to meet the needs of their students better.
5. **Ethical considerations:** Teachers play a critical role in ensuring the ethical use of AI-based assessments. They must ensure that the assessment is fair, reliable, and valid and that student data is protected and used responsibly.
6. **Providing feedback:** Teachers can provide additional feedback to students beyond the automated feedback generated by AI-based assessments. This feedback can be more personalized and help students better understand their strengths and weaknesses and identify areas for improvement.
7. **Individualizing instruction:** Teachers can use the results from AI-based assessments to individualize instruction for students based on their learning needs. For example, if a student struggles with a particular concept, the teacher can provide additional resources or work with the student to develop a personalized learning plan.
8. **Monitoring progress:** Teachers can use AI-based assessments to monitor student progress over

time. By tracking student performance, teachers can identify trends and patterns and adjust their teaching strategies accordingly.

9. **Fostering critical thinking:** While AI-based assessments are designed to evaluate student knowledge, teachers can use them to foster critical thinking skills. By asking students to reflect on their responses and analyze the feedback provided, teachers can help students develop their ability to think critically about their learning.
10. **Ensuring accuracy:** Finally, teachers play a crucial role in ensuring the accuracy of AI-based assessments. They must ensure that the data is reliable and valid and that the assessment measures what it intends. Teachers can also identify and correct any errors or inaccuracies in the assessment process.

CHALLENGES OF USING ARTIFICIAL INTELLIGENCE-POWERED TOOLS IN EDUCATIONAL ASSESSMENT

AI powered tools in educational assessment can be beneficial in various ways, including enhancing the assessment process's accuracy, speed, and efficiency. However, several challenges are associated with using AI in educational assessment. Some of these challenges include:

1. **Lack of stakeholders' participation in developing artificial intelligence tools:** The lack of stakeholder participation in developing AI tools for education can be a significant challenge in AI adoption. When AI tools are developed without input from educators, students, parents, and other stakeholders, the resulting tools may not be tailored to the specific needs of the education system. This can lead to a lack of relevance or adoption of AI tools in classrooms. According to Luckin and Cukurova (2019), developers of AI have limited knowledge about learning sciences and insufficient pedagogical understanding for the successful integration of AI into teaching. Consequently, AI developers frequently overlook the expectations of teachers, who are the end-users of AI in education (Cukurova & Luckin, 2018). Teachers are highly important in AI-based teaching (Seufert et al., 2020); therefore, it is essential to take their opinions, past experiences, and expectations into account to ensure the successful integration of AI in schools (Holmes et al., 2019).
2. **Lack of transparency:** One of the primary challenges associated with AI in educational assessment is the lack of transparency in decision-making. It can be challenging to understand how AI algorithms arrive at their conclusions, making

- it difficult for educators to assess the accuracy and fairness of the assessments.
3. **Bias:** AI algorithms can be biased, leading to inaccurate and unfair assessments. This is because AI systems are only as good as the data on which they are trained. If the data used to train the AI system is biased, then the system will produce biased outcomes.
 4. **Lack of human interaction:** The use of AI in educational assessment can lead to a lack of human interaction in the assessment process, which can be detrimental to students' learning experiences. Educators may miss important cues and nuances that can only be detected through human interaction.
 5. **Limited scope:** AI-powered tools are typically designed to focus on specific assessment areas, such as lower-order cognitive abilities (such as knowledge and comprehension, usually measured with objective test items). This means that they may not be suitable for assessing other important areas of learning, such as critical thinking, creativity, and problem-solving. Furthermore, AI-tools, due to their inanimate nature, may not be well-suited to measure and grade affective attributes (such as students' feelings, attitude, interest and anxiety) and psychomotor skills (such as driving, footballing, cooking, weaving, laboratory work, and dancing, among others). Therefore, the teacher must cover up in these areas and provide the relevant feedback to the AI model.
 6. **Ethical concerns:** The use of AI in educational assessment raises ethical concerns, such as issues around data privacy and ownership, as well as concerns around using technology to replace human educators.
 7. **Limited understanding:** Another challenge with using AI-powered tools in educational assessment is that educators may have a limited understanding of how the technology works. Educators may lack the technical expertise to fully comprehend the algorithms used in AI tools and how they affect assessment outcomes.
 8. **Inadequate training:** Educators must be trained to understand how to use the technology and effectively interpret the results. Without adequate training, educators may be unable to use the tools effectively, leading to inaccurate assessments.
 9. **Integration with existing systems:** Integrating AI-powered assessment tools with existing systems can be a challenge. Educators must invest in new technologies and infrastructure and ensure the new tools work seamlessly with existing assessment processes.
 10. **Cost:** The development and implementation of AI-powered assessment tools can be costly. Some schools or educational institutions may not have the financial resources to invest in these technologies, which can lead to unequal access to these tools.
 11. **Resistance to change:** Adopting AI-powered assessment tools may face resistance from educators, students, and parents. They may not trust the technology or may prefer traditional assessment methods. Introducing these new tools into the educational assessment process can make it challenging.
 12. **Student motivation and engagement:** AI-powered assessment tools may affect student motivation and engagement in learning. Students may perceive the assessments as impersonal or become overly reliant on technology, leading to a lack of effort in their learning.
 13. **Standardization:** AI-powered assessment tools are often designed to provide standardized tests that can be used to compare students' performance. But focus on standardization may not account for the diversity of students' learning experiences and backgrounds, leading to a one-size-fits-all approach to education.
 14. **Technical difficulties:** AI-powered assessment tools require a stable and reliable technology infrastructure. Technical difficulties, such as power outages, internet disruptions, or software malfunctions, can disrupt the assessment process, leading to inaccurate or incomplete assessments.
 15. **Data management:** AI-powered assessment tools generate large amounts of data that must be stored, managed, and analyzed. Educators must ensure that the data is accurate, secure, and compliant with privacy regulations.
 16. **Feedback and support:** AI-powered assessment tools can provide quick feedback on students' performance but may not offer the personalized feedback and support students need to improve their learning. Educators need to balance the benefits of quick feedback with the importance of providing meaningful feedback that supports students' learning and development.

STRATEGIES TO ADDRESS THE CHALLENGES OF USING ARTIFICIAL INTELLIGENCE IN EDUCATIONAL ASSESSMENT

There are several ways to address the challenges of using AI-powered tools in educational measurement and assessment. Below are some ways in which the challenges earlier identified could be addressed to

promote equity, transparency, and quality in educational assessment:

1. **Developing transparent and ethical artificial intelligence algorithms:** To ensure that AI-powered tools are transparent, educators and developers must work together to develop algorithms free from bias and discrimination. This can involve designing algorithms based on diverse data sets and incorporating ethical considerations into the design process.
2. **Using personalized and adaptive assessment approaches:** To capture the nuance of students' learning experiences, educators should consider using personalized and adaptive assessment approaches that can account for individual differences in students' learning styles and needs. This can involve using various assessment methods, including open-ended questions and performance tasks, to better understand students' knowledge and skills.
3. **Providing training and support for educators:** To use AI-powered tools effectively, educators need proper training and support. This can involve providing educators with professional development opportunities to use the technology and ongoing support to ensure they can effectively integrate it into their teaching practice.
4. **Collaborating with students:** To ensure that AI-powered tools are meeting the needs of students, educators should collaborate with them to gather feedback on their experiences with the technology. This can involve soliciting student input on the usability and effectiveness of the tools and incorporating their feedback into the design and development process.
5. **Ensuring accessibility for all students:** To ensure that AI-powered tools are accessible for all students, educators should consider the needs of students with disabilities and provide accommodations as needed. This can involve designing tools compatible with assistive technology, providing alternative formats for materials, and ensuring that the tools comply with accessibility standards.
6. **Incorporating human inputs and oversight:** While AI-powered tools can provide valuable insights into student learning, it is important to remember that they are not infallible. Educators should incorporate human input and oversight into the assessment process to ensure that the results are accurate and fair. This can involve reviewing and verifying the results generated by the AI algorithm and adjusting them as needed to ensure that they are fair and accurate.
7. **Regular evaluation and update of technology:** AI technology is constantly evolving, and educators

need to stay updated with the latest advancements in the field. This can involve regularly evaluating and updating the technology to ensure that it meets the needs of students and educators and remains transparent, fair, and compliant with data privacy regulations.

8. **Ensuring student data security and privacy:** AI-powered tools in educational assessment often involve collecting and analyzing student data. It is important to ensure that data are secure, protected, and only used for its intended purpose. This can involve implementing appropriate data security measures, such as encryption and access controls, and ensuring compliance with data privacy regulations such as Family Educational Rights and Privacy Act (FERPA).
9. **Educating students and parents on using artificial intelligence-powered tools:** To ensure that students and parents are comfortable using AI-powered tools in educational assessment, educators should provide education and training on the technology. This can involve explaining how the technology works, what data is being collected, and how it improves learning outcomes.

CONCLUSIONS

In conclusion, this paper has explored the various applications of AI in educational assessment and the challenges that need to be addressed to enhance its effectiveness. The use of AI in educational assessment has both benefits and limitations.

On the one hand, AI-based assessment can provide more accurate, objective, and efficient grading, freeing up time for teachers to focus on more meaningful interactions with their students. It can also identify areas of weakness and strength, allowing educators to tailor their teaching methods to individual needs. However, AI-based assessment is not a panacea and should not replace human judgment entirely.

AI algorithms can be biased and may not consider non-cognitive factors that can influence academic performance. Additionally, some students may be uncomfortable with being assessed by machines, which could impact their motivation and engagement. Therefore, while AI can enhance educational assessment, it should be used alongside human judgment and ethical considerations.

While using AI-powered tools in educational assessment presents several challenges, many ways exist to address these challenges and ensure that the technology is used ethically and effectively. By incorporating human input and oversight, regularly evaluating and updating the technology, and ensuring the security and privacy of student data, we can create a more equitable and effective educational assessment system.

As such, educators, policymakers, and stakeholders must work together to develop strategies that maximize the benefits of AI in educational assessment while mitigating the associated risks. AI in educational assessment can ultimately transform education, improve learning outcomes, and equip students with the skills needed to succeed in the 21st century.

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: Authors stated that the study did not require an ethical approval since it is based on existing literature.

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429. <https://doi.org/10.30935/cedtech/13152>
- Bassey, B. A., Ubi, I. O., Anagbogu, G. E., & Owan, V. J. (2020). Permutation of UTME multiple-choice test items on performance in use of English and mathematics among prospective higher education students. *The Journal of Social Sciences Research*, 6(4), 483-493. <https://doi.org/10.32861/jssr.64.483.493>
- Chatterjee, J., & Dethlefs, N. (2023). This new conversational AI model can be your friend, philosopher, and guide ... and even your worst enemy. *Patterns*, 4(1), 100676. <https://doi.org/10.1016/j.patter.2022.100676>
- Cukurova, M., & Luckin, R. (2018). *Measuring the impact of emerging technologies in education: A pragmatic approach*. Springer. https://doi.org/10.1007/978-3-319-53803-7_81-1
- Delgado, H. O. K., de Azevedo Fay, A., Sebastian, M. J., & Silva, A. D. C. (2020). Artificial intelligence adaptive learning tools. *BELT-Brazilian English Language Teaching Journal*, 11(2), e38749-e38749. <https://doi.org/10.15448/2178-3640.2020.2.38749>
- Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615-622. <https://doi.org/10.5114/biolsport.2023.125623>
- Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education*, 26(2), 544-560. <https://doi.org/10.1007/s40593-016-0106-z>
- Elliott, D., & Soifer, E. (2022). AI technologies, privacy, and security. *Frontiers in Artificial Intelligence*, 5, 826737. <https://doi.org/10.3389/frai.2022.826737>
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421. <https://doi.org/10.30935/cedtech/13036>
- Hawes, D., & Arya, A. (2023). Technology solutions to reduce anxiety and increase cognitive availability in students. *IEEE Transactions on Learning Technologies*, 16(2), 278-291. <https://doi.org/10.1109/TLT.2023.3239985>
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. *Center for Curriculum Redesign*. <https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf>
- Hou, J., Li, Z., & Liu, G. (2022). Macro education approach to improve learning interest under the background of artificial intelligence. *Wireless Communications and Mobile Computing*, 2022, 4295887. <https://doi.org/10.1155/2022/4295887>
- Hrastinski, S., Olofsson, A. D., Arkenback, C., Ekström, S., Ericsson, E., Fransson, G., Jaldemark, J., Ryberg, T., Öberg, L.-M., Fuentes, A., Gustafsson, U., Humble, N., Mozelius, P., Sundgren, M., & Utterberg, M. (2019). Critical imaginaries and reflections on artificial intelligence and robots in post-digital K-12 education. *Post-Digital Science and Education*, 1(2), 427-445. <https://doi.org/10.1007/s42438-019-00046-x>
- Hu, Y., & Min, H. K. (2023). The dark side of artificial intelligence in service: The “watching-eye” effect and privacy concerns. *International Journal of Hospitality Management*, 110, 103437. <https://doi.org/10.1016/j.ijhm.2023.103437>
- Huang, A. Y., Lu, O. H., & Yang, S. J. (2023a). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194, 104684. <https://doi.org/10.1016/j.compedu.2022.104684>
- Huang, S.-P. (2018). Effects of using artificial intelligence teaching system for environmental education on environmental knowledge and attitude. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(7), 3277-3284. <https://doi.org/10.29333/ejmste/91248>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023b). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112-131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)

- Ingraham, J., Garg, V., Barzilay, R., & Jaakkola, T. (2019). Generative models for graph-based protein design. In H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc.
- Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(2), 100115. <https://doi.org/10.1016/j.tbench.2023.100115>
- Joshua, M. T. (2012). *Fundamentals of tests and measurement in education*. University of Calabar Press.
- Karsenti, T. (2019). Artificial intelligence in education: The urgent need to prepare teachers for tomorrow's schools. *Formation et Profession [Education and Profession]*, 27(1), 112-116. <https://doi.org/10.18162/fp.2019.a166>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khan, I., Ahmad, A. R., Jabeur, N., & Mahdi, M. N. (2021). An artificial intelligence approach to monitor student performance and devise preventive measures. *Smart Learning Environments*, 8(1), 1-18. <https://doi.org/10.1186/s40561-021-00161-y>
- King, M. R., & ChatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, 16(1), 1-2. <https://doi.org/10.1007/s12195-022-00754-8>
- Kumar, N. S. (2019). Implementation of artificial intelligence in imparting education and evaluating student performance. *Journal of Artificial Intelligence*, 1(01), 1-9. <https://doi.org/10.36548/jaicn.2019.1.001>
- Lin, P. Y., Chai, C. S., Jong, M. S. Y., Dai, Y., Guo, Y., & Qin, J. (2021). Modelling the structural relationship among primary students' motivation to learn artificial intelligence. *Computers and Education: Artificial Intelligence*, 2, 100006. <https://doi.org/10.1016/j.caiei.2020.100006>
- Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6), 2824-2838. <https://doi.org/10.1111/bjet.12861>
- Luo, Y., Han, X., & Zhang, C. (2022). Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses. *Asia Pacific Education Review*. <https://doi.org/10.1007/s12564-022-09749-6>
- Mena-Guacas, A. F., Urueña Rodríguez, J. A., Santana Trujillo, D. M., Gómez-Galán, J., & López-Meneses, E. (2023). Collaborative learning and skill development for educational growth of artificial intelligence: A systematic review. *Contemporary Educational Technology*, 15(3), ep428. <https://doi.org/10.30935/cedtech/13123>
- Mohamed, M. Z. B., Hidayat, R., Suhaizi, N. N. B., Sabri, N. B. M., Mahmud, M. K. H. B., & Baharuddin, S. N. B. (2022). Artificial intelligence in mathematics education: A systematic literature review. *International Electronic Journal of Mathematics Education*, 17(3), em0694. <https://doi.org/10.29333/iejme/12132>
- Naqvi, A. (2020). *Artificial intelligence for audit, forensic accounting, and valuation: A strategic perspective*. John Wiley & Sons. <https://doi.org/10.1002/9781119601906>
- Nasution, N. E. A. (2023). Using artificial intelligence to create biology multiple choice questions for higher education. *Agricultural and Environmental Education*, 2(1), em002. <https://doi.org/10.29333/agrenedu/13071>
- Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53(4), 914-931. <https://doi.org/10.1111/bjet.13232>
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Owan, V. J. (2020). Computer-administered testing practice in higher education in an era of severe acute respiratory syndrome-related diseases outbreaks. In V. C. Emeribe, L. U. Akah, O. A. Dada, D. A. Alawa, & B. A. Akuengwu (Eds.), *Multidisciplinary issues in health, human kinetics and general education practices* (pp. 429-442). University of Calabar Press. <https://doi.org/10.2139/ssrn.3906923>
- Owan, V. J., Duruamaku-dim, J. U., & Eneje, S. (2019). Mode of test administration, birth variables, and students' academic achievement in mathematics in Obubra Local Government Area of Cross River

- State, Nigeria. *Prestige Journal of Counselling Psychology*, 2(2), 60-77.
- Owan, V. J., Owan, M. V., & Ogabor, J. O. (2023). Sitting arrangement and malpractice behaviors among higher education test-takers: On educational assessment in Nigeria. *Journal of Applied Learning & Teaching*, 6(1), 1-15. <https://doi.org/10.37074/jalt.2023.6.1.5>
- Papapicco, C. (2020). Google mini: Italian example of artificial pro-sociality. *Online Journal of Communication and Media Technologies*, 10(3), e202015. <https://doi.org/10.29333/ojcmi/7995>
- Ren, X. (2020). Artificial intelligence and depression: How AI powered chatbots in virtual reality games may reduce anxiety and depression levels. *Journal of Artificial Intelligence Practice*, 3(1), 48-58.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 1-22. <https://doi.org/10.37074/jalt.2023.6.1.29>
- Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2022). Assessing behavioral data science privacy issues in government artificial intelligence deployment. *Government Information Quarterly*, 39(4), 101679. <https://doi.org/10.1016/j.giq.2022.101679>
- Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre-and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 106552. <https://doi.org/10.1016/j.chb.2020.106552>
- Suh, W., & Ahn, S. (2022). Development and validation of a scale measuring student attitudes toward artificial intelligence. *Sage Open*, 12(2), 21582440221100463. <https://doi.org/10.1177/21582440221100463>
- Tovani-Palone, M. R. (2023). Some challenges and limitations of using ChatGPT in medicine. *Electronic Journal of General Medicine*, 20(5), em503. <https://doi.org/10.29333/ejgm/13263>
- Vasconcelos, M. A. R., & dos Santos, R. P. (2023). Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(7), em2296. <https://doi.org/10.29333/ejmste/13313>
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- Xia, Q., Chiu, T. K. F., Lee, M., Temitayo I., Dai, Y., & Chai, C. S. (2022). A self-determination theory design approach for inclusive and diverse artificial intelligence (AI) K-12 education. *Computers & Education*, 189, 104582. <https://doi.org/10.1016/j.compedu.2022.104582>

<https://www.ejmste.com>