International Journal of Education, Culture, and Society

e-ISSN : 3024-8973
p-ISSN : 3024-8981

Index: Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/IJECS.v4i3.9375

Comparing Mathematics Achievement across High-Stakes Examinations: Evidence from WAEC/NECO and JAMB in Nigeria

Hassan Muhammad
Federal University of Education Kontagora, Niger State, Nigeria
profmaths781@fcekg.edu.ng

Article Info:

Submitted:	Revised:	Accepted:	Published:
Mar 28, 2026	Jun 25, 2026	Jul 7, 2026	Jul 12, 2026

Abstract

This study compares mathematics performance across essay-based secondary school examinations administered by the West African Examinations Council (WAEC) and the National Examinations Council (NECO) and the multiple-choice university entrance examination administered by the Joint Admissions and Matriculation Board (JAMB) in Nigeria. Using an ex post facto comparative design, archival mathematics examination records from 600 secondary school graduates were analyzed to determine the comparability of performance across the three examination systems. Mathematics performance in WAEC and NECO was operationalized using official grade classifications, whereas JAMB performance was assessed using subject-level aggregate score ranges. Data were analyzed through descriptive statistics, score standardization, and paired-samples *t* tests. The findings showed that a substantial proportion of candidates who attained credit-level grades in WAEC and NECO Mathematics obtained relatively low aggregate scores in JAMB Mathematics. Standardized mean comparisons further indicated that the same candidates performed better in WAEC and NECO than in JAMB, while the paired-samples analysis confirmed that the differences were statistically significant and

associated with a large effect size. These results demonstrate systematic variation in mathematics performance between essay-based and multiple-choice high-stakes examinations. The study contributes empirical evidence to the mathematics education literature on assessment comparability in multi-examination contexts and underscores the need for cautious interpretation of examination outcomes generated by assessment systems with different formats and purposes, particularly when such results inform high-stakes educational decisions.

Keywords: Assessment Comparability; High-Stakes Examinations; Mathematics Performance; Multiple-Choice Assessment; Standardized Testing

INTRODUCTION

The result of mathematics in high-stakes tests is a determining factor that determines the educational pathways and access to higher education among students (Alfieri et al., 2010). National examination results in Nigeria play a key role in determining certification, advancement and admission into university. Mathematics performance of the candidates is evaluated by various examination systems which vary significantly in terms of structure, purpose, and form of assessment (Zana et al., 2024). Western Africa The West African Senior School Certificate Examination (WASSCE), which is administered by the West African Examinations Council (WAEC), and the Senior School Certificate Examination, which is administered by the National Examinations Council (NECO) use mainly essay-based and structured-response items to assess mathematics learning in secondary schools. On the contrary, the Unified Tertiary Matriculation Examination (UTME) conducted by the Joint Admissions and Matriculation Board (JAMB) is based almost solely on multiple-choice questions to evaluate the readiness of candidates to tertiary education (Fokuo et al., 2022). Although they are commonly used, these examinations are based on various assessment systems, which creates the issue of comparing mathematics performance across systems (Cantley and McAllister, 2021).

This paper concentrates on differences in observed mathematics performance in high-stakes examinations that are based on different item formats, namely essay-based and multiple-choice testing. Cases where applicants with credit-level grades in high school math tests have low scores in the UTME are often emphasized publicly and in academic literature (Raji, 2021). These inconsistencies carry serious consequences on the admissions

and the understanding of mathematical readiness of students. Nevertheless, these observations are usually treated anecdotally, and there is little empirical research on them based on candidate-level examination records.

In mathematics education research, the aspect of assessment format has long been considered as a factor that can also affect the observed performance (O'Connor et al., 2019). Other researchers have proposed that essay-based tests could permit longer rationality, illustration of methodology actions, and part credit, but multiple-choice tests could stress recognition, speed, or strategic reacting. Meanwhile, empirical evidence on the implications of assessment format is inconclusive, and it still remains a matter of debate how much various formats produce systematically different mathematics performance results. Notably, a significant portion of evidence available is based on classroom-based tests or researcher-created tests as opposed to nationally administered, high-stakes tests.

Although the result of WAEC, NECO, and JAMB is central to educational decision-making in Nigeria, there is scanty empirical research that explicitly compares the outcome of mathematics performance performance in these exams using archival records of their candidates. Specifically, existing literature is deficient in investigating whether the performance patterns in essay-based mathematics examinations in secondary schools correlate with multiple-choice university entrance examinations of the same group of students. This gap needs to be filled to enhance knowledge on comparability of assessment and to guide interpretations of the results of mathematics examination in various testing formats.

To fill this gap, the current study observed patterns of mathematics performance in the context of the WAEC/NECO and JAMB exams among the secondary school graduates in Nigeria. The study compared mathematics performance outcomes observed through assessment systems of varying formats using records of archival examination, and specifically the direction and magnitude of the performance differences. The study offers empirical evidence in a multi-examination setting, which adds to the current discourse in mathematics education about comparability of assessments and the meaning of performance among the different high-stakes examination models.

Statement of the Problem

The use of high stakes mathematics tests is very important in establishing the educational opportunity and advancement of students. The performance of mathematics in

Nigeria is tested by various national examination systems, the most common one being WAEC/NECO at secondary school level and JAMB at the university entry level. Although candidates often achieve credit level grades in both WAEC and NECO mathematics, a significant proportion of those who take JAMB mathematics achieve low scores, which undermines the performance results of testing systems with significant differences in assessment format and scoring schemes (Raji, 2021; Cantley and McAllister, 2021).

Though the literature on the topic of differences between essay-based and multiple-choice assessments has been widespread in mathematics education (O'Connor et al., 2019), there is a dearth of empirical studies directly comparing the performance of candidates in WAEC/NECO and JAMB mathematics examinations. Specifically, the use of archival candidate-level data on the issue of whether observed performance on secondary school mathematics essays can be linked to performance on multiple choice university entrance tests is inadequate. Lack of such evidence restricts informed decoding of examination results and inhibits the comprehension of the possible relation of the format of assessment to the observed discrepancies in mathematics performance in high-stakes settings.

Significance of the Study

This research gives empirical information regarding differences in mathematics performance between WAEC/NECO and JAMB exams based on archival records of candidates. The results can be used in the area of mathematics education research to reinforce the discussion of comparability in assessment and the interpretation of performance based on different forms of examination (Zana et al., 2024). In practice, the research provides an informational source to the examination bodies, policymakers, and educators on the consequences of using the results of various high-stakes assessment systems in educational decision-making. The research is also used as a point of reference in future studies on the effect of assessment format in large-scale, nationally administered examinations.

Objectives of the Study

The objectives of the study were as follows:

i. To analyse the performance trends of mathematics in WAEC/NECO and JAMB exams among secondary school graduates in Nigeria.
ii. To compare the observed mathematics performance results between essay-based (WAEC/NECO) and multiple-choice-based (JAMB) examinations.

iii. To establish the existence of statistically significant difference in the performance of the candidates in mathematics in both formats of the examinations.

Research Questions

The study was directed by the following research questions:

i. What are the mathematics performances trends in WAEC/NECO and JAMB exams amongst secondary school graduates in Nigeria?
ii. What is the difference between the performance of mathematics in essay-based WAEC/NECO examinations and the multiple-choice-based JAMB examination?
iii. Is the difference between the performance of candidates in mathematics in WAEC/NECO and JAMB examinations statistically significant?

Mathematical Ability as an Assessment Construct

Mathematical ability is now being conceptualised as a unified construct that includes conceptual knowledge, procedural skill, reasoning and the ability of mathematics to be applied in new and diverse situations. The modern understanding of assessment as a learning process highlights that the valid measurement of this construct must be done through assignments that go beyond simple correctness to involve how learners understand, relate and justify mathematical concepts. Consequently, the design of assessment can be important in deciding which aspects of mathematical ability are brought into the realm of visibility because of test performance.

Studies in the large-scale assessment environment indicate that most mathematics tests favor close measures of competence, especially those that are linked to efficient scoring and comparability. Papanastasiou and Michaelides (2024) prove that when the test items are concentrated in the lower levels of cognitive demand, the assessment is likely to underestimate the ability of students to think complexly and solve strategic problems. In the same vein, Smith, Wang, and Liu (2020) note that examinations that are filled with practices run the risk of confusing mathematical capability with familiarity and speed, as opposed to understanding.

In addition to the cognitive demand, the inclusion of mathematical practices in the assessment activities also has obtained growing interest. Park et al. (2023) emphasize that the mathematical skill of students is directly connected to the activities that they perform including modelling, representation, and justification. In cases where testing does not

inspire these practices, the performance results can be influenced by the test format limitations and not the competence. In this view, mathematical aptitude cannot be entirely determined based on correctness but should be construed alongside the possibilities given to the students to show reasoning and use.

Moreover, the increased focus on mathematical literacy highlights the necessity of the assessment that would measure the degree to which learners apply mathematics to make sense of the real-life situation. Amalina and Vidakovich (2022) argue that mathematical ability does not merely mean the mastery of formal procedures, but also interpretation, decision making and contextual reasoning. Evaluations that fail to consider these dimensions can give a partial or skewed view of the abilities of learners.

Combined, these views indicate that mathematical ability is not a stable characteristic that is manifested in a consistent way across the assessment forms. Instead, it is constructed in a mutually dependent manner between the learners and the cognitive, contextual, and structural aspects of the assessment tasks. This knowledge gives a critical base to the analysis of variations in performance among various mathematics examination formats.

Construct Alignment and Validity of Mathematics Examinations of Large Scales

Validity in large-scale mathematics tests is the extent to which test results are valid in indicating the intended construct of mathematical ability. Construct alignment is a very important aspect of validity and it is necessary to have coherence between the curriculum goals and the cognitive demands of the test items and the inferences of the assessment results. This alignment is not easily achieved due to the fact that the traditional high-stakes mathematics examinations tend to concentrate on individual skills, as opposed to problem-solving and reasoning skills (Burkhardt and Schoenfeld, 2018). Such restricted range of content knowledge in mass-scale tests does not reflect the capacity of students to utilize mathematical thinking in practice, which is a fundamental point of mathematical literacy (Amalina & Vidakovich, 2022).

The variability in test item design adds to the problem of alignment in that most tests have low cognitive demand, not assessing higher-order problem-solving skills. To a greater extent, as Anghel et al. (2024) indicate, the more effective solution would be to make assessment items aligned with the cognitive models that capture the entire spectrum of mathematical competencies, such as problem-solving, reasoning, and application.

Moreover, massive testing tends to disregard the role of student practices, which play a crucial role in the development of mathematics in students (Park et al., 2023). The fact that the tasks which make students apply mathematical modeling and justification are not included worsens the situation with the lack of alignment between the tests and the overall learning objectives of mathematical reasoning.

Finally, construct alignment in large scale mathematics testing is crucial in enhancing validity of such tests. Cognitive correspondence between test questions and real-life mathematical practices should be put more into the focus to measure problem-solving and reasoning abilities better. Future studies are aimed at creating assessment models that would more effectively combine cognitive complexity with valid representations of mathematical competence (Provasnik, 2021).

The following is a neatly integrated replacement section of Q1 standard, which is a combination of previous Sections 2.3 (Cognitive Complexity) and 2.6 (Construct Misalignment) into a single logical section with no duplication of content, tightened logic, and improved flow preferred by the reviewer.

Mathematics Assessment Cognitive Complexity and Construct Misalignment.

In mathematics assessment cognitive complexity and construct alignment go hand in hand, especially in large-scale examinations where validity of inferences can be determined by the extent to which the assessment tasks capture the entire scope of mathematical skills. Cognitive complexity is the range of mental activities needed to approach mathematical problems, with one end having simple recall and routine actions, and the other end having very complicated activities such as analysis, synthesis, reasoning, and evaluation. When tests are based on the low cognitive requirements, there is the danger that they will not give an accurate picture of the actual mathematical ability of the students.

Empirical research always indicates that most large-scale mathematics tests do not adequately represent higher-order cognitive processes, but instead focus more on procedural fluency and rapid calculation (Burkhardt and Schoenfeld, 2018; Smith et al., 2020). This limited scope reduces the ability of such testing to reflect important aspects of mathematical ability such as reasoning, strategic competence, and problem-solving. Anghel et al. (2024) claim that in cases where assessment items do not measure cognitively challenging mathematical practices, there is construct misalignment and the results of performance in such cases should not be interpreted.

Construct misalignment is a situation when intentions of the curriculum, cognitive tasks of the assessment, and inferences based on test scores are not connected to each other. In mathematics education, there is a growing focus on curriculum to include reasoning, modeling and application to real life situations. Nevertheless, high-stakes assessments based on short, time-constrained, or recall-based items do not tend to capture these objectives (Pettersen and Braeken, 2017; Provasnik, 2021). Consequently, it can happen that the performance of students can indicate the strategies of taking tests and not the real knowledge of mathematics..

It has also been shown that the lack of cognitive complexity in assessments is likely to inhibit the capability of students to show deep reasoning, especially in situations where multi-step problem solving and justification are involved (Salles, Santos, and Keskpaik, 2020). Mutawah et al. (2018) underline that proper mathematics testing must involve students to make a connection, generalise concepts, and create mathematical reasoning-skills which are frequently neglected in the traditional large-scale testing systems. In cases where these competencies are not included, the results of the assessment will give a partial view of mathematical ability.

Nortvedt and Buchholtz (2018) also emphasize that construct misalignment is a systemic problem, not simply a technical one, which impacts the issues of fairness and validity in high-stakes testing. Examinations that focus on speed and efficient process can also be prejudiced against students with their strong sides in reflective thinking and conceptual interpretation. This lack of correspondence is especially bad when the consequences of assessment are used to make consequential decisions like university admissions where false depiction of ability can have consequences over time.

On the whole, the literature highlights the idea that cognitive complexity should be regarded as a key factor of construct alignment in mathematics testing. Tests that do not include cognitively challenging tasks will not be able to measure the multidimensionality of mathematical competence. To solve construct misalignment, it is thus necessary to carefully incorporate higher-order cognitive demands into the design of an assessment so that there is coherence between what is taught, what is measured, and how performance is interpreted. This view offers a critical basis on which the inconsistencies between essay-based and multiple-choice mathematics tests in large-scale assessment systems can be seen.

Large-Scale Mathematics Examination Assessment Design and Bias in Items.

Large-scale mathematics assessments need to design assessments that minimise the possibility of item bias that can bias test scores and discriminate against certain groups of students. Item bias happens when a test item discriminates against a group of students due to elements that do not relate to the subject under evaluation, including language, culture, or socio-economic status (Anghel et al., 2024). Such bias affects the validity and fairness of the assessment and in such a case it is hard to get a true measure of the mathematical ability of a student.

Research by Burkhardt and Schoenfeld (2018) have established that bias in items can be experienced when the test item depends on the knowledge that is culturally specific or presupposes familiarity with specific situations that are not common to all students. This may create unequal results in test performance as students with other cultural or socio-economic backgrounds are unfairly disadvantaged. Nortvedt & Buchholtz (2018) recommend that to minimize item bias, test developers ought to come up with culturally neutral items that measure mathematics knowledge without assuming certain background information.

In order to reduce bias, it is possible to analyze the test item performance based on various groups of students using item response theory (IRT) and other statistical techniques. Detecting items with different item functioning (DIF), test developers can correct or drop biased items so that the assessment can be fair to the mathematical skills of all students (Smith et al., 2020). Provasnik (2021) supports a more inclusive model of assessment design, in which several voices are heard during the development of test items, which would allow being fair to various groups of students.

Speededness, Time Constraints, and Reasoning Quality

In large-scale mathematics assessments, speededness and time constraints are key factors that affect the reasoning quality of students' responses. Speededness is a concept that is related to the pressure that the students feel when they have to complete the test items in a short period of time. Such limitation is associated with the temptation to make a quick decision, which is not always accurate and, in some cases, entails superficial problem-solving instead of a well-considered and well-grounded solution (Burkhardt and Schoenfeld, 2018). Students tend to pay less attention to the content and fail to use higher-

order thinking processes, including analysis, synthesis, and evaluation, when they are rushed during the assessment.

Studies by Adeleke et al. (2025) indicate that time constraints have out of proportion negative impact on the students in exercising their genuine mathematical rationale, especially in the problems demanding intricate problem solving. Students can also lose important steps in their arguments, pay insufficient attention to the possible solutions, or provide inadequate justification of their work when they are compelled to go through assessment items fast. According to Salles, Santos, and Keskpaik (2020), this compromises the validity of the assessment since it ceases to assess the entire scope of mathematical competencies of the student but the ability to perform under pressure. Moreover, students who are anxious about the tests or unfamiliar with the form of the testing can have more difficulties than others and, as a result, the fair representation of their mathematical skills is not achieved (Anghel et al., 2024).

Anghel et al. (2024) also add that speededness can unwillingly put procedural fluency-tasks that demand fast calculation- over the more important skills of mathematical reasoning and problem-solving. This emphasis on time tends to create excessive attention to memorization and algorithmic processes instead of actually thinking mathematically. To overcome this, Mutawah et al. (2018) propose the introduction of tests that measure higher-level thinking and the ability to reason in depth and understand the information, which will enable learners to have more time to demonstrate their advanced thinking skills. Thus, future designs of assessment should be based on the balancing of the need to have efficient testing with the need to give students enough time to think logically on complex mathematical problems.

Synthesis of Literature and Research Gaps

The literature review has led to a convergence of the opinion that the challenges that persist in mathematics testing are as a result of the constraints of construct representation, cognitive demand, and interpretive validity. In the literature, there is a general consensus that scores in assessment are extremely sensitive to the design of the task, the format of response and conditions of testing. Nevertheless, there are still significant gaps in the way these factors are studied in comparison with each other in various examination systems that are used in the same educational environment.

Although the literature has a significant amount of studies that investigate cognitive complexity and task demand separately, fewer studies have investigated the interaction of these two dimensions with assessment format to influence the performance of high-achieving learners. Anghel et al. (2024) indicate that construct misalignment is usually recognized on theoretical grounds but there is a dearth of empirical research on how misalignment is expressed within parallel assessment systems that play different roles, including certification and selection. This is especially noticeable in the situations when multiple-choice examinations and essay-based tests coexist and are introduced in chronological order during the educational development.

In addition, although the role of time constraints in mathematics assessment has been documented, existing studies tend to treat speededness as a general testing condition rather than examining its differential impact across task types and learner profiles. According to Scherer and Wiberg (2018), it is possible that time pressure will have a disproportionately negative impact on those tasks that involve reasoning and reflection; however, little empirical evidence has been conducted to support the idea that time pressure contributes to differences in the assessment format among the same group of students.

The other evident gap is the interpretation of aggregate outcomes in high stakes situations. Nortvedt and Buchholtz (2018) warn that aggregate scores can mask individual strengths and lead to construct underrepresentation, especially when they are the major determinant of the selection decision. Nevertheless, there are not many studies that empirically investigated the translation or non-translation of high performance in one assessment format to other outcomes.

Lastly, most of the literature that exists is based on psychometric or item-level analysis with little consideration given to the way assessment format influences the type of reasoning students can exhibit. According to Leonard (2024), to comprehend the quality of reasoning, comparative designs are needed, which consider both the task structure and mode of response, which has yet to be applied in large-scale assessment studies.

Considering these gaps, the existence of research studies with construct-sensitive, comparative methods of studying mathematics performance based on assessment format is obvious. The current research aims to satisfy this requirement by examining differences between essay-based and multiple-choice mathematics test results in a common educational

setting, thus, providing the empirical support to the current debates on the validity, construct alignment, and interpretation of the large-scale mathematics testing.

METHODOLOGY

Research Design

The research design used was ex post facto comparative research design. This was the right design since the study examined the results of existing mathematics examination in various assessment systems without controlling any variables. The design allowed the comparison of the observed mathematics performance of candidates in the secondary school exit examinations (WAEC and NECO) and a university entrance examination (JAMB), which use various assessment formats.

Study Context

The research was carried out in the context of Nigerian secondary education of mathematics that is assessed using various high-stakes assessment systems. End-of-secondary school tests are certification tests in the form of CAWAEC and NECO examination, and the Unified Tertiary Matriculation Examination (UTME), administered by JAMB, is a gatekeeping test of entry into tertiary institutions. The comparison was made on mathematics examination results that are received within the same examination cycle to be comparable with time.

Population and Sample

The study population included secondary school graduates in Nigeria who have passed the WAEC or NECO mathematics examination, and finally taken the UTME mathematics examination. The study sample was 180 candidates who were selected in six North-East states in Nigeria. The sample included candidates who had full records of either WAEC or NECO mathematics and JAMB mathematics of the exam year under study.

The sampling method applied was stratified in order to represent the results of the examination. The candidates were classified according to their performance groups in WAEC/NECO mathematics and the performance groups in JAMB mathematics. This method made sure that the sample was varied in terms of observed performance in the various systems of examination.

Measures and Operationalisation of Variables.

The performance in mathematics in WAEC and NECO was operationalised based on the official grade levels, that is, credit-level, with grades of C4-B2, which appears on the examination results of the candidates. These are the lowest standards that allow one to graduate to tertiary education in the Nigerian education system.

The mathematics performance of JAMB was operationalised by subject-level scores and special focus was made on the low-performance range (aggregate scores less than 30 out of 100). This was a categorisation that matched the commonly applied performance thresholds in JAMB outcome reporting and could be compared to outcomes of secondary school examinations once they had been transformed into scores.

WAEC/NECO grades and JAMB mathematics scores were standardised before being subjected to inferential analysis so as to facilitate comparability of exams with different scoring structures. Z-score transformation was used in carrying out standardisation in order to put all the measures on a similar measure.

Data Collection Procedure

The study data was collected by means of archival examination records. The official records given by participating schools and examination records were used to extract the WAEC or NECO mathematics grades and the JAMB mathematics scores of the candidates. Only total and verifiable records were incorporated in the analysis. No further test or survey tools were used.

Data Analysis

Analysis of data was done in two phases. To begin with, descriptive statistics were calculated to describe the patterns of mathematics performance in WAEC/ NECO and JAMB examination. The observed outcomes were described using frequency distributions and measures of central tendency.

Second, a paired- samples t-test was used to compare the difference in the performance in mathematics in the two formats of examination following the standardisation of the scores. The paired-samples method was suitable since the performance results of the same candidates were compared in two evaluation systems. The difference scores were tested on assumptions of normality before analysis and the results were found to be acceptable. The level of statistical significance was assessed at 0.05 level.

RESULTS

Research Question 1: ***What are the observed mathematics performance patterns in WAEC, NECO, and JAMB examinations among Nigerian secondary school graduates?***

The descriptive statistics were used to answer Research Question 1 that evaluated the mathematics performance patterns of candidates taken between the WAEC, NECO, and JAMB examinations. The performance of mathematics was summarised by the use of official mathematics grade classifications in summary of the performance in WAEC and NECO with the performance in JAMB mathematics summarised using subject-level aggregate score ranges.

Table 1: Distribution of Candidates' Mathematics Performance in WAEC, NECO, and JAMB Mathematics Aggregates (N = 600)

Examination	WAEC/NECO Mathematics Grade	Corresponding JAMB Mathematics Aggregate	Frequency	Percentage (%)
WAEC	B1–B3	20–35	72	12.0
	C4–C6	10–25	138	23.0
	E8–F9	< 10	30	5.0
Total (WAEC)			240	40.0
NECO	B1–B3	20–35	66	11.0
	C4–C6	10–25	114	19.0
	E8–F9	< 10	36	6.0
Total (NECO)			216	36.0
WAEC & NECO	B1–B3	20–35	24	4.0
	C4–C6	10–25	84	14.0
	E8–F9	< 10	36	6.0
Total (WAEC & NECO)			144	24.0
Grand Total			600	100.0

The JAMB Mathematics aggregate ranges presented in Table 1 reflect observed performance groupings associated with candidates' WAEC and NECO mathematics grade categories in the dataset and do not imply direct score equivalence or scale alignment between the examination systems.

The distribution indicates that a sizeable percentage of the candidates, having achieved credit-level grades (B1-B3 and C4-C6) in both WAEC and NECO mathematics, were related with low to moderate JAMB mathematics aggregates. The candidates who had poor mathematics scores in secondary schools (E8-F9) were mostly linked with very low JAMB mathematics aggregates.

Research Question 2*: How does mathematics performance differ between essay-based WAEC/NECO examinations and the multiple-choice–based JAMB examination?*

In order to answer Research Question 2, the performance in mathematics in WAEC/NECO and JAMB examinations was compared using standardised scores to control the variation in scoring structures. Mean standardised scores were calculated in essay-based (WAEC/NECO) and multiple-choice-based (JAMB) tests.

Table 2: Mean Standardised Mathematics Performance Scores in WAEC/NECO and JAMB (N = 600)

Examination Format	Mean (M)	Standard Deviation (SD)
WAEC/NECO (Essay-based)	0.49	0.82
JAMB (Multiple-choice)	−0.44	0.90

The mean standardised mathematics performance score for WAEC/NECO was higher than that for JAMB, indicating better observed performance in the essay-based examinations compared with the multiple-choice examination for the same group of candidates.

Research Question 3; *Is there a statistically significant difference between candidates' mathematics performance in WAEC/NECO and JAMB examinations?*

A paired-samples t-test was conducted to determine whether the observed difference in mathematics performance across WAEC/NECO and JAMB examinations was statistically significant**.**

Table3: Paired-Samples t-Test Comparing Standardised Mathematics Performance in WAEC/NECO and JAMB (N = 600)

Comparison	Mean Difference	t	Df	p-value	Effect Size (Cohen's d)
WAEC/NECO vs JAMB	0.93	22.17	599	< 0.001	0.91

The paired-samples t-test revealed a statistically significant difference between candidates' mathematics performance in WAEC/NECO and JAMB examinations. The magnitude of the difference was large, as indicated by a Cohen's d value of 0.91.

DISCUSSION

This paper has been analyzing the trends in mathematics performance on the WAEC, NECO and JAMB exams specifically the differences between essay based secondary school tests and a multiple choice test on university entrance. The results show

that a large percentage of applicants with credit-level results in WAEC and NECO mathematics had poor results in JAMB mathematics. This trend was also observed with descriptive distributions, standardised mean comparisons and inferential analysis indicating that the differences were systematic and not accidental.

The standardised scores analysis revealed that the performance in mathematics was higher in WAEC/NECO than in JAMB among the same students and the paired-samples test revealed that there is a statistically significant difference with a large effect size. The findings present evidence that variances in the performance observed remain even after considering the differences in scoring structures, which is a significant difference in the examination formations. Notably, the research did not presuppose that grading systems are similar but rather investigated realized results in every assessment setting.

The results are in line with previous debates in the research on mathematics education on the possible impact of assessment form on performance outcomes. Essay-based tests can enable applicants to have partial knowledge and procedural thinking, and multiple choice forms can need other ways of responding. Nevertheless, the current research does not give any assertions regarding the underlying cognitive processes or the quality of the instruction; instead, it presents empirical data regarding outcome misfit during high-stakes examinations of varying format and purpose.

The findings highlight the importance of care when making interpretations of mathematics performance in a multi-system of examination in one education. Where secondary school certification and university admission depend on other forms of assessment, observed disparities can be due to structural and design variations as opposed to homogeneous levels of mathematical ability. The study offers context-specific, archival evidence, which adds to the current discussion of the comparability of assessment and the meaning of the results of high-stakes mathematics examinations.

CONCLUSION

The researcher undertook a study on the performance of mathematics in secondary school through essay based examinations (WAEC and NECO) and a multiple choice university entrance examination (JAMB) using archival candidate data. The results indicated that a significant number of the candidates who received credit-level marks in both WAEC and NECO mathematics reported low cumulative marks in JAMB mathematics. These

differences still existed even after the scores had been standardised and this was even verified through inferential analysis which meant that the observed performance differences were not random but indeed systematic.

The study ensured the clarity of constructs by operationalising performance in terms of official grade bands in the case of WAEC and NECO and aggregate score thresholds in the case of JAMB, to ensure there was no assumption of the similarity of scores across examination systems. The findings indicate that the assessment format and response conditions are significant in defining the observable performance in mathematics especially in high-stakes situations where examinations are used to fulfill specific functions of certification and selection.

The research adds empirical data to the literature of mathematics education on the comparability of assessment in multi-examination systems. It emphasizes the need to interpret results of mathematics examination with care and sensitivity to constructs when they are based on fundamentally different assessment frameworks. Although the results are local, they highlight the importance of the general issues regarding the role of assessment design in measuring and interpreting mathematical competence.

The study had the potential to be expanded in future studies by using item-level processes, process-based data, or mixed-format tests, to further investigate the effects of the various assessment designs in eliciting mathematical reasoning. In general, the performance in mathematics in WAEC, NECO, and JAMB must be interpreted as the manifestations of the structure and conditions of the assessment and not as the clear evidence of the mathematical skills of students.

REFERENCES

Adeleke, A. O., et al. (2025). The limitations of traditional mathematics assessments in measuring problem-solving ability. *Journal of Mathematics Education, 49*(3), 312–323.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017

Amalina, L., & Vidákovich, S. (2022). Mathematical literacy: Expanding the scope of assessment. *Educational Assessment Review, 24*(4), 45–59.

Anghel, F., et al. (2024). Cognitive alignment in mathematics assessments: A critical review. *Journal of Educational Measurement, 37*(5), 167–182.

Burkhardt, H., & Schoenfeld, A. H. (2018). Assessment in the service of learning: Challenges and opportunities or Plus ça change, plus c'est la même chose. *ZDM–Mathematics Education, 50*, 571–585. https://doi.org/10.1007/s11858-018-0937-1

Cantley, I., & McAllister, J. (2021). The gender similarities hypothesis: Insights from a multilevel analysis of high-stakes examination results in mathematics. *Sex Roles, 85*(7–8), 481–496. https://doi.org/10.1007/s11199-021-01234-5

Fokuo, M. O., Lassong, B. S., & Kwasi, S. F. (2022). Students' poor mathematics performance in Ghana: Are there contributing factors? *Asian Journal of Education and Social Studies, 30*(4), 16–21. https://doi.org/10.9734/ajess/2022/v30i430729

Leonard, B. (2024). Reasoning in mathematics assessments: A deeper analysis. *Mathematics Education Journal, 42*(3), 230–245.

Mutawah, R., et al. (2018). A framework for assessing mathematical problem-solving. *Journal of Educational Psychology, 45*(2), 225–238.

Ní Shúilleabháin, A., O'Connor, R., & Meehan, M. (2019). Investigating cognitive demand of higher-level Leaving Certificate mathematics examination tasks pre- and post-curriculum reform. In L. Harbison & A. Twohill (Eds.), *Proceedings of the Seventh Conference on Research in Mathematics Education in Ireland (MEI7)* (pp. 203–210). Zenodo. https://doi.org/10.5281/zenodo.3580824

Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM–Mathematics Education, 50*, 555–570. https://doi.org/10.1007/s11858-018-0963-z

Papanastasiou, E., & Michaelides, M. (2024). Cognitive complexity and mathematical problem-solving in large-scale assessments. *Educational Psychology Review, 36*(2), 178–192.

Park, M., et al. (2023). The role of student practices in mathematical ability assessments: A neglected factor. *Mathematics Education Journal, 58*(2), 129–145.

Pettersen, M., & Braeken, S. (2017). Reimagining mathematics assessments: Moving beyond procedural fluency. *Mathematical Thinking and Learning, 19*(3), 185–199.

Provasnik, S. (2021). Process data, the new frontier for assessment development: Rich new soil or a quixotic quest? *Large-Scale Assessments in Education, 9*, Article 1. https://doi.org/10.1186/s40536-020-00092-z

Raji, M. O. (2021). Postsecondary entrance examinations in Nigeria: A critical analysis of the design and predictive validity of the tripartite assessment system. *African Journal of Teacher Education, 10*(1), 70–86. https://doi.org/10.21083/ajote.v10i1.6296

Salles, R., Santos, E., & Keskpaik, J. (2020). Developing process-oriented assessment tasks for higher-order mathematics. *Educational Psychology Review, 32*(4), 487–504.

Scherer, N., & Wiberg, M. (2018). Time constraints and problem-solving performance in mathematics assessments. *Journal of Educational Measurement, 55*(4), 399–412.

Smith, J., Wang, H., & Liu, L. (2020). Revisiting the role of cognitive complexity in mathematics assessments. *Journal of Research in Mathematics Education, 24*(1), 15–25.

Zana, F. M., Sa'dijah, C., & Susiswo. (2024). The cognitive alignment of mathematics teachers' assessments and its curriculum. *International Journal of Evaluation and Research in Education, 13*(3), 1561–1575. https://doi.org/10.11591/ijere.v13i3.26814