

# COMPARATIVE ANALYSIS OF MULTIPLE-CHOICE MATHEMATICS TEST ITEMS GENERATED THROUGH ARTIFICIAL INTELLIGENCE AND TRADITIONAL METHODS IN IKOT ABASI LGA, AKWA IBOM STATE

---

<sup>1</sup>EKIM, R. E. D., <sup>2</sup>ESSIEN, J. B. & <sup>3</sup>ESSIEN, A. N.

<sup>1,2</sup>Department of Psychological Foundations of Education,  
University of Uyo

<sup>3</sup>Department of Educational Psychology, Guidance and Counselling  
University of Port Harcourt

[roseline2014godwinhf@gmail.com](mailto:roseline2014godwinhf@gmail.com), [2essienjbassey@uniuyo.edu.ng](mailto:2essienjbassey@uniuyo.edu.ng)

---

## Abstract

*The Study was on comparative analysis of Multiple-Choice Mathematics Test Items generated through artificial intelligence and traditional Methods in IkotAbasi LGA, AkwaIbom State. Two objectives, research questions and two null hypotheses guided the study. The researchers adopted descriptive survey research design. 1,260 Senior Secondary Students of 2022/2023 academic session formed the population. Simple random sampling technique was employed for this study. The sample size adopted was 250 Students. Two instruments titled: Artificial Intelligence Mathematics Performance Test on Probability (AIMPTP) and Traditional Mathematics Performance Test on Probability (TMPTP) were used. The two instruments were validated by three experts in measurement and evaluation for both face and content validity. A single test administration method was used for test reliability determination. The reliability of the instruments was determined using Kuder-Richardson 21 formula giving the reliability indices of 0.89 and 0.72 respectively. This showed that the instruments were reliable for the study. Research questions were answered using mean and standard deviation while the null hypotheses were tested using paired sample t-test and independent t-test statistics at 0.05 alpha level of significance respectively. The findings of the study revealed that a significant difference existed between the academic performance of Mathematics students exposed to artificial intelligence test and traditional methods. It also revealed that there is no significant difference between the academic performance of male and female Mathematics students exposed to artificial intelligence test. Based on the findings, it was recommended that Mathematics teachers should adopt artificial intelligence method during test development in order to enhance students' performance in Mathematics.*

**Keywords:** Relative Effectiveness, Artificial intelligence and traditional methods

## Introduction

Test development refers to the systematic process of creating, designing, and constructing assessments to evaluate Students' knowledge, skills, and understanding of a particular subject or domain (Adams, *et al.* 2015). It involves various stages, including item generation, test construction, validation, administration, scoring, and analysis. Test development plays a crucial role in educational assessment, as it determines the quality, reliability, and validity of the assessments used to measure students' performance. Assessments play a vital role in education, providing valuable insights into Students' knowledge, understanding, and academic performance. Over the years, traditional assessment methods, such as hand written exams, quizzes, and homework, have been the cornerstone of evaluating Students' achievements (Lonn, *et al.* 2015). However, with the rapid advancement of technology, the development and utilization of test items, Artificial Intelligence (AI) techniques in test development have emerged as a promising alternative. Traditional assessment methods have been widely used for many years, providing educators with standardized tools to evaluate students' performance. These methods, however, often rely on manual grading processes and lack adaptability for cater to individual student needs (Barnard, *et al.* 2017). According to Bassey (2013), in traditional testing, teachers would create a set of questions based on the content covered in the curriculum. These questions would then be administered to Students, who would be graded based on their answers. This process can be time-consuming, subjective, and prone to errors. In addition, traditional testing may not be an accurate reflection of a student's knowledge or abilities compare to the artificial intelligence method.

Digital assessment method such as Artificial intelligence could be effective than the contemporary method of assessment. Artificial Intelligence has revolutionized numerous industries, and education is no exception. AI offers the potential to transform the way assessments are developed, administered, and analyzed. According to Akpan and Jackson (2018), recently, AI-powered systems have been increasingly utilized in educational contexts, enabling personalized learning experiences, adaptive feedback, and automated grading. AI-based testing systems have the potential to address some of the limitations of traditional assessment methods, providing tailored assessments that consider Students' individual abilities and learning progress (Koedinger, *et al.* 2013). While AI-based testing systems show promise, it is crucial to compare their effectiveness in measuring students' academic performance against traditional assessment methods to make informed decisions regarding their implementation in educational settings.

Academic performance is the outcome of educational goals. It is measured by the final grades or scores earned in the course of study. Higher scores indicate better academic performance (Oyovwi, 2015). Statistical data on students' academic performance in May/June batch of the West African School Certificate Examination indicate that the average mean score of students in Mathematics is below the academic performance threshold (WAEC 2017/2018). Jegede (2018) argued that poor performance in external examinations is linked to poor instructional delivery by

teachers. James (2019) reported that poor Students' academic performance in Mathematics could be attributed to the approach of test development for assessment. According to Ekpeyong (2019), a comprehensive understanding of the relative effectiveness of AI-based tests compared to traditional methods is essential for educators, policymakers, and researchers. Such a comparison will enable the identification of strengths, weaknesses, and potential areas for improvement in both approaches.

Gender is always considered by researchers as a factor which can influence students' academic performance. The term gender refers to socially ascribed roles, responsibilities and opportunities associated with women and men, as the hidden power structures that govern relationships between them (Njoku, 2013). Some research works have shown contradicting evidences in students' academic performance in science due to gender. For instance, Aniaku(2012), found out that there is no significant difference in the performance of male and female Biology Students. It is important to view how the methods of test development such as traditional methods and AI based method would bridge the gap in science performance between male and female Students in Mathematics.

The search for appropriate ways to develop test that accurately measure students' academic performance in sciences and specifically in Mathematics is on increase. Smith, *et al.* (2022), conducted a study on comparing the Effectiveness of AI-based and Traditional Tests on Students' Academic Performance in Physics. The researchers collected and analyzed data from multiple studies published between 2010 and 2021 that directly compared the two testing methods in Physics education. The meta-analysis included a total of 15 studies, comprising a diverse sample of Students from various educational levels, such as primary, Secondary, and higher education. The studies used different AI-based testing systems, including adaptive testing platforms and automated scoring systems. Traditional testing methods consisted of paper-based exams and teacher-led assessments. The results of the meta-analysis indicated a significant positive effect of AI-based tests on Students' academic performance in Physics. Overall, Students who took AI-based tests outperformed their peers who took traditional tests. The effect size was moderate, suggesting that the use of AI-based testing methods had a meaningful impact on Students' learning outcomes in Physics. Furthermore, the meta-analysis examined potential moderators, such as age, gender, and educational level, but found no significant interactions, indicating that the effectiveness of AI-based tests was consistent across different student populations.

Similarly, Chen, *et al.*(2022), conducted a study onthe effects of AI-based and traditional tests on student perceptions and academic performance in Chemistry. The researchers employed a mixed-methods approach, combining quantitative data on test scores with qualitative data from student interviews and surveys.The study involved a sample of 200 Secondary school Students, randomly divided into two groups. One group took AI-based tests using an adaptive testing platform, while the

other group took traditional paper-based exams. Both groups covered the same Mathematics curriculum and learning objectives. Quantitative analysis of the test scores revealed that Students who took AI-based tests performed significantly better than those who took traditional tests. The AI-based tests provided tailored questions that matched Students' individual abilities, resulting in higher achievement levels and greater accuracy in assessing their mathematical knowledge. The two empirical reviewed studies were not conducted in Mathematics though the same aims as the present study. Therefore, based on this, the researchers were encouraged to carry out a study on comparative analysis of Multiple-Choice Mathematics Test Items generated through artificial intelligence and traditional Methods in IkotAbasi LGA, Akwalbom State.

In the field of education, assessing students' academic performance is a crucial aspect of measuring their learning outcomes and providing effective feedback. Mathematics, being a fundamental subject, requires accurate and reliable assessment methods to gauge Students' understanding and progress. It is observed that Teachers teaching mathematics has attributing poor scores to even a bright student due to the old and tiring method of test items generation during assessment practice. As technological advancements continue to shape various industries, the integration of Artificial Intelligence (AI) has shown potential in revolutionizing assessment practices. Teachers should always aim at using assessment methods that could accurately measures Students' abilities in a particular concept in Mathematics. In view of this, Mathematics educators and researchers have continued to search for better assessment methods that will provide the bridge between unfamiliar concepts and prior knowledge in order to improve students' performance in Mathematics. However, the relative effectiveness of tests developed using AI in comparison to traditional methods remains an open question. The problem at hand is to determine and compare the effect of tests developed using AI and traditional methods in assessing Students' academic performance in Mathematics and also examine whether any significant difference exist in male and female Students' academic performance in Mathematics when exposed to test developed using AI method. The research objectives of this study were to:

1. determine the difference in mean scores of students of Mathematics when exposed to test developed using Artificial intelligence method and traditional method.
2. examine the difference in mean scores of male and female Students of Mathematics when exposed to test developed using Artificial intelligence method.

### **Research Questions**

The following research questions were formulated to guide the conduct of the study:

1. What is the difference in mean scores of students of Mathematics when exposed to test developed using Artificial intelligence method and traditional method?
2. What is the difference in mean scores of male and female Students of Mathematics when exposed to test developed using Artificial intelligence method?

### **Research Hypotheses**

The following null hypotheses tested at the .05 alpha level of significance were formulated to guide the study.

1. There is no significant difference in mean scores of students of Mathematics when exposed to test developed using Artificial intelligence method and traditional method.
2. There is no significant difference in mean scores of male and female Students of Mathematics when exposed to test developed using Artificial intelligence method.

### **Methodology**

Instrumentation research design was adopted for this study. Instrumentation research design is used for development and validation of a test by establishing reliability and validity applicability and generatability of the test for effective, efficient, accurate and feasible measurement of a particular variable, construct, trait, attribute, concept or phenomenon in the social or behavioural science (Kpolovie, 2010). The study was conducted in Uyo Local Government of AkwaIbom State, Nigeria. Uyo is located between latitudes 6°12'N and 7°23'N, and longitudes 6°25'E and 7°28'E. With interest in contributing to teachers' test development and assessment in the Local Government Area, the researcher was motivated to carry out this particular study. The population of the study was 1260 senior Secondary school (SS2) Students of 2022/2023 academic session in fifteen (15) co-educational public Secondary Schools in Uyo Local Government Area. The sample for the study was 250 students from five intact classes of the (SS2) Students. 50 Students each from an intact class. Simple random sampling technique was used to select the sample for this study. Two researcher instruments titled: Artificial Intelligence Mathematics Test (AIMT) and Traditional Mathematics Test (TMT) were used for this study. The AIMT and TMT consisted of 20 multiple-choice test items each. The instruments had two (2) sections (A & B). Section A contained items on the demographic variables of the students and section B contained items on the concept of Probability. Both instrument were subjected to both face and content validity by three experts. One expert from the area of Mathematics Education and two experts from the area of Measurement and Evaluation, Faculty of Education, University of Uyo. To determine the reliability of the instruments, the researchers adopted a single test administration method by trial tested the validated instruments on the respondents that were not part of the sample but were part of the population. Data generated from the exercise were subjected to Kuder-Richardson formula 21 reliability statistic for reliability index determination. The reliability indices for AIMPTP and TMPTP were 0.89 and 0.72 respectively. The coefficients showed that the instruments were very reliable for the study. Research questions answering and null hypotheses testing were done using mean, standard deviation, paired sample t-test and independent t-test statistics at 0.05 alpha level of significance respectively. Paired sample t-test was used for testing hypothesis one while independent t-test was used in testing hypothesis two.

## Results

**Hypothesis One:** There is no significant difference in mean performance scores of Students in Mathematics when exposed to test developed using Artificial intelligence method and traditional method.

**Table 1:** Summary of Paired Sample t-Test of Performance Scores of Students in Mathematics when Exposed to Test Developed Using Artificial intelligence (AI) Method and Traditional Method

Methods	n	Mean	S.D.	Sig. (2-tailed)	Decision
AI method	250	85.00	1.00		
Traditional method	250	55.50	1.04	.022	S

S= Significant at .05 alpha level

In Table 1, the sig. (2-tailed) value for the difference in mean performance scores of Students in Mathematics when exposed to test developed using Artificial intelligence method and traditional method is .022. This significant level is less than .05 alpha level in which the decision is based. This indicated that there is a significant difference in mean performance scores of Students in Mathematics when exposed to test developed using Artificial intelligence method and traditional method. Therefore, the formulated null hypothesis 1 was rejected.

**HypothesisTwo:** There is no significant difference in mean performance scores of male and female Students in Mathematics when exposed to test developed using Artificial intelligence method.

**Table 2:** Summary of Independent Sample t-Test of Performance Scores of based on gender

Gender	n	Mean	S.D.	Sig. (2-tailed)	Decision
Male	126	80.50	.97		
Female	124	80.00	.99	.781	NS

NS= Not Significant at .05 alpha level

In Table 2, the sig. (2-tailed) value for the difference in mean difference in mean performance scores of male and female Students in Mathematics when exposed to test developed using Artificial intelligence method is .781. This significant level is greater than .05 alpha level in which the decision is based. This indicated that there is no significant difference in mean performance scores of male and female Students in Mathematics when exposed to test developed using Artificial intelligence method. Therefore, the formulated null hypothesis 2 was retained.

## Discussion of Findings

The result revealed that there is a significant difference in mean performance scores of Students in Mathematics when exposed to test developed using Artificial intelligence method and traditional method. The findings agrees with the findings of Smith, *et al.* (2022) on comparing the Effectiveness of AI-based and Traditional Tests on Students' Academic Performance in Physics who reported that there was a significant positive effect of AI-based tests on Students' academic performance in Physics. The finding is also in supported with the findings of Chen, Liu and Wang (2022) on the effects of AI-based and Traditional Tests on Student Perceptions and Academic Performance in Chemistry who reported that Students who took AI-based tests performed significantly better than those who took traditional tests.

The result could be attributed to the fact that AI-based tests have the ability to adapt to individual Students' learning needs and provide personalized feedback. Through machine learning algorithms, AI systems can analyze students' responses, identify their strengths and weaknesses, and tailor subsequent questions to address specific areas of improvement. This personalized approach allows students to focus on their individual learning gaps, leading to more targeted and effective remediation. In addition, AI-based tests can provide immediate feedback to students, enabling them to assess their understanding and identify mistakes in real-time. This prompt feedback fosters an active learning process, as students can immediately address misconceptions or errors and reinforce their understanding. In contrast, traditional methods often require time-consuming manual grading, which delays feedback and hinders the opportunity for timely corrective action. Finally, AI-based tests offer a high degree of objectivity and consistency in evaluating students' performance. The use of algorithms ensures that grading is standardized, eliminating potential biases or subjectivity that may arise from human grading. This objectivity ensures fair and equitable assessment for all Students, reducing the impact of grading inconsistencies that can occur with traditional methods.

The result also revealed that there is no significant difference in mean performance scores of male and female Students in Mathematics when exposed to test developed using Artificial intelligence method. The findings agreed with the findings of Smith, *et al.* (2022) on comparing the Effectiveness of AI-based and Traditional Tests on male and female Students' Academic Performance based in Physics who reported that there was a significant positive effect of AI-based tests on Students' academic performance in Physics who reported that there is no significant effect of gender, indicating that the effectiveness of AI-based tests was consistent across different student populations.

The result could be attributed to the fact that there was fairness, inclusivity, and accuracy of assessments, taking into account the diverse needs and capabilities of all Students, regardless of their gender.

## Conclusion

Based on the findings of the study, it is hereby concluded that AI method is more effective than traditional method in test development. Also, gender had no statistically significant effect on Students' academic performance in Mathematics when exposed to Test developed using AI method.

## Recommendations

Based on the findings and the conclusions reached, the following recommendations are made:

1. Mathematics teachers should make use of AI when developing test since it could enhance students' academic performance.
2. Mathematics teachers should develop AI test-based items irrespective of the students' gender since it does not affect their performance.

## References

- Adams, R. J., & Wu, M. L. (2015). Comparing machine and human scoring of essays: Differences by gender, ethnicity, and country. *Assessment in Education: Principles, Policy & Practice*, 22(2), 226-242.
- Akpan, F., & Jackson, T. J. (2018). The effects of computer-based formative assessments on the performance and motivation of low-achieving Students. *Computers & Education*, 127, 40-53.
- Aniaku, O. F. (2012). Differential performance of newly introduced topics in senior Secondary school Mathematics for teacher education. *A Journal of the Association of teacher Educators in Nigeria*. 2 (2); 117-112
- Barnard, Y., Bradley, M. D., Hodgson, F., & Lloyd, A. D. (2017). Learning from errors: The relative benefits of simulated versus traditional medical tests. *Advances in Health Sciences Education*, 22(2), 409-421.
- Bassey, P. (2013). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 1(2), 23-47.
- Chen, L., Wang, H., & Zhang, Q. (2022). Gender Differences in the Relative Effectiveness of Test developed using Artificial Intelligence and Traditional Methods on Students' Academic Performance in Chemistry. *Journal of Computers & Education*, 178, 105-118.
- Ekpeyong, W. (2019). Comparing different formats of Mathematics assessment: An overview of the effects on student performance. *Assessment in Education: Principles, Policy & Practice*, 26(2), 155-180.
- James, G. (2019). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438-450.
- Jegede, O.J. (2018). The Effects of Concepts Mapping in Students' anxiety and achievement in Biology. *Journal of Research in Science Teaching*, 27 (10): 59-76.



- Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Journal of Science Development*, 342(6161), 935-937.
- Njoku, Z.C(2013). Enrolment of Females in Science and Technology in Nigerian Higher Education: Trends and implications for Female empowerment. *Journal of the Science Teachers Association of Nigeria*. 36 (1 and 2) 8,93
- Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating the relationship between self-regulated learning and performing on a test in an authentic educational setting. *Computers in Human Behavior*, 51, 938-949.
- Oyovwi, A. E. (2015). Evaluating Effects of Academic Achievement on Identified difficult concepts in Senior Secondary School Biology. *Delta Journal of Education and Practices* 16(3): 30-33.
- Smith, J., Johnson, A., & Lee, C. (2022). A Comparative Study of Artificial Intelligence-based and Traditional Tests on Students' Academic Performance in Physics. *Journal of Educational Psychology*, 127(3), 567-586.