The Use of AI for Test Items Setting and Marking Schemes in the University Setting

[1]Ibrahim Shuaibu, PhD, [2]Ahmad Sabo

[1]Department of Education, Sule Lamido University Kafin Hausa, Jigawa State

[2]Department of Curriculum and Instructions, School of Education,
Aminu Kano, College of Islamic and Legal Studies, Kano
Email: *ibrahim.shuaibu@slu.edu.ng*

ABSTRACT

It has been established that Artificial Intelligence (AI) driven systems, otherwise known as tools, can be used to enhance educational delivery and improve efficiency. In this paper, the paper examines the use of AI in the examination process especially within the university level. The paper discusses the use of AI as a learning assistant, its uses in the examination process, AI tools that can be used for examination setting, such as examsoft, exam pro among others. It further examines the benefits associated with the use of AI in the examination and concludes by identifying certain limitations and challenges that could be encountered by the users. The paper recommended that AI here serves as educational assistant not a replacement for the mandatory human interface.

Keywords: artificial intelligence, Test Marking scheme

INTRODUCTION

Artificial intelligence is transforming education by enabling personalized learning, improving administrative efficiency, and promoting equity, while requiring ethical oversight and policy support. AI-driven systems can tailor educational content to individual student needs, providing customised learning pathways and pacing based on student performance and preferences otherwise known as personalised learning. AI assessment systems are digital platforms that leverage machine learning **(**ML) and natural language processing (NLP) to evaluate student work, employee skills, or candidate performance. They automate routine tasks such as grading multiple-choice questions, coding assignments, and essays, while providing personalized, real-time feedback to learners or candidates, significantly reducing the workload for educators and HR professionals. These systems also detect patterns in responses, track progress, and generate insights for informed decision-makingAI tools like machine learning algorithms and natural language processing assist in automating assessment tasks, such as grading exams, assignments, and even essays, reducing faculty workload and delivering faster feedback to students. This is a reflection of AI ability to automate grading and feedback, a key component of the university education system. Another area in which AI has transformed education, at the university level is that AI-powered virtual teaching

assistants and chatbots offer round-the-clock student support, answering questions, tutoring, and even offering supplemental learning resources.

In the university system today, AI tools analyse vast datasets to identify gaps in curricula, ensuring that academic programs remain relevant and aligned with industry needs and this is known as 'Curriculum Development'. Furthermore, universities now leverage on AI to analyse student data and predict trends like dropout rates, allowing for early intervention strategies and improving student retention. It is imperative to note also that, AI enhances academic research through tools that assist with literature review, data analysis, and even collaborative research efforts across institutions.

Artificial Intelligence, as a Learning Assistant can be used to personalised Learning support for student. For example, AI-powered tools would analyse student data and adapt lessons, resources, and feedback to individual learning styles and progress. This in turn will create a tailored learning experience, helping students grasp difficult concepts and pace their studies effectively. Similarly, AI has a 24/7 Availability operational ability. This is because, AI virtual assistants and chatbots are available around the clock to answer students' questions, provide explanations, and offer guidance on course materials, reducing dependency on faculty for routine inquiries. Furthermore, through the use of AI, Automated Feedback is ensured.

More so as AI tools can review assignments quizzes, and even complex tasks like essays, providing instant, detailed feedback. This speeds up the learning process and help students correct mistakes immediately. Another important area where AI is useful as learning assistant at the university level, is its ability to serve as, Learning Analytics Learning analytics improves item difficulty, item discrimination, and item distractor in as sessments by utilizing item analysis techniq ues. AI tracks student performance, engagement, and behaviour patterns, helping both students and instructors identify areas that need attention. Universities use these insights to tailor support strategies and improve outcomes. It is noteworthy that, it can be deployed as a tool for Supplemental Tutoring. AI-driven tutoring systems provide students with extra help in difficult subjects by offering targeted practice problems, explanations, and study resources, acting as a personal tutor.

AI in Examination Process

AI in the examination process, particularly at the university level, enhances efficiency and accuracy in test administration and evaluation. AI tools can automate exam creation, generating diverse question types tailored to different learning outcomes. These systems often incorporate machine learning to ensure test validity and can adapt questions based on student performance for formative assessments. In grading, AI tools like automated essay scoring systems quickly assess written responses, providing objective and timely feedback. This reduces the burden on instructors and ensures consistency in grading. However, integrating AI raises concerns about fairness, data security, and the need for oversight to maintain academic integrity. Let us look at the following illustrations in the area of AI involvement in examination process:

Automated Test Generation: AI tools like Quizlet or ExamSoft can generate exam questions based on predefined course objectives and student learning patterns. For instance, in a university physics class, an AI system could generate multiple-choice, short-answer, and problem-solving questions from a pool of content, ensuring varied difficulty levels and topics. This

saves professors time in creating exams and can even offer adaptive testing, where questions adjust in difficulty based on a student's performance, providing a personalised assessment experience.

AI-based Essay Grading: Tools like Grammarly and Turnitin now include AI-powered features that can assess student essays or long-form responses by evaluating grammar, coherence, and argument strength. In a literature course, for example, AI could grade essays based on set rubrics, offering instant feedback on structure, argument logic, and plagiarism detection. This not only speeds up the grading process but also ensures consistency across all student submissions, reducing the likelihood of subjective bias in evaluation. Apart from the foregoing, it is imperative to know that AI can be used for content generation, adaptive examination generation and personalisation of test creation. The following illustrations would explain these more:

AI Content Generation: AI tools can automatically generate test questions based on course content, learning objectives, and specific guidelines set by instructors. These tools analyse the curriculum, textbooks, and class materials to create relevant questions of varying formats such as multiple-choice, short answer, true/false, and even complex problem-solving exercises. For example, AI-powered platforms like QuestionPro or ExamSoft can generate a wide array of questions from a single chapter, ensuring coverage of key concepts. AI can also balance question difficulty, ensuring a range of easy, moderate, and challenging questions.

Adaptive Test Generation: Adaptive testing leverages AI to tailor exams based on a student's real-time performance. For example, if a student answers several easy questions correctly, the AI system automatically increases the difficulty of subsequent questions. Conversely, if a student struggles with difficult questions, the test will provide easier ones. This adaptive generation allows for a more personalised assessment experience. AI ensures the test remains challenging and fair, optimising learning outcomes. Adaptive tests are often used in large-scale assessments like the GRE or GMAT, but universities are increasingly adopting these AI-driven methods to individualise exams and gauge student comprehension.

Personalisation of Test Creation: AI enables the personalisation of test creation by considering individual learning styles, progress, and areas needing improvement. By analysing a student's past performance, behaviour during quizzes, and engagement with study materials, AI can tailor exams specifically for each student. For example, in a university-level biology course, if a student excels in genetics but struggles with molecular biology, AI-generated tests can focus more on the latter to help address learning gaps. Personalised assessments help students focus on weak areas and reduce unnecessary stress on topics they have mastered. This targeted approach to testing not only improves performance but also enhances long-term knowledge retention. The implication of the above is that AI transforms the exam process by generating varied and comprehensive content, adapting tests to real-time performance, and personalising assessments to individual students' needs. These developments help create fairer, more efficient, and more effective educational assessments Darad (2025).

Marking and Assessment with AI: Artificial Intelligence (AI) has significantly transformed the landscape of marking and assessment, particularly at the university level, by enhancing efficiency, accuracy, and consistency. The use of AI in these areas reduces the workload for educators,

improves feedback quality for students, and ensures greater objectivity in evaluation. As universities embrace digital transformation, AI-powered tools are increasingly integrated into grading systems, providing both automated and semi-automated solutions for assessment.

Automated Grading for Objective Tests: One of the most prominent uses of AI in university assessments is the automated grading of objective questions, such as multiple-choice, true/false, and matching tests. AI-driven software can instantly evaluate large volumes of answers, providing immediate feedback to students. Tools such as Grade Scope and Exam Soft can scan and assess digital answer sheets quickly and efficiently, ensuring consistent results across all student submissions. This feature is especially valuable in large classes or for standardised tests, where manual grading would otherwise be time consuming. The automation of grading for objective questions not only speeds up the process but also allows for large-scale assessments to be conducted more frequently, giving students regular and instant feedback. This immediacy is crucial for improving learning outcomes, as students can quickly identify areas where they need to improve.

AI for Essay and Open-Ended Responses: AI has also made strides in grading essays and other openended responses, areas traditionally marked by human graders. Tools like Grammarly and Turnitin now incorporate AI algorithms to assess grammar, coherence, and structure, while more sophisticated systems such as E-rater (used by ETS) analyse argument quality, style, and even critical thinking. These tools can provide detailed feedback on a student's writing in a matter of minutes, flagging common issues like grammar errors, unclear ideas, or lack of supporting evidence.

AI-based essay scoring systems use machine learning models trained on thousands of essays to understand patterns in high-quality and low-quality writing. These systems offer consistent grading, eliminating human biases that can often affect subjective assessments. However, they do have limitations, particularly when dealing with nuanced or creative responses that require deeper understanding and contextual evaluation. In such cases, AI is best used as a preliminary tool for grading, with final reviews left to human assessors.

Adaptive Assessment and Real-Time Feedback: AI not only marks existing assessments but also helps create adaptive assessments that adjust based on student performance in real-time. Adaptive testing is a personalised form of assessment where the difficulty of questions varies depending on how well the student is performing. For example, if a student answers several easy questions correctly, the system will present them with harder questions to gauge deeper understanding. This form of AI-powered assessment is particularly useful for diagnostic purposes, helping educators identify knowledge gaps or areas where students need further support. In universities, adaptive assessments are increasingly used in formative evaluations to guide learning rather than summative assessments that focus solely on end results. By providing real-time feedback, AI enhances learning by allowing students to immediately see their mistakes and make corrections during the learning process.

Objectivity and Consistency in Grading: One of the key advantages of AI in assessment is the consistency and objectivity it brings to grading. Human assessors can be influenced by subjective factors such as bias, fatigue, or emotional state, which may lead to inconsistent results. AI, however, applies the same grading criteria uniformly to all students, ensuring

fairness in evaluation. This consistency is especially important in large-scale assessments, where multiple graders are involved, and maintaining uniformity in grading can be challenging (Oniye, 2024).

Challenges and solutions: Despite its benefits, AI-based marking and assessment are not without challenges. The primary concern revolves around the potential lack of nuance in AI's grading capabilities, particularly for open-ended or creative responses. AI may struggle to fully understand context, creativity, or cultural nuances, potentially leading to inaccurate assessments. Moreover, the use of AI raises ethical concerns about data security and the transparency of algorithms used in grading. It is crucial for universities to strike a balance between automated systems and human oversight, ensuring that AI complements, rather than replaces, traditional evaluation methods. The following illustrations would help us to comprehend deeper, the use of AI in marking and Assessment:

AI-Assisted Essay Grading in a Literature Course: In a university literature course, students are often required to submit essays that analyse texts, explore themes, or critique literary techniques. Traditionally, grading such essays can be time-consuming for instructors, especially in large classes. An AI tool like Erater or Grammarly can be used to assess the essays based on grammar, coherence, organisation, and argument structure. The AI system scans each submission and provides an initial score, focusing on technical aspects such as sentence structure, logical flow, and use of language. It can also flag areas that need improvement, like ambiguous phrasing or unsupported arguments. The AI provides detailed feedback to students instantly, helping them improve their writing skills. The final score can then be reviewed and adjusted by the instructor, ensuring that creative or nuanced interpretations are fairly evaluated. This approach streamlines the grading process while maintaining academic rigor (Carl, 2023).

Automated Grading of Objective Exams in a Large Biology Class-In a first-year biology course with 300 students, the final exam consists of multiple-choice, true/false, and matching questions that test knowledge of cellular biology and genetics. Using an AI-powered assessment platform like GradeScope or ExamSoft, the instructor uploads the students' answers to the system, which instantly grades all submissions with high accuracy. The AI system can also analyse student performance trends, identifying which questions were most challenging for the class and offering insights into topics that may require additional review. With automated grading, the instructor can provide timely feedback to students, and the saved time can be used to offer more personalised guidance during office hours or group discussions. Additionally, the AI ensures consistency in grading, eliminating any potential human error or bias.

AI Tools for Examination Question Setting and Marking scheme

AI tools examination setting and marking in university education, enhancing both efficiency and accuracy. For exam creation, tools like Exam Soft, Question Pro and Q correct automate the process by generating diverse question formats such as multiple-choice, short-answer, and essay questions. These tools can adapt questions to align with specific learning outcomes and difficulty levels, reducing the time educators spend on manual exam design. Moreover, AI can analyse past student performance to create personalised assessments, ensuring comprehensive coverage of key topics. (Nasr, 2021).

In marking, AI tools like Grade Scope and Turnitin streamline the grading process for both objective and open-ended questions. Objective tests such as multiple-choice tests though primarily assess specific knowledge or skills, often focusing on factual recall rather than deeper understanding or critical thinking. This makes them less effective for evaluating complex concepts or higher-order skills such as analysis, synthesis, and evaluation yet are instantly graded, providing immediate results and detailed feedback to students. For essays and written responses, AI systems like E-rater or Grammarly evaluate grammar, coherence, and structure, offering preliminary scores that educators can review. This reduces grading time, maintains consistency, and minimises human error or bias (Nicol, & Macfarlane-Dick,2024). Additionally, AI tools offer insights into student performance, helping educators identify common areas of struggle, thereby allowing targeted interventions. While AI facilitates efficient assessment, human oversight remains crucial to address its limitations, especially in nuanced or creative responses. As a university teacher, the following illustrations would help you to understand better, the use of specific AI tools in examination setting and marking:

Examination Setting with Exam Soft: In a university-level psychology course, the instructor uses ExamSoft to set mid-term exams. The AI-powered platform allows the instructor to upload course materials and automatically generate a range of questions, including multiple-choice, short-answer, and case-based problem-solving questions. The instructor can specify which learning objectives each question should address, and the system ensures that the exam covers the right balance of topics. Additionally, ExamSoft allows the instructor to set different difficulty levels for different sections of the exam, ensuring that students are adequately challenged. The platform also generates multiple versions of the test to reduce the chances of cheating during the exam.

AI-Powered Grading with Grade Scope: In a large undergraduate engineering course, the final exam consists of both objective and open-ended problem-solving questions. The instructor uses GradeScope, an AI tool for automated grading, to evaluate the objective portions (multiple-choice, true/false) of the test. For the open-ended questions, the AI assists by grouping similar student responses, enabling the instructor to grade more efficiently by applying the same rubric to clustered answers.

Benefits of Using AI in Examination Marking scheme

The benefits of utilising specific AI tools in the examination process in our universities are diverse especially bearing in mind that, now more than ever before universities are experiencing higher enrolment of university degree seekers a development that continues to challenge the manpower availability in the system. These benefits can be appreciated further if we look at them from the following perspectives:

The integration of AI into the university examination process offers numerous benefits, enhancing efficiency, accuracy, and personalisation. One of the primary advantages is automated exam creation, where AI tools such as Exam Soft can generate a variety of questions, including multiple-choice, short answer, and essay questions, based on specific learning outcomes and difficulty levels. This saves educators significant time in designing assessments, allowing them to focus more on content delivery and student engagement. Moreover, AI systems can generate personalised assessments, tailoring questions to a student's performance history and learning needs.

In terms of grading and assessment, AI tools like Grade Scope or Turnitin streamline the process by automating the evaluation of objective and open-ended responses. AI instantly grades multiple-choice and true/false questions, providing immediate feedback to students. For essays or problem-solving questions, AI systems assess aspects like grammar, coherence, and structure, offering preliminary scores that educators can review and adjust if necessary. This not only accelerates grading but ensures consistent, objective evaluations, reducing potential bias or human error.

Another significant benefit is the real-time feedback AI offers to both students and educators. AI systems provide detailed insights into student performance, highlighting areas of strength and weakness. This enables instructors to tailor future instruction and interventions more effectively, enhancing learning outcomes. Additionally, AI allows for more frequent assessments, promoting continuous learning and improvement. Overall, AI makes the examination process more efficient, fair, and adaptable to diverse learning needs.

Limitations and Challenges of AI in Examination Process

The use of AI in the university examination process brings numerous benefits, there are several limitations and challenges that need to be addressed to ensure effective implementation.

Lack of Nuance in Grading: One of the primary limitations of AI is its difficulty in assessing the nuanced and subjective elements of student responses, particularly in essays or creative work. AI tools like Turnitin or E-rater are efficient at evaluating technical aspects such as grammar, spelling, and structure, but they struggle to understand deeper contexts, complex arguments, or creative insights. This lack of nuance can result in oversimplified assessments, where the AI fails to recognise innovative thinking or interpretive richness that a human grader might appreciate. For example, in disciplines like literature or philosophy, where subjective interpretation is key, AI can miss critical elements that contribute to a student's overall performance.

Bias and Fairness Concerns: Though AI is often touted as a more objective alternative to human grading, it is not entirely free from bias. AI systems are trained on large datasets, and any biases present in the training data can be perpetuated in the grading process. For instance, if the training data predominantly represents one demographic or writing style, the AI may unfairly penalise students who do not conform to those norms, potentially leading to biased assessments. This could be particularly problematic for students from diverse cultural or linguistic backgrounds who may approach questions in ways not anticipated by the AI system.

Data Privacy and Security: Another major concern with AI in examinations is the issue of data privacy. AI systems collect and process large amounts of student data to improve their performance. However, the collection and storage of this data pose risks, especially if security measures are not robust. Universities must ensure that AI tools comply with data protection regulations to safeguard student information from breaches or misuse. This concern extends to the ethical use of data and how it might be employed beyond grading, such as in predictive analytics or profiling students' academic performance.

Over-reliance on Technology: AI is a valuable tool for enhancing the efficiency of the examination process, but over-reliance on AI can diminish the role of human judgment. The integration of human moderation in AI content is a complex ye

essential aspect of online content management While AI excels in automating routine tasks, it lacks the human ability to empathise, contextualize, or engage with students on a personal level. A balance between AI tools and human oversight is essential to ensure that the assessment process remains fair, ethical, and pedagogically sound. Although AI has the potential to revolutionise the university examination process, its limitations ranging from a lack of nuance in grading to data security concerns highlight the need for careful implementation. Universities must balance the use of AI with human involvement to address these challenges and ensure a fair and comprehensive assessment system.

CONCLUSION

This section has examined the use of AI systems in the examination process with particular reference to the university level of education. The specific uses of AI in the examination process like using it as a form of learning assistant, in the examination process were presented. Furthermore, the chapter presented to the reader, how AI tools could be used for examination setting (item generation), marking (grading, sorting) as well as the relevant tools that could be used in these processes. The section highlighted the benefits, university lecturers stand to gain when they deploy AI tools in the examination process. The chapter was rounded up with a clear insight into potential limitations and challenges associated with the use of AI systems in the examination process. A key takeaway from this chapter, is the cautionary note presented by the author to the effect that the use of AI tools in the education system should be seen as a kind of technological assistant to the lecturer and not a replacement for the human oversight functions in the preparation and execution stages of the examination.

RECOMMENDATION

It is recommended among others, that universities in Nigeria and beyond should embrace the use of AI systems in order to enhance service delivery, improve efficiency and ensure real-time feedback of examination scores/grades to the students. It is crucial that users of AI should ensure ethical compliance in their choice of AI tools at all levels of the examination process. Furthermore, it is hereby recommended Users of the AI systems must take due cognizance the limitations and challenges inherent in total reliance on AI tools as a replacement for the human component in chain of examinational preparation, delivery, communication of result and performance evaluation.

REFERENCE

Carl, P. (2023). What is ProProfs Quiz Maker? Using AI tools in assessment/Teaching and Learning-UCL. Retrieved from https://www.ucl.ac.uk/..ai-hub

Mouza, C., & Lavigne, K. (2013). A framework for designing and implementing effective technology-enhanced learning experiences. Journal of Educational Technology & Society, 16(4), 12-23.

Miller, A. I. (2019). AI and the Future of Work: The Implications of Artificial Intelligence in Higher Education. *International Journal of Educational Technology in Higher Education*, 16(1), 1-17.

Nicol, D. J., & Macfarlane-Dick, D. (2024). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199-218.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199-218.

Njolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company.

Weller, M. (2020). 25 Years of EdTech. Athabasca University Press.

Nasr, Y. (2021). Qcorrect,11 best exam and assessment platforms. Recommended assessment tool for online exam-Sites. Retrieved from https://sites.reading.ac.uk.

Oniye, A.O (2024). Guide to examination setting and marking: Being text of workshop presentation for academic staff of the University of Kigali on preparation of assessment items and marking guide on 11th January 2024.