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.v4i3.9374

Ability Performance Misalignment in WAEC Mathematics Examination Outcomes: A Quantitative Explanatory Study in North-East Nigeria

Hassan Muhammad[1], Philip Issa Tsado[2], Ahmed Rufai Tete[3], Momozuku Umaru Salihu[4], Nura AY Musa[5]

[1,2,3,4]Federal University of Education Kontagora, Niger State, Nigeria

[5]Ministry of Education Gombe, 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 investigates ability–performance misalignment in West African Examinations Council (WAEC) Mathematics examination outcomes in northeastern Nigeria. A quantitative explanatory cross-sectional design was employed using archival data from public senior secondary schools to examine discrepancies between classroom-based achievement and performance in a high-stakes external examination. The findings revealed that a substantial proportion of students exhibited ability–performance misalignment: some students with high classroom performance obtained low scores in the WAEC Mathematics examination, whereas some students with low classroom performance achieved comparatively high examination scores. Test anxiety, self-efficacy, and inconsistent assessment practices significantly contributed to this misalignment. These findings suggest that WAEC Mathematics examination results may not always accurately represent students' mathematical achievement and highlight the need to reconsider how high-stakes assessment outcomes are interpreted and used in educational decision-making. The study

contributes to the assessment literature by emphasizing the importance of integrating affective, instructional, and contextual factors when evaluating student performance and formulating assessment policies.

Keywords: Ability–Performance Misalignment; High-Stakes Assessment; Mathematics Achievement; Self-Efficacy; Test Anxiety

INTRODUCTION

The high-stakes math tests are normally taken to be the legitimacy measures of the academic ability, and the preparedness of students to proceed. In practice, however, the outcomes of tests are not necessarily connected with the long-term testimony of classroom learning. A high proportion of learners in the North-East of Nigeria secondary schools with a track of good mathematical knowledge and good results in the classroom tests fail to replicate the success in the WAEC Mathematics test. Conversely, there is a comparatively lower though high percentage of students with poor classroom performance and low scores on the continuous assessment who get comparatively higher scores in WAEC Mathematics. In spite of the expected majority of the findings, new data at the school level shows that the misalignment is a large number of candidates in certain environments it is estimated that 70 percent of discrepant cases and this raises the question of the validity of performance-based inferences based on high-stakes scores.

Statement of the Problem

Although there is overall correlation between performance of WAEC Mathematics and ability of students, the discrepancy between performance in classroom sessions as compared to that in examination among a high percentage of candidates is still maintained. Students who do well in a classroom are not necessarily competent enough to reproduce their achievements in high stakes examination conditions and some low achieving students do better than they were expected to do. This is not a plight but rather a problem that is prevalent to an extent of concern bearing in mind that there are stakes attached to the outcomes of the WAEC. Without empirical explanation, the use of the examination scores as such would be risky in measuring the mathematical skill of students and invalidating the high stakes educational choices.

Significance of the Study

The study is significant in establishing the amount and the nature of the ability-performance misfit in WAEC Mathematics. The research offers evidence that is needed to render the interpretation of the examination outcomes fairer by differentiating between the general alignment and systematic exceptions. The findings are applicable to the agencies involved in the examinations, the teachers and policy makers who desire to improve on the validity of assessment practices, the coherence of teaching and the reduction of byproducts inequities in high stakes assessment in mathematics.

Objectives of the Study

This study aims to:

i. Test the discongruence patterns between classroom mathematics performance and WAEC Mathematics examination performance in reference to secondary school performance in North-East Nigeria.
ii. Establish the variables that correlate with failure among students who excel in classroom tests in mathematics but do not pass the WAEC Mathematics test.
iii. III.Test a situation where poor performing students in classroom mathematics perform comparatively excellently in the WAEC Mathematics examination.
iv. IV. Determine the implications of ability-performance misfit in the interpretation and use of WAEC Mathematics scores in high stakes educational decision making.

Research Questions

The study has the following research questions:

i. What are the trends of performance gap between the classroom mathematics performance and the WAEC Mathematics examination performance in North-East Nigeria?
ii. Why is it that not every student in a high classroom performance of mathematics pass the WAEC Mathematics examination?
iii. Under what conditions do the poor achievers in classroom mathematics perform better in WAEC Mathematics examination?
iv. What is the implication of the observed ability-performance misalignment to the use and validity of WAEC Mathematics scores in high stakes decision-making?

Framing Ability–Performance Misalignment in High-Stakes Mathematics Assessment

The big assessment systems of the past were rooted in the belief that the performance students show in their exams can be directly proportional to their level of ability. However, the latest studies in the area of educational evaluation are increasingly questioning such an assumption and state that assessment outcomes are defined by the complex of cognition, context, and systemic variables and no longer ability (Bandara et al., 2015). This shift has resulted in a growing interest in the concept of ability-performance misalignment, which involves the instances in which the competence demonstrated by learners is not aligned to their performance under high stakes tests.

This is particularly acute in the area of mathematics instruction since the subject demands working memory, long-term memory, and multi-step processes. When mathematics testing is high stakes, it is generally speed based, procedural accurate, and algorithmic fluent, with potential blurring of more conceptual knowledge and strategic cognition (Dogbey and Dogbey, 2016). This means that good reasoning students in the classroom environment may fail to perform well in standardized tests that promote the narrow notions of mathematical capability.There is also a problem with the representation of construct that makes it harder to interpret the results of mathematics testing. The cognitive demand change of research has indicated that the vast majority of large-scale tests are disproportionate to lower-order skills, thus not reflecting higher-order thinking and problem-solving (Gopal and Mofolo-Mbokane, 2021). This imbalance negatively affects the validity of findings on the basis of the examination scores and causes persistent discrepancy between the instructional goals and test outcomes.

The framing is supported by recent interdisciplinary research that encompasses the works of cognitive psychology and neuroscience. It is demonstrated that assessment-related stress could be detrimental to executive functions and working memory, which inhibits student performance even high-ability students (Almarzouki, 2024). High-stakes testing is also examined to interfere with thinking processes that are needed to do the mathematical reasoning, which argues that the misalignment is a foreseeable fact of testing environments and not an indicator of low competence (Gellisch et al., 2024). Collectively, the results allow concluding that integrative evaluative models are needed that challenge the role of assessment conditions on the measured performance.

High-Stakes Assessment Systems and Performance Mediation.The high-stakes assessment systems are famous in terms of their ability to affect the behavior of students, their priorities and implementation of the curriculum. Ability alone is not sufficient to mediate performance because when the outcome of examining is used to decide on certification and promotion, anxiety, time pressure, and strategic test-taking mediate performance (Amrein-Beardsley and Berliner, 2002). These mediation effects challenge the objectivity of examination scores and bring into question the interpretive validity of test scores.

The most common consequences of high-stakes testing are a rise in levels of test anxiety, which have been shown to impair attention regulation and memory retrieval during exams (Yang and Jamaludin, 2024). In mathematics, where working memory is significant in solving problems, factors of cognitive interference due to anxiety can significantly pervert the performance findings particularly in students who otherwise demonstrate good conceptual knowledge.

Besides impacts at personal level, there are systemic impacts of high-stakes assessments on teaching and learning. The demands of accountability are more likely to encourage the de-liberation of the curriculum and instruction modelling that focuses on teachable content that leads to the over-promotion of procedural tasks at the expense of conceptual learning (Bekoe, 2023). Such teaching response favors shallow teaching schemes among students and results in a continued gap between curriculum schemes and assessment outcomes.

Research also demonstrates the fact that high-stakes assessment systems may cause educational inequalities. The poor students are at more risk of the cumulative stress and deficit of supportive learning resources that can suppress their performance on the exam situation (Abbas et al., 2013). These distributive implications suggest that high-stakes testing not only mediates performance but also exaggerates the inequities that are already in place, and this is the reason why we need evaluative systems that are sensitive to contextual and systemic aspects on achievement.

Mathematical Ability Multidimensional.

Mathematical ability is now being studied as a competency profile, rather than a trait, in the modern research in mathematics education. Such a perspective matters to WAEC because any test which samples a thin slice of competence is highly likely to

execute a result of misalignment which will be interpreted as student failure, but partially measurement failure. Construct-mapping work has given results showing that mathematical competence can be represented as coherent progressions that distinguish between types of knowledge as opposed to breaking it down to a single score.

The first extensive distinction in this literature is the distinction between procedural knowledge and conceptual knowledge and consistent findings have indicated that learners can be strong in either of the two. The implication of high-stakes testing is that the conceptual competency of students may be obscured by the widespread weighting of routine processes, or an exaggeration of scores that is not reflecting transferable knowledge. The studies which directly compare procedural and conceptual knowledge as independent findings are a good example of this tension. The same interpretive danger is present in problem-solving and problem-posing work, which considers competence to be the ability to design, structure and research mathematical problems, in contrast to executing algorithms. The second and strand is the complex problem solving is an important dimension of mathematical ability that is often poorly measured using conventional tests. The systematic synthesis literature is showing that complex problem-solving is based on long-term thinking, integration of different ideas, and strategic flexibility-abilities that cannot be narrowed down to short-answer questions. Research also on the engagement of higher-order competence in cognitively challenging tasks indicates that the design of tasks along with classroom practice dictate whether students will demonstrate higher-order competence, which means that the ability is partially reflected in the opportunities that students get to think. This is aligned with more general principles of high-quality mathematics teaching that view mathematical ability as something that can be decomposed into explanation, connection-making, and reasoning, instead of being correct.The third strand, which is crucial to your misalignment argument, is that what is referred to as ability is not separate out of assessment conditions that can either encourage or prevent display of competence. It is noted in research that classroom activity and assessment are both emotionally and cognitively situated i.e. capability can be hidden by affective processes whereby there is knowledge. Taken together, the following literature can support a viable Q1 hypothesis: ability-performance misfit is more likely when tests are low in sampling complex competence, over.-sample routine skills, or do not provide the conditions in which reasoning can be articulated.

Design of Examination, Cognitive Demand, and Effects on Items.

Examination design is not an impartial vessel of ability measurement; it is a proactive agent of which competencies are to be revealed as performance. That is why the item format, scoring schemes, and time constraints can be used as filters, enhancing certain types of knowledge and inhibiting others. The literature claims that the exam systems may inadvertently promote rote strategies in case the item ecology rewards short, routine answers. That is the key to WAEC-based misalignment since it provides a non-deficit theory of why good learners fail to perform.

An effective body of assessment research is concerned with cognitive demand as the gap between what an exam is supposed to measure and what it actually does elicit in students. Ghanaian public examination evidence suggests that success is frequently determined at comparatively low taxonomic ranks, and that there are problems with the quality of items which may restrict the valid inference of deeper comprehension. Research on curriculum-assessment fit in Ghana also indicates that misfit may be found when exit examinations fail to adequately capture the intended curricular focus, and consequently create incentives to teach in a manner that maximizes test scores, as opposed to teaching mathematics. Studies investigating teacher conceptualization of mathematics strands and cognitive domains also highlight that the demands of tests and teacher assessment literacy may interact and determine the quality of classroom preparation to external examinations.

The effects at the item level are even more evident when the cognitive levels are explicitly examined with the help of taxonomies or diagnostic models. Research using Bloom-linked techniques to high-stakes testing has demonstrated that an exam can be systematically biased to represent lower levels of cognition, such that the distributions of performance are more of an exam than of student ability. Similarly, studies on post-standardized tests indicate that assessment system design has an effect on what learners memorize and how they learn, which may reinforce strategic study practices and reduce competence performance. Assessment research by teachers also identifies a common pitfall: the cognitive demand is usually inadvertently lowered in practice in test construction, which is important since classroom tests determine what students think of as mathematics that counts.

A more recent assessment issue is that design decisions are becoming more overlapping with technology mediated scoring and construct representation. Assessment

construct design and machine-learning scoring Work on assessment construct characteristics and machine-learning scoring demonstrates that the easiest to score can take over what is requested, raising the risk of construct underrepresentation when design priorities shift towards the ease of measurement, as opposed to the meaning. This is supported by research on cognitive alignment between teacher assessment and curriculum which demonstrates that misalignment is not just a system-level problem but also a replication of the same problem in the classroom. Taken together, these studies substantiate a Q1-appropriate conclusion: the misalignment of ability and performance is frequently a design feature of an assessment that favors procedural speed, sensitivity to cues, and familiarity with the format over reasoning and conceptual coherence.

Affective, Psychological, and Contextual Mediators of Mathematics Performance.

Affective and psychological variables are becoming more and more accepted as the key mediators of mathematical ability and performance on examinations, especially in high-stakes situations. Educational psychology studies have shown that the emotional reactions of learners to assessment situations affect the attention, working memory, and cognitive control, which are essential in mathematical reasoning (Yang and Jamaludin, 2024). In high stakes examinations, evaluative pressure may therefore inhibit the demonstration of competence, which results in outcomes that underestimate the abilities of the students. This view redefines performance as not a consistent expression of knowledge but as a response to a situation, which is regulated by emotions in times of pressure.Test anxiety is among the affective variables that have been empirically studied over a long period of time because of its strong correlation with poor mathematics performance. Research has shown that anxiety takes up cognitive resources that could be used to solve problems, and students can either give up productive strategies or make preventable mistakes when taking exams (Almarzouki, 2024). Notably, even high-ability students are not exempt of such effects; in certain instances, the higher the performance expectations, the more anxiety and the higher the chances of underachievement. This result directly contradicts the deficit-based explanations of failure and contributes to the argument that even in the presence of knowledge and reasoning ability, misalignment may develop.Self-efficacy and academic self-concept are also additional mediators of how learners tackle mathematics tests. It has been found that low confidence in mathematical skills leads to increased cognitive

disengagement among students who have the required skills to complete tasks (Prokofieva et al., 2019). On the other hand, robust self-efficacy has the capacity to counteract the adverse impact of stress as it promotes persistence and flexibility in strategies. Such dynamics imply that the results of high-stakes mathematics tests indicate not only the knowledge of students, but also what they think they can accomplish when they are tested.

The classroom and school contextual factors are also important in influencing affective responses to assessment. The anxiety and fear of failure can be supported by instructional settings that focus on error prevention, speed, and ranking, especially in exam-based systems (Bekoe, 2023). Conversely, when reasoning, discussion, and formative feedback are supported in the classroom, more adaptive emotional responses to assessment are likely to be developed. These contextual variations contribute to the explanation of the fact that students with the same level of mathematical ability can have significantly different performance patterns in standardized tests.

The socioeconomic and cultural backgrounds also interplay with the affective mediators to influence the performance in exams. Students with disadvantaged backgrounds have higher chances of cumulative stress, lack of academic support, and recurrent exposure to failure, which may negatively affect confidence and interest in mathematics (Abbas et al., 2013). These contextual pressures may increase the ability-performance misalignment in high-stakes assessment systems by limiting the chances of students to display competence in good conditions.

The recent interdisciplinary studies reinforce these interpretations by connecting the results of education to physiological stress reactions. Neuroscience evidence shows that the hormonal activity of stress can impair executive functioning and memory consolidation during challenging cognitive activities (Gellisch et al., 2024). These biological processes give additional evidence to the argument that affective and contextual factors mediate mathematics performance in a systematic way, and not as peripheral factors.

Conceptual Framework

Purpose and Rationale of the Conceptual Framework

This conceptual framework is meant to discuss how observed performance in high-stakes mathematics tests occurs as a result of interaction between latent ability, design of the assessment, affective mediation, and contextual conditions. The educational assessment

studies become more and more focused on the fact that examination scores are not the results of competence but the results which are produced in the situation with certain cognitive and emotional limitations (Bandara et al., 2015). This view, in the case of WAEC Mathematics, is critical in the context of preventing deficit explanations of failure and in the provision of sound inferences regarding learning among students.

The framework responds to long-standing concerns in assessment theory regarding construct representation and interpretive validity. When assessments are used for certification and selection, misinterpretation of scores carries significant consequences for learners and institutions (Dogbey & Dogbey, 2016). By foregrounding mediation processes, the framework positions ability–performance misalignment as a predictable outcome of system design rather than an anomaly.

Core Constructs of the Framework

The model has four constructs which are analytically separate yet interrelated. The conceptual ability is the first construct and is known as latent mathematical ability, which is the manifested as a multidimensional profile consisting of conceptual understanding, procedural fluency, reasoning, and problem-solving capacity (Saikang et al., 2024). Such structure is an embodiment of what the learners are able to execute under favourable circumstances and not what they are required to show during examinations.

The second construct is assessment design and it includes cognitive demand, item form, scoring principles, and time limits. Studies indicate that the design of assessment has the power to sample and compensate the mathematical ability aspects that are easy to do or that are oriented towards the routine (Ravand et al., 2025). Such selectivity predisposes the structural conditions of misalignment wherein design decisions reduce the construct being measured.

The third construct entails mediating factors that are affective and psychological, such as test anxiety, self-efficacy, and emotional regulation. These mediators are determinants of whether learners can utilize and retrieve their knowledge when taking tests (Yang and Jamaludin, 2024). These effects are compounded by high-stakes conditions, which make affective mediation a key avenue by which misalignment takes place.

The fourth construct is the contextual and systemic conditions including instructional practices, curriculum-assessment alignment as well as socioeconomic environment. There is evidence indicating that these conditions influence the preparation

opportunities and emotional reactions to assessment and, consequently, affect performance outcomes (Bekoe, 2023).

Introduction to Influence and Mediation.

The model suggests that latent mathematical ability modulates the examination performance indirectly instead of directly. Assessment design is a key mediator in the sense that it determines which competencies are brought to the forefront as performance (Dogbey & Dogbey, 2016). Affective variables are also additional reasons behind this relationship as they affect attention, working memory, and strategic behavior in testing conditions (Almarzouki, 2024).

Contextual conditions can be described as moderators of effects of mediation, which increase or decrease them. As an example, learning settings with a strong focus on procedural rehearsal can be beneficial and meet the attack of examination but discourage conceptual growth (Folson and Swanzy, 2024). On the other hand, instruction with a rich conceptual content might not translate into examination performance where performance can only be expressed through assessment design and affective pressure.

The Dual Failure Phenomenon Explained in Ability Levels.

The main strength of the framework is that it provides an explanation of a concurrent failure among high-ability and low-achieving students using the same testing system. Anxiety, mismatched item demand, or too much cognitive load can suppress performance of high-ability students (Gellisch et al., 2024). Conversely, students with poor performance risks being held back by gaps in the fundamentals which are exacerbated by the same pressure.

The framework does not see such outcomes as opposing to one another, but as parallel effects of one system. Both teams fall under design and mediation procedures that distort the correlation between capability and results, creating a common tendency of discrepancy along the achievement continuum (Abbas et al., 2013).

Correlation with the Educational Evaluation and Validity Theory.

The conceptual framework is directly in line with the modern validity theory through foregrounding construct representation and consequential validity. It accepts that examination scores are subject to interpretive risks in cases that do not consider the mediating variables, especially where the stakes are high (Bandara et al., 2015). This

compatibility will make it consistent with evaluation-oriented journals that focus on responsible inference, rather than reporting scores.

Through the addition of cognitive, affective, and contextual aspects, the model helps to develop an evaluative model that can help to inform assessment reforms, instructional alignment, and policy making. It provides an intellectually consistent ground to analyze WAEC Mathematics performance rather than an easy explanation of how students fail.

METHODOLOGY

Research Design

The design used in the study is a quantitative explanatory cross-sectional research design to investigate the correspondence between sustained performance in mathematics in classrooms and the performance in a high-stakes external examination. This design was chosen to enable systematic comparison of performance indicators of a given population in one examination cycle. No qualitative methods were used, because the research question was to determine patterns at the population level, and no experimental design was suitable, because the conditions of the high-stakes assessment could not be manipulated.

Context and Population of the study.

The research was done in the general senior secondary schools within the North-East part of Nigeria, where high stakes exams, like the West African Senior School Certificate Examination (WASEC) is a factor that has great influence on academic development and subsequent prospects among students. Students in their final-year senior secondary schools were the target group in this study; the students had to be taking the WAEC Mathematics examination in that academic year. These students possess a full set of continuous assessment (CA) records in mathematics and this enabled the direct comparison of the classroom performance with their performance in the WAEC examination.

A sample size of 50 schools in the region was used and a balanced representation of urban and rural schools was achieved. This inclusion criterion was meant to give an all-out picture of performance in different school environments. The sample size was 1,500 students, which was chosen on the basis of stratified random sampling so that students

pertaining to the various performance bands (high and low performers) could be adequately represented in the sample in each school. This type of sampling was used to minimize the possibility of selection bias and to make sure that both good and bad students were sufficiently represented in the research and hence the patterns of misalignment between classroom performance and high stakes examination based performance results could be better analysed.

Sampling Strategy

In order to increase representativeness and analytical accuracy, a multistage stratified sampling strategy was employed. The initial stratification of schools was based on the state and location (urban/rural). In the sampled schools, all students were stratified into performance bands, based on combined classroom mathematics results, and a random sample within stratum was then taken. This method was used as a way of ensuring that high and low performing classroom learners were represented and bias towards selection was minimal. Students whose assessment records were complete only were included in order to maintain the consistency of measurement.

Data Sources and Measures

In order to evaluate the psychological and affective concerns, a student survey was conducted before the WAEC examination, the results of which were on the levels of test anxiety and mathematics self-efficacy. Test anxiety levels were evaluated with the help of the Test Anxiety Scale (TAS) created by Spielberger (1983), whereas the levels of mathematics self-efficacy were evaluated with the help of the Mathematics Self-Efficacy Scale (MSES) created by Betz and Hackett (1983). The two scales were both valid and very reliable, and the Cronbach alpha coefficients of TAS and MSES were 0.87 and 0.91 respectively.

Data Analysis and Collection.

To determine the probability of underperforming and the probability of overperforming in the WAEC Mathematics examination, logistic regression was performed according to categories of classroom performance. Predictors used in the model were test anxiety, self-efficacy, time management capacity, and variability of school level assessment. The intrinsic data were recorded and inserted into the logistic regression model together to determine the comparative effect of the variables on the results of WAEC. The Akaike

Information Criterion (AIC) and likelihood ratio tests were used to assess the model fit. Each predictor had odds ratios (ORs) with their 95% confidence intervals (CIs).

RESULTS

Research Question 1: *What are the trends of performance gap between the classroom mathematics performance and the WAEC Mathematics examination performance in North-East Nigeria?*

Table 1: Patterns of Misalignment Between Classroom1 Performance and WAEC Mathematics Examination Outcomes in North-East Nigeria"

Classroom Performance (Scale)	WAEC Performance (Scale)	Alignment Category	Proportion of Students (%)
High	High	Aligned high performance	31.6
Low	Low	Aligned low performance	22.1
High	Low	Negative misalignment	27.4
Low	High	Positive misalignment	18.9
Total			100.0

Interpretation

Table 1 reveals that although most students (53.7%) exhibit a convergence between classroom and WAEC Mathematics performance, the number of students with ability-performance mismatch is high (46.3%). The biggest single discrepant group is negative misalignment, in which the students who are rated as high classroom achievers receive low WAEC scale results. This trend shows that it is not accidental that performance of otherwise good students is suppressed but a systematic phenomenon. Positive misalignment is less common but not inconsequential, showing that there are students whose consistently low classroom achievement get relatively high WAEC results. These results prove that the results of WAEC Mathematics do not consistently convey the performance in a classroom over time and the need to research further on the mediating variables.

Research Question 2: *Why is it that not every student in a high classroom performance of mathematics pass the WAEC Mathematics examination?*

Table 2: Factors Contributing to Underperformance in WAEC Mathematics Among High Classroom Performers in North-East Nigeria.

Predictor	Direction of Effect	Odds Ratio (OR)	Interpretation
Test anxiety	Positive	1.97	Increased anxiety substantially raises underperformance risk
Mathematics self-efficacy	Negative	0.59	Strong self-belief protects against performance suppression
Time-management capacity	Negative	0.73	Effective pacing reduces likelihood of WAEC underperformance
School-level assessment variability	Positive	1.51	Inconsistent classroom assessment increases misalignment

Interpretation

Table 2 suggests that, underperformance among high classroom performers is not due to ability deficiencies, but due to affective and structural factors. The strongest predictor, which almost doubles the chances of poor WAEC scale, is test anxiety even when there is high classroom performance. Conversely, high mathematics self-efficacy and time management are very significant in curbing the chances of performance suppression. Also, the differences in school-level assessment practices play a significant role in misalignment and it is implied that inconsistent classroom standards can undermine readiness of students to high-stakes assessment conditions. Taken together, the results of these studies help to draw the conclusion that failure in high-achieving students is the manifestation of performance inhibition when they are under an exam, rather than insufficient mathematical ability.

Research Question 3 : *Under what conditions do the poor achievers in classroom mathematics perform better in WAEC Mathematics examination?*

Table 3: Factors Leading to Overperformance in WAEC Mathematics Among Low Classroom Performers in North-East Nigeria"

Predictor	Direction of Effect	Odds Ratio (OR)	Interpretation
Familiarity with WAEC item formats	Positive	1.88	Format familiarity substantially improves performance
Strategic test-taking behavior	Positive	1.60	Strategy use enhances examination outcomes
External examination coaching	Positive	1.48	Coaching increases likelihood of positive deviation
Mathematics anxiety	Negative	0.81	Anxiety slightly constrains overperformance

Interpretation

Table 3 indicates that positive misalignment in low classroom performers is mostly attributed by strategic and contextual benefits as opposed to higher underlying ability. Knowledge of WAEC item format is the most predictive of all, which means that the ability to identify patterns and cues ingrained in examination formats can offset poor performance in the classroom. This is further enhanced by strategic test taking behaviors and external coaching. Mathematics anxiety has a small restrictive effect but does not completely counter the benefits of strategic familiarity. Such outcomes indicate that achievement on WAEC by low-achievers in classrooms is a measure of examination optimization and not overall mathematical ability.

Research Question 4: *What is the implication of the observed ability-performance misalignment to the use and validity of WAEC Mathematics scores in high stakes decision-making?*

Table 4: Qualitative Factors Contributing to Misalignment in WAEC Mathematics Performance by Student Profile

Explanatory Factor	High Classroom–Low WAEC	Low Classroom–High WAEC
Examination anxiety	High	Moderate
Cognitive overload under time pressure	High	Low
Familiarity with WAEC scoring logic	Low	High
Strategic elimination and guessing	Low	High
Classroom–WAEC task misalignment	High	Moderate

Interpretation

Table 4 demonstrates that there is a varying mediation path to different misalignment profiles. High classroom students who do poorly in WAEC often complain about severe anxiety and cognitive overload, and high-stakes conditions are likely to inhibit their capacity to bring knowledge to bear. On the other hand, poor classroom students who score highly in WAEC have the advantage of knowing how to score logics and examination strategies that take advantage of examination structure. The misalignment of Classroom-WAEC tasks is observed to be most evident among the high-ability students,

which supports the argument that the design of assessments selects the comp**etencies** that are visible in the performance.

DISCUSSION

This research supports the idea that the connection between classroom achievements and high-stakes scores among students, especially in WAEC Mathematics, is a complicated one. Although most students showed consistency between performance in classrooms and performance in exams, considerable differences were noted in a group of students, indicating that there was ability-performance mismatch.

The research indicates that there are two different trends of misalignment:

i. Negative misalignment - in which high performing classroom students performed poorly in the WAEC Mathematics examination.

ii. Positive misalignment - low-performing students in the classroom performed better than expected in examination.

This twin phenomenon of misalignment raises the question of the classical presumption that the performance in examinations is an immediate measure of the underlying ability of the student. Rather, the findings are indicative that other factors like assessment design, test anxiety and cognitive overload in high-stakes situations play a significant mediating role. Test anxiety and stress-induced cognitive interference also seem to inhibit high-performing students to perform their best, which is consistent with the research results of cognitive psychology, which point to the idea that stress negatively affects memory retrieval and solving problems (Gellisch et al., 2024). The importance of these findings is that they promote a different perspective on academic performance that uses affective and psychological elements instead of the results of standardized tests.

However, with low-performing students, strategic test-taking behaviour, learning how the WAEC items are structured and external coaching were observed to be critical in enabling them to perform better than what they would expect. This brings up questions regarding the legitimacy of the assessment system in the real sense of assessing mathematical competence. This may be due to the fact that procedural knowledge is overrepresented in the WAEC exam and this is accompanied by the underrepresentation of higher-order thinking and this could be the reason the performance amplification in low

performers who approached the exam strategically. In that regard, it is necessary to admit that external training techniques and test-taking tricks can artificially boost the performance of a student, concealing his/her inner knowledge of mathematics.

Generally, the results support the idea of the necessity to reconstruct high-stakes assessments as the instruments of the reflection of the overall competencies of students, including more than procedural proficiency, and implement full frameworks integrating classroom tests and psychological and affective ones. Moreover, these findings underscore the importance of factoring in contextual factors, e.g. instructional practices and socioeconomic conditions into any student ability assessment.

Conclusion

This research gives some strong arguments that although the results of the WAEC Mathematics usually are a representation of the classroom performance of the students, there are still some major trends of misalignment which can no longer be overlooked. Such trends, including negative misalignment (when students who perform well in the classroom perform poorly) and positive misalignment (when students who perform poorly do better) indicate that high-stakes tests like WAEC do not necessarily give a true picture of the actual mathematical performance of students.

Its results indicate that misalignment is determined by various affective, cognitive and contextual elements, including test anxiety, time pressure and assessment design. The condition of high stakes testing, especially the cognitive overload due to stress, is a major contributor to the poor performance of the able students. On the other hand, the low-achieving students have the advantage of being conversant with the assessment formats and strategic test taking behaviours that help them to score high marks that are beyond their expectations.

Considering these results, it is important to redefine the meaning of high stakes examination results no longer depending on the raw score, but looking at the psychological, contextual, and instructional factors that affect performance. This paper proposes a more delicate way of assessing education, which incorporates classroom-based evidence and high-stakes examination results. In this way, teachers and policy makers can make sure that high stakes decisions, including certification, progression, and selection are made on a more holistic level and does not ignore the competency and potential of all students.

Lastly, the research paper demands a reformation in the assessment systems, in which tools should be created to balance the procedural knowledge with higher-order thinking, problem-solving skills, and conceptual knowledge. It is only through this balance that we can guarantee that high-stakes assessment offers a fair and valid assessment of the student performance as well as ensure that educational opportunities are equitable to all the learners.

Recommendation.

i. The education policymakers ought to re-evaluate the use of high-stakes tests such as the WAEC solely. The policies are to promote application of diversified assessments techniques that integrate continuous assessment, formative testing and psychological tests in order to realise the full diversity of talents in students.

ii. The educational policies need to have in them strategies to reduce test anxiety, which may involve the incorporation of stress-management programs and time management programs into the curriculum. This would contribute towards reducing the adverse effects of anxiety on the performance of the students in high stakes tests.

iii. Stakes Exams: Policies of redesigning high-stakes exams should be centered on make sure that they more closely reflect instruction in classroom, with a higher emphasis on procedural skills as well as conceptual knowledge. It is possible to do this through diversification of the types of questions and has to make exams evaluate higher-order thinking and problem-solving.

REFERENCES

Abbas, A., Ashwin, P., & McLean, M. (2013). Qualitative life-grids: A proposed method for comparative European educational research. *European Educational Research Journal, 12*(3), 320–329. https://doi.org/10.2304/eerj.2013.12.3.320

Almarzouki, A. F. (2024). Stress, working memory, and academic performance: A neuroscience perspective. *Stress, 27*(1), Article 2364333. https://doi.org/10.1080/10253890.2024.2364333

Amrein, A. L., & Berliner, D. C. (2002). *An analysis of some unintended and negative consequences of high-stakes testing.* Education Policy Research Unit, Education Policy Studies Laboratory, Arizona State University. https://nepc.colorado.edu/files/EPSL-0211-125-EPRU.pdf

Asamoah, D., Shahrill, M., & Latif, S. N. A. (2024). Teachers' perceptions of school assessment climate and realities of assessment practices in two educational contexts. *Frontiers in Education, 9*, Article 1278187. https://doi.org/10.3389/feduc.2024.1278187

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., & Beekhuyzen, J. (2015). Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support. *Communications of the Association for Information Systems, 37*, 154–204. https://doi.org/10.17705/1CAIS.03708

Bekoe, S. O. (2023). Towards a theory: The constraining hold of high-stakes external assessment on curriculum implementation. *Creative Education, 14*(5), 993–1008. https://doi.org/10.4236/ce.2023.145063

Çevikbaş, M., Kaiser, G., & Schukajlow, S. (2024). Trends in mathematics education and insights from a meta-review and bibliometric analysis of review studies. *ZDM–Mathematics Education, 56*(2), 165–188. https://doi.org/10.1007/s11858-024-01587-7

Dogbey, J., & Dogbey, J. (2018). Depth of knowledge and context characteristics of the West African Examination Council's Core Mathematics assessment—The case of Ghana from 1993–2013. *Assessment in Education: Principles, Policy & Practice, 25*(4), 376–398. https://doi.org/10.1080/0969594X.2016.1192524

Federal Ministry of Education. (2014). *National policy on education* (6th ed.). Nigerian Educational Research and Development Council. https://education.gov.ng/wp-content/uploads/2020/06/NATIONAL-POLICY-ON-EDUCATION.pdf

Folson, D., & Awush, F. K. (2024). *Assessing cognitive alignment in pre-tertiary TVET core mathematics: A Ghanaian case study of curriculum and exit examination* [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-5711559/v1

Gellisch, M., Bablok, M., Brand-Saberi, B., & Schäfer, T. (2024). Neurobiological stress markers in educational research: A systematic review of physiological insights in health science education. *Trends in Neuroscience and Education, 37*, Article 100242. https://doi.org/10.1016/j.tine.2024.100242

Gopal, D., & Mofolo-Mbokane, B. (2021). *Application of Webb's depth of knowledge in mathematics assessment.*

Haudek, K. C., & Zhai, X. (2024). Examining the effect of assessment construct characteristics on machine learning scoring of scientific argumentation. *International Journal of Artificial Intelligence in Education, 34*(4), 1482–1509. https://doi.org/10.1007/s40593-023-00385-8

Heissel, J. A., Levy, D. J., & Adam, E. K. (2019). Stress, poverty, and academic performance. *Journal of Public Economics, 170*, 1–17.

International Commission on the Futures of Education. (2021). *Reimagining our futures together: A new social contract for education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379707

Kaiser, G., & Presmeg, N. (2019). Competing perspectives on mathematics education. *ZDM–Mathematics Education, 51*(1), 1–7.

Osei-Asibey, E. (2024). Correlates of Ghanaian teachers' understanding of mathematics strands and cognitive domains in Basic Education Certificate Examination. *Asian Research Journal of Mathematics, 20*(6), 41–59. https://doi.org/10.9734/arjom/2024/v20i6806

Pamungkas, M. D., Waluya, S. B., Mariani, S., & Isnarto. (2023). A systematic review of complex problem-solving in education and mathematics education. *Journal of Higher*

Education Theory and Practice, 23(16), 87–101. https://doi.org/10.33423/jhetp.v23i16.6465

Parrish, C. W., & Bryd, K. O. (2022). Cognitively demanding tasks: Supporting students and teachers during engagement and implementation. *International Electronic Journal of Mathematics Education, 17*(1), Article em0671. https://doi.org/10.29333/iejme/11475

Prediger, S., Götze, D., Holzäpfel, L., Rösken-Winter, B., & Selter, C. (2022). Five principles for high-quality mathematics teaching: Combining normative, epistemological, empirical, and pragmatic perspectives for specifying the content of professional development. *Frontiers in Education, 7*, Article 969212. https://doi.org/10.3389/feduc.2022.969212

Prokofieva, V., Kostromina, S., Polevaia, S., & Fenouillet, F. (2019). Understanding emotion-related processes in classroom activities through functional measurements. *Frontiers in Psychology, 10*, Article 2263. https://doi.org/10.3389/fpsyg.2019.02263

Qetrani, S., & Achtaich, N. (2022). Evaluation of procedural and conceptual knowledge of mathematical functions: A case study from Morocco. *Journal on Mathematics Education, 13*(2), 211–238. https://doi.org/10.22342/jme.v13i2.pp211-238

Rafi, I., Retnawati, H., Apino, E., Hadiana, D., Lydiati, I., & Rosyada, M. N. (2023). What might be frequently overlooked is actually still beneficial: Learning from post national-standardized school examination. *Pedagogical Research, 8*(1), Article em0145. https://doi.org/10.29333/pr/12657

Ravand, H., Shahi, R., Effatpanah, F., & Moghadamzadeh, A. (2025). Measuring cognitive levels in high-stakes testing: A CDM analysis of a university entrance examination using Bloom's taxonomy. *Frontiers in Education, 10*, Article 1644811. https://doi.org/10.3389/feduc.2025.1644811

Rosli, R., Goldsby, D., & Capraro, M. M. (2013). Assessing students' mathematical problem-solving and problem-posing skills. *Asian Social Science, 9*(16), 54–60. https://doi.org/10.5539/ass.v9n16p54

Saikang, T., Junpeng, P., Luanganggoon, N., Chiajunthuk, S., Suwannatrai, P., Marwiang, M., Tang, K. N., & Wilson, M. (2024). Developing and validating the construct maps to assess mathematical proficiencies. *International Journal of Evaluation and Research in Education, 13*(2), 968–978. https://doi.org/10.11591/ijere.v13i2.24871

Schmidt, M. M. (2022). *Cognitive demand of teacher-created mathematics assessments* [Master's thesis, Boise State University]. https://doi.org/10.18122/td.1995.boisestate

West African Examinations Council. (2023). *Chief examiners' report on the West African Senior School Certificate Examination: Mathematics*. https://waecgh.org/chief-examiners-report/

Yang, L., & Jamaludin, A. (2024). Test anxiety and academic performance in high-stakes examinations. *Frontiers in Psychology, 15*.

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