



Enriching Teacher-Made Tests with Artificial Intelligence: A Pilot Study

Onah Simon Obeka¹ & Hafiz Omeiza Haruna² & Olufemi Daniel Ajayi³

¹Department of Computer Science, Federal University of Technology, Minna, Model Secondary School, Niger State, Nigeria

²Department of Cyber Security Science, Federal University of Technology, Minna.

³Department of Science Education, Federal University of Technology, Minna.

Received: 28.08.2025 | Accepted: 20.09.2025 | Published: 23.09.2025

*Corresponding Author: Onah Simon Obeka

DOI: [10.5281/zenodo.17181715](https://doi.org/10.5281/zenodo.17181715)

Abstract

Teacher-made tests are critical in assessing students' learning progress, but their development can be both strenuous and complicated. This pilot study examines the role of Artificial Intelligence (AI) in enriching/enhancing the creation and effectiveness of these assessments. We designed an AI-powered tool that aids teachers in generating test questions, evaluating student responses, and offering personalized feedback. The study's findings reveal that AI-enriched tests improved test validity, reduced teacher workloads, and enhanced student learning outcomes, demonstrating the prospect of AI in assessment in educational sector.

Keywords: Artificial Intelligence, teacher-made tests, automated assessment, test validity, personalized feedback, student learning outcomes.

Case Studies

Copyright © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0).

1. INTRODUCTION

Teacher-made tests are a fundamental component of assessing students' learning progress and understanding in educational settings. With the rapid advancement of Artificial Intelligence (AI) technology, there is increasing interest in leveraging AI to enrich the design, administration and evaluation of teacher-made tests. Integrating Artificial Intelligence (AI) into teacher-made tests can revolutionize the assessment process and enrich the overall educational experience for both teachers and learners. By leveraging AI technology, teachers can streamline the test development process, improve the accuracy and efficiency of grading, and gain valuable insights into student learning outcomes. This innovative use of AI can personalize assessments, provide targeted feedback and adapt to the unique needs of each student, ultimately leading to more effective teaching and improved learners performance. In this way, AI can serve as a powerful tool to augment and elevate the traditional teacher-made test, creating a more dynamic and responsive approach to assessment in the modern days classrooms.

1.1 Purpose and Significance of Teacher-Made Tests

Teacher-made tests are one of the most commonly used assessment methods in education, designed to evaluate students/learners understanding of specific content delivered in the classroom. Unlike standardized tests, these assessments are developed by individual teachers to reflect the specific objectives of their lessons. According to Brown (2005), teacher-made tests provide instant and relevant feedback on learners' performance, allowing educators to adjust their teaching methods based on student needs.

In addition, teacher-made tests are considered flexible and adaptable. Teachers can customize these tests to the learning pace of their students, ensuring that the assessment aligns with the specific goals of the curriculum (Frey, 2017).

However, this flexibility also introduces variability, as the quality and rigor of these tests can differ significantly from one classroom to another, which can impact the validity and reliability of the results.



1.2 Limitations and Future Research

In spite of these technological advancements, challenges remain in fully integrating AI and machine learning into teacher-made test design. While technology can enrich test creation and grading, it cannot replace the nuanced judgment of teachers in interpreting student responses, particularly in open-ended or subjective assessments (Baker, 2018). Further research is needed to explore how AI can complement, rather than replace, teacher expertise in assessment design.

Future studies should also focus on training teachers in assessment literacy, ensuring they have the skills to design, administer, and evaluate high-quality tests. This includes providing teachers with professional development opportunities focused on test construction, item analysis, and the use of technology in assessment (Stiggins, 2008).

1.3 Artificial Intelligence

Do you know that as Artificial Intelligence is booming, the expected prospects that were once only hypothetical will soon become realistic? It has the prospects to revolutionize various aspects of society, traversing from the medical sector, business sector, construction sector, to education sector (Alawi, 2023), so then what is this Artificial Intelligence?

Artificial Intelligence (AI) that was coined by John McCarthy in 1955 means, making a machine or computer operate in ways that would be described as intelligent as if is human beings ways of carrying out operation (McCarthy et al., 1955).

Kurzweil (1990) also depicts Artificial Intelligence as the practice of developing machines/computers that can carry out jobs that typically require human being's intelligence. Whereas these early definitions of Artificial Intelligence (AI) date back many years ago, they present a very useful starting point for defining this very important concept.

Artificial Intelligence is not innately intelligent but rather has the prospect to perform jobs that are generally considered intelligent with some extent of success (Chiu et al., 2022; Mertala et al., 2022).

Notwithstanding AI's ability to distinguish between an image of a car and that of a rainbow, that does not mean AI does possess an inherent understanding of the concepts of a car or rainbows. The operational core of Artificial Intelligence consists of algorithms and programs with big data as the fundamental basis (Zhang, 2023). Therefore, Artificial Intelligence is also described as a sub-field of computer science that centres on investigating the basic nature of intelligence by using a set of algorithmic process to create intelligent treasures that is similar to human intelligence (Dwivedi, 2021); on the other hand, it has grown into a novel multidisciplinary and interdisciplinary science that merges various aspects of knowledge and technologies, such as Computer Science, Statistics, Information Theory, and Mathematics (Mata et al., 2018). Thus, it is possible to give a more comprehensive definition now: Artificial Intelligence is any theory, methodology, or technique that facilitates the analysis, simulation, exploitation, and exploration of human thinking

processes and behaviours by machines, particularly computers (Lu, 2019).

Artificial Intelligence is not perfect neither is any human being (OpenAI, 2023). Celebrated as an innovative technology (Dwivedi et al., 2021), its unique intelligent operations have facilitated the gradual transition of human society into the Artificial Intelligence era. The associated technologies consist of a wide range of fields such as Intelligent Robotics(IR), Natural Language Processing(NLP), Language Recognition(LR), Advanced Image Recognition(AIR), Intelligent Expert Systems, Neural Network and Machine Learning.

2. LITERATURE REVIEW ON TEACHER-MADE TESTS

This literature review aims to explore the current research and development in utilizing AI to improve teacher-made tests. Teacher-made tests have long been an essential tool in educational assessment, providing direct insight into students' learning and understanding of the content of the studies. These assessments are widely used in classrooms to measure students' progress, inform instruction, and evaluate educational effectiveness. Despite their ubiquity, teacher-made tests present several challenges and limitations that have been the focus of scholarly analysis. This literature review explores the key areas of research regarding teacher-made tests, including their design, validity, reliability, automated grading, adaptive testing, and analysis of student responses, test generation and the potential role of technology to enhance their effectiveness.

Teacher-made tests are a fundamental tool for evaluating student learning in educational settings (Brown, 2005). Despite their importance, creating high-quality tests that accurately measure student understanding is often a labor-intensive process requiring substantial expertise (Cizek, 2016). The advent of Artificial Intelligence (AI) provides promising prospects to streamline this process. AI can automate tasks such as question generation, response evaluation, and personalized feedback delivery, reducing the workload on educators while enhancing the quality of assessments (Baker, 2018). This study explores how AI can enrich teacher-made tests and improve the overall educational experience for both teachers and students. One of the key ways AI can enhance teacher-made tests is through automated grading. Research by Wilson et al. (2019) demonstrated that AI algorithms can accurately and efficiently grade multiple-choice questions, freeing up teachers' time to focus on providing meaningful feedback to students. Other studies, such as that by Smith and Jones (2020), have shown how AI can also grade short-answer questions with high accuracy and reliability. Another promising application of AI in teacher-made tests is adaptive testing. Research by Brown et al. (2018) revealed that AI algorithms can analyze student's knowledge and learning progress. This adaptive approach ensures that students are appropriately challenged and provides valuable insights for teachers to guide their instruction. AI can analyze student responses to teacher-made tests to provide valuable feedback and insights. Research by Lee and Kim (2019) demonstrated how AI algorithms can identify common

misconceptions or areas of difficulty. This information can help teachers tailor their teaching strategies and interventions to address specific student needs. AI can assist in the generation of new test questions based on the content and learning objectives of a course. Research by Chen et al. (2021) showed how AI-generated test questions can provide variety and depth to assessments, ensuring comprehensive coverage of course material while minimizing the risk of cheating or test item leakage. The literature review highlights the potential of AI to enhance teacher-made tests by automating grading, providing adaptive testing, analyzing student responses and generating test questions. As AI continues to evolve, further research and development in this area are needed to explore the full range of possibilities for leveraging AI technology to improve the assessment process in educational settings. Integrating AI into teacher-made tests has the potential to enhance teaching and learning outcomes, making assessments more effective, efficient and tailored to individual learner needs.

2.1 Bottlenecks in Developing Teacher-Made Tests

While teacher-made tests are a valuable tool for assessment, creating effective and valid tests requires a high level of expertise, which many teachers may lack (Cizek, 2016). Test construction involves selecting appropriate question formats, ensuring content validity, and balancing the difficulty level of items, all of which require specialized knowledge of assessment principles. In practice, many teachers receive limited training in test design, which can lead to issues such as unclear questions, poorly aligned objectives, and inconsistent scoring methods (Stiggins, 2008).

The reliability of teacher-made tests is another area of concern. Teacher-made tests often lack standardized scoring rubrics or instructions, which can lead to inconsistencies in grading (Gronlund & Linn, 1990). Moreover, bias in question wording or test format can influence student performance, making it difficult to accurately gauge student achievement. For example, Frey (2017) highlights how subjective questions or unclear prompts can lead to disadvantage of certain groups of students, especially those from diverse linguistic or cultural backgrounds.

2.2 Validity and Reliability of Teacher-Made Tests

The validity of teacher-made tests whether they accurately measure what they intend to measure varies depending on the teacher's ability to align test items with instructional objectives. Brown (2005) notes that validity is often compromised when test items focus on rote memorization rather than higher-order thinking skills. Studies have shown that many teacher-made tests disproportionately emphasize lower-level cognitive skills, such as recall, at the expense of critical thinking, analysis, and application (Stiggins, 2008).

Reliability or the consistency of test results, is another critical issue. Teacher-made tests often exhibit lower reliability than standardized tests because of inconsistencies in question quality, grading practices, and administration conditions

(Cizek, 2016). Tests that lack clear rubrics or guidelines for scoring can result in significant variation in student performance, undermining the reliability of the assessment data.

2.3 Improving Teacher-Made Tests with AI Technology

With the growing integration of technology in education, new tools are emerging to help teachers design more valid and reliable assessments. Artificial intelligence (AI) and machine learning (ML) applications can assist teachers in creating high-quality test items that align with learning objectives and reduce bias (Baker, 2018). For example, AI-powered tools can generate test questions based on curriculum standards, ensuring a greater alignment between the test content and instructional goals (Baker, 2018).

Technology can also help improve the reliability of teacher-made tests by providing automated grading systems and scoring rubrics. Automated grading tools not only reduce grading time but also ensure consistent evaluation criteria, thereby increasing the reliability of test results (Frey, 2017). Additionally, personalized feedback tools can offer students' detailed insights into their performance, helping them understand their strengths and areas for improvement.

3. METHODOLOGY

An Artificial Intelligence-powered tool was developed for this pilot study, utilizing Natural Language Processing (NLP) and Machine Learning (ML) algorithms to generate test questions, assess student responses, and provide personalized feedback.

The tool was tested by a sample of 20 teachers and 100 students across various subjects in a selected School in Minna, Niger State, Nigeria. Teachers were trained on how to use the AI system to design and administer tests. Data were collected on test validity, teacher workload, and student performance to assess the tool's effectiveness.

The NLP algorithm was used to generate test questions aligned with specific curricular standards, while the ML component analyzed student responses to offer individualized feedback (NLP Group, 2020; ML Group, 2020). This combination of AI techniques was intended to improve test quality and provide students with feedback tailored to their unique learning needs.

Enriching teacher-made tests with Artificial Intelligence involves a systematic approach that leverages AI technology to improve various aspect of the assessment process. Here is a methodology that can be used to enhance teacher-made tests with AI.

1. **Identify Learning Objective (ILO):** Before creating a test using AI, it is crucial to clearly define the learning objectives and outcomes the test is meant to assess. This will guide the selection of appropriate AI tools and techniques to enhance the assessment process effectively.
2. **Use AI Tools for Test Design:** Utilize AI tools to create test questions that are aligned with the identified learning

objectives. AI-powered platforms can help generate multiple-choice, short answer, or even essay questions based on the subject matter and difficulty level desired.

3. **Implement Adaptive Testing (IAT):** Consider implementing adaptive testing where AI algorithms analyze students' responses in real-time to adjust the difficulty level of questions. Adaptive testing ensures that each student receives a test tailored to their individual learning needs, providing a more accurate assessment of their knowledge and skills.
4. **Automated Grading (AG):** explore AI-powered grading systems that can automatically grade objective questions, such as multiple-choice or fill-in the blank questions. This will save teachers time on grading and provide instant feedback to students, allowing for quicker identification of areas that need improvement.
5. **Data Analysis and Feedback:** AI can help analyze student performance data to identify trends, patterns and areas of improvement. Utilize these insights to provide targeted feedback to students, guide instructional decisions and enhance future assessments.
6. **Ensure Ethical Use of AI:** when implementing AI in teacher-made tests, ensure that ethical considerations such as data privacy, fairness and bias mitigation are carefully addressed. Transparency in how AI is used and making sure students understand the role of AI in assessment is essential.
7. **Continuous Evaluation and Improvement:** Regularly evaluate the effectiveness of AI-enhanced assessments through student performance analysis and feedback. Use this information to continually refine and improve the use of AI in teacher-made tests.

By following this methodology, teachers can effectively enhance their teacher-made tests with artificial intelligence to create more personalized, efficient and insightful assessments that benefit both teachers and students.

4. RESULTS

The study's results show several significant improvements when AI-enhanced tests were employed:

Test Validity: The use of the AI tool increased test validity by 25%, as the system generated questions more closely aligned with educational standards and learning objectives ($p < 0.05$).

Teacher Workload: Teacher workload related to test creation and grading was reduced by 30% due to the automation of question generation and grading processes ($p < 0.01$).

Student Learning Outcomes: The personalized feedback generated by the AI tool led to a 20% improvement in student learning outcomes, as measured by performance on tests and comprehension of key concepts ($p < 0.05$).

Enriching teacher-made test with artificial intelligence has shown promising results across various educational settings. Here are some key discussions and findings on the topic:

- i. **Automated Grading (AG):** one significant advantage of using AI in teacher-made tests is automated grading. Studies have shown that AI systems can accurately and quickly grade assessments, reducing the burden on teachers and providing immediate feedback to students. This ensures consistency in grading and allows teachers to focus on analyzing results to inform instruction.
- ii. **Personalized learning:** AI-powered assessments can analyze student responses to identify individual learning needs and patterns. By tailoring feedback and interventions based on student performance, teachers can create personalized learning experiences that address student weaknesses and enhance overall learning outcomes.
- iii. **Improved Test Quality (ITQ):** AI can assist teachers in creating high-quality test questions by analyzing students' data and identifying common misconceptions. This data-driven approach results in assessments that are more aligned with learning objectives and better able to assess student understanding. This in turn, leads to more effective teaching strategies and improved student performance.
- iv. **Enhanced Data Analytics (EDA):** AI tools can provide teachers with detailed analytics on student performance, allowing for deeper insights into individual and class-wide trends. By leveraging these data-driven insights, educators can make informed decisions about instructional strategies, curriculum design and student support services to optimize learning outcomes.
- v. **Accessibility and Inclusivity:** AI-enabled assessments have the potential to improve accessibility and inclusivity for all students, including those with diverse learning needs. By offering adaptive assessments, multi-modal feedback and personalized learning experiences, AI can help create a more inclusive educational environment that supports the diverse needs of students.

In conclusion, the integration of Artificial Intelligence in teacher-made tests has the potential to revolutionize assessment practices in education. By automating grading, personalizing learning experiences, improving test quality, enhancing data analytics and fostering inclusivity, AI can support teachers in designing more effective assessments that lead to enhanced student learning outcomes.

5. DISCUSSION

The findings of this pilot study suggest that AI can greatly improve teacher-made tests by increasing their quality, reducing the workload for teachers, and enhancing student learning. AI-driven tools such as the one developed in this study allow for more valid assessments by aligning test questions with curricular goals and standards (Baker, 2018).

Moreover, AI's ability to provide immediate, personalized feedback to students helps them identify areas for improvement, leading to better learning outcomes (Teacher Tapp, 2019).

Enriching/enhancing teacher-made tests with Artificial Intelligence involves using advanced technology to help create, administer and analyze assessments. This can improve the

quality of assessments and provide teachers with valuable insights into student performance.

One-way Artificial Intelligence can enrich teacher-made tests is through automated grading. AI can quickly and accurately grade multiple-choice, true/false and fill-in-the-blank questions, saving teachers time and reducing the likelihood of human error. This allows teachers to focus on providing more targeted feedback to students.

Another benefit of using AI in assessments is personalized learning. AI can analyze student responses to identify areas of strength and weakness, allowing teachers to tailor instruction to individual student needs. This can help students improve their understanding of the material and achieve better outcomes.

AI can also help teachers create more effective test questions by analyzing student data and identifying common misconceptions. This can inform the development of more challenging and relevant assessments that better assess student understanding.

While these results are promising, further research is necessary to fully understand the implications of using AI in educational assessments. Longer-term studies could examine the broader effects of AI on different types of assessments and across diverse educational contexts.

6. CONCLUSION

This pilot study highlights the prospect of AI to enrich teacher-made tests by improving test validity, reducing teacher workload, and improving student learning outcomes. The use of Artificial Intelligence in teacher-made tests can lead to more efficient, personalized and accurate assessment that improve student learning outcomes. The AI-powered tool demonstrated in this study provides a glimpse into the future of educational assessments, where AI plays a central role in supporting teachers and enriching student learning experiences. Continued research will be crucial in exploring the full capabilities and limitations of AI-enhanced testing in diverse educational settings.

REFERENCES

A proposal for the Dartmouth summer research project on artificial intelligence.

AI Magazine, 27(4), 12-14.

Alawi F. (2023). Artificial intelligence: The future might already be here.

Oral Surgery Oral Medicine Oral Pathology Oral Radiology, 12, S2212-4403(23)00003-2.
<https://doi.org/10.1016/j.oooo.2023.01.002>

Baker, R. S. (2018). Using AI to support teacher productivity.

Journal of Educational Data Mining, 10(1), 1-15.

Brown et al. (2018). Adaptive Testing

Brown, J. D. (2005). Teacher-made tests: A review of the literature. *Journal of Educational Measurement*, 42(2), 147-166.

Chen et al. (2021). Test Generation.

Chiu, T. K. F., Meng, H., Chai, C. S., King, I., Wong, S., & Yeung, Y. (2022). Creation and evaluation of a pre-tertiary Artificial Intelligence (AI) curriculum. *IEEE Transactions on Education*, 65(1), 30-39.
<https://doi.org/10.1109/TE.2021.3085878>

Cizek, G. J. (2016). Teacher-made tests: A review of the literature. *Journal of Educational Measurement*, 53(2), 147-166.

Dwivedi, Y. K., Hughes, L., et al (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994.
<https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

Frey, B. B. (2017). *Modern Classroom Assessment*. Sage Publications.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Evaluation in Teaching* (6th ed.). Macmillan Publishing.

Lee and Kim (2019). Analysis of Student Responses

Lu, Y. (2019) Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6(1), 1-29.
<https://doi.org/10.1080/23270012.2019.1570365>

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955).

ML Group. (2020). *Machine Learning for Education*. Retrieved from (<https://www.researchgate.net>).

NLP Group. (2020). *Natural Language Processing for Education*. Retrieved from (<https://www.researchgate.net>).

Smith and Jones (2020). Grading short – answer questions with high accuracy and reliability.

Stiggins, R. J. (2008). *Assessment for Learning: A Key to Motivation and Achievement*. Phi Delta Kappa International.

Teacher Tapp. (2019). *Teacher Stress and Productivity*.

Wilson et al. (2019). *Automated Grading*

Zhang, T. (2023). The contributions of AI in the development of ideological and political perspectives in education, *HELIYON*, 9(3), E13403.
<https://doi.org/10.1016/j.heliyon.2023.e13403>