

ORIGINAL ARTICLE

Item Statistics Disparity between 2023 WASSCE and NECO SSCE Mathematics Large-Scale Assessments

Rioborue Alexander Oghenerume^{1b}

MEd, Measurement and Evaluation, University of Benin, Benin City, Nigeria.
ORCID: 0009-0008-7808-4810

Funding Information

No funding was received for the study.

Conflict of Interest

No conflict of interest is present in the conduction or the reporting of this study.

ABSTRACT

WAEC and NECO (2023) unveiled 16.29% of WASSCE Mathematics learners' results were withheld owing to assessment malpractice, while only 84.38% of WASSCE Mathematics learners merited credit, and 61.60% of NECO SSCE Mathematics learners merited credit in the Mathematics assessments overseen by both assessment bodies. These issues, with others, necessitated this exploration to assess the disparity in item statistics between 2023 WASSCE and NECO SSCE Mathematics large-scale assessments via IRT Three-Parameter Logistic Model. Three postulations were posed. This exploration applied survey research design. While this exploration's populace entailed 31,182 SS 3 Mathematics learners in 550 public and 850 private senior high schools in Enugu State, a multistage sampling tactic opted for 9,724. 2023 WASSCE and NECO SSCE Mathematics fixed-choice tests were applied as instruments for this exploration. Data amassed were probed via 3PLM within IRT framework, using flexMIRT psychometric software to produce the slopes, thresholds, and lower asymptotes, and independent samples t-test to verify all postulations for significance. The outcomes of this exploration proved a significant disparity ensued in the lower asymptotes amid 2023 WASSCE and NECO SSCE Mathematics fixed-choice items. Nonetheless, no significant disparities emerged in the thresholds and slopes amid these assessments' items. The realistic connotation of these outcomes inferred that harnessing IRT Three-Parameter Logistic Model could better spot Mathematics fixed-choice items formed by WAEC and NECO that are vulnerable to conjecturing, conceding for their deletion or amendment to mitigate guessing by learners in successive assessments, thereby boosting the entire attribute of their items.

Keywords: Item Statistics Disparity, WASSCE, NECO SSCE, Mathematics, IRT Three-Parameter Logistic Model

Received: 15/09/2024

Accepted: 03/03/2025

*Corresponding Author

Rioborue Alexander Oghenerume, English Language Teaching, Çanakkale Onsekiz Mart University, Türkiye
Email: raoghenerume@gmail.com

INTRODUCTION

Mathematics is a deep-rooted domain of erudition that cuts across diverse intellectual disciplines, and it entails methodical analysis of numbers, quantities, shapes, and patterns, in conjunction with their intricate associations, expending logic and abstract cognition. According to Rosen (2019), Mathematics has a wide range of applications and seeks to provide solutions to real-life issues. Irrefutably, Mathematics is a mandatory subject at all levels of Nigerian education from primary school through secondary school (Adeniyi & Akinoso, 2020; Okafor & Anaduaka, 2013). Mathematics is a core aspect of the school curricula, set by the Nigerian Educational Research and Development Council (NERDC) reflecting its significance in nurturing critical thinking abilities and preparing learners for various fields, including science, engineering, and technology (Federal Republic of Nigeria, 2013). While Mathematics serves as a baseline for learners to gain entry into the university to study any course of their choice, it has an intertwined relationship with other subjects, and as such, it grooms learners with diverse dexterities they need to gain in order to thrive in their varied chosen careers. To facilitate the transition of Mathematics learners from secondary school to university, and to ensure their critical thinking abilities corroborate with curricula benchmarks, two significant assessment bodies—the National Examinations Council (NECO) and the West African Examinations Council (WAEC)—coordinate annual large-scale assessments using standardised Mathematics tests. These large-scale assessments appraise the learners' mastery of a broad range of mathematical topics and verify their eligibility for the secondary school leaving certificate. Oghenerume and Uyi-Osaretin (2024) divulged that these assessment bodies are predominantly in the dominance of Senior Secondary Certificate Examinations (SSCE). As NECO superintends the National Examinations Council Senior School Certificate Examination (NECO SSCE) for Nigerian learners, WAEC superintends the West African Senior School Certificate Examination (WASSCE) for Nigerian, Gambian, Ghanaian, Sierra Leonean, and Liberian learners (Oghenerume & Uyi-Osaretin, 2024; Oghenerume & Egberha, 2024). The NECO SSCE is analogous to the WASSCE, but it is chiefly reserved for Nigerian learners (Oghenerume & Egberha, 2024; Oghenerume & Uyi-Osaretin, 2024).

These Mathematics standardised tests go through item statistics analytics to measure the power of every single item and likewise detect any that could either necessitate an amendment or riddance. They inevitably entail fixed-choice and free-response items that measure the learners' erudition and wittiness of the discipline. Typically, in objective tests, the fixed-choice items assess learners' aptitude to logically apply a wide range of mathematical principles within the syllabi. Some of these fixed-choice items involve complex problem-solving scenarios or require learners to interpret data or graphs. Predominantly in essay tests, the free-response items sternly necessitate the learners to crack complicated difficulties that unveil their mathematical dexterity. Oghenerume (2022) explicated that item statistics are quantitative measures that capture the numerical properties and characteristics of test items, and these quantitative measures encompass various aspects of test items, including their parameters. For any item to be deemed an appropriate or fitting item, it ought to comply with the three parameters: slope, threshold, and lower asymptote. Slope denotes an item's adeptness to split among test-takers with differing points of the construct being assessed, threshold is how challenging an item is for test-takers, and lower asymptote quantifies the possibility that a test-taker with minimal adeptness would reply an item correctly purely by chance (Oghenerume, 2022). This is because items should not excessively discriminate, be overly difficult, or vulnerable to guessing for learners (Oghenerume, 2022; Oghenerume & Uyi-Osaretin, 2024).

Item statistics disparity denotes discrepancies or variations in the statistical properties of test items, predominantly



when comparing tests or large-scale assessments across diverse contexts, groups, or assessment bodies, with these disparities arising in terms of slopes, thresholds, lower asymptotes, and item-total correlation. The adeptness of an item to distinguish between upper and lower achieving learners may vary across large-scale assessments. An item that meritoriously distinguish between learners in one large-scale assessment (e.g., WASSCE) may function less well in another (e.g., NECO SSCE), possibly owing to disparities in the large-scale assessment configuration or curriculum target. Items may diverge in threshold across diverse assessments or for dissimilar populations. For instance, a similar item that appears in WASSCE and NECO SSCE together may have a dissimilar proportion of precise rejoinders due to discrepancies in content emphasis, directive, or learner groundwork. On objective tests, the odds that a learner can gamble the accurate rejoinder may diverge between assessments. This could be a result of disparities in the structure of the distractors (imprecise alternatives) or the quantity of rejoinder choices given. Additionally, item statistics disparity can occur due to variations in item-total correlation. The association between separate items and overall assessment feat can diverge between diverse assessments or groups. This can signify that items may be more or less aligned with the overall assessment intents in diverse contexts. Entrancingly, Aborisade and Fajobi (2020) previously reported a significant disparity in the slopes of items in the fixed-choice Mathematics tests formed by WAEC and NECO for the year 2017. Oghenerume and Egberha (2024) and Oghenerume and Uyi-Osaretin (2024) reported significant disparities in the lower asymptotes of items in the Data Processing and English Language fixed-choice tests, respectively, for the year 2023 in WASSCE and NECO SSCE. Could these three reports be affiliated with the fixed-choice items in the Mathematics large-scale assessments of WASSCE and NECO SSCE for the year 2023? This was one of the lacunas this recent study sought to tackle. While the previous explorations focused solely on public senior high schools, this current study included both public and private senior high schools, filling another lacuna. This allowed for a broader exploration of item statistics disparities, resulting in a more robust and generalisable report.

Literature Review

This latest exploration pivoted on Three-Parameter Logistic Model (3PLM) by Birnbaum (1968) within Item Response Theory (IRT) framework by Thurstone (1925) as its theoretical framework. Louis Thurstone introduced Item Response Theory (IRT) and set up its conceptual underpinning in his exploration, "A Method of Scaling Psychological and Educational Tests" (Oghenerume & Egberha, 2024; van der Linden; 2010). In this influential work, Thurstone provided a procedure for accurately positioning the items on the Binet scale (van der Linden, 2010; Oghenerume, 2022; Oghenerume & Egberha, 2024). Globally, IRT is extensively applied across various fields, such as education, psychology, and healthcare (Gerber & Price, 2018; Linn, 2010; Carlson & von Davier, 2017; Glas, 2019). In education, IRT supports large-scale assessments like PISA and NAEP (Carlson & von Davier, 2017). It also facilitates adaptive testing, which provides more efficient assessments tailored to individual aptitudes (van der Linden, 2010; Carlson & von Davier, 2017). In psychology, IRT refines the precision of measurement scales for latent traits, such as personality features (Zanon et al., 2016; Hambleton et al., 2000; Hambleton & Jodoin, 2003). Additionally, in healthcare, IRT refines diagnostic tools and patient-reported outcomes by opting for the most informative items for evaluation, ensuring that assessments are both reliable and valid (Gerber & Price, 2018; Hambleton, 2005). According to Cai and Huang (2022), Item Response Theory (IRT) can be viewed minimally as a set of psychometric models for categorical item-level response data. The data amassed from respondents to items with multiple options or classifications which are conventionally discrete and qualitative, signifying diverse levels or substitutes are known as categorical item-level response data (von Davier et al., 2021). Objective test items in the form of fixed-choice options with predefined classifications or distinct answer

choices, such as strongly agree, agree, disagree, and strongly disagree are examples of item-level response data (Oghenerume, 2022; von Davier et al., 2021; Carlson, 2020).

Birnbaum (1968) expanded on IRT by proposing the Three-Parameter Logistic Model (3PLM), which posits that the probability of a correct response to an item depends on three parameters: slope 'a' (discrimination), threshold 'b' (difficulty), and lower asymptote 'c' (guessing). Based on the analysis provided by Glas (2019), the Three-Parameter Logistic Model (3PLM) should be considered the optimal choice for interpreting data from large-scale assessments. As reported by Zanon et al. (2016) and Oghenerume and Egberha (2024), the Three-Parameter Logistic (3PLM) authorises items to diverge in their knack to distinct between test-takers with divergent proficiency altitudes and in their arduousness, likewise accounting for the possibility that test-takers with very low proficiency might conjecture the precise rejoinders to items (Baker & Kim, 2023; Oghenerume & Uyi-Osaretin, 2024; Baker & Kim, 2017; Reise & Moore, 2023; Carlson, 2020; Bock & Gibbons, 2021). Additionally, Birnbaum (1968) put forward the psychometric formula of the Three-Parameter Logistic Model (3PLM), which depicts the possibility that an unintentionally chosen test-taker with competence 'θ' at measure 'k' will perfectly respond to item 'j':

$$P(x_j=1 | \theta_k, a_j, b_j, c_j) = c_j + \frac{(1-c_j)}{1+e^{-Dz_j(\theta_1-b_j)}}$$

Wherein

x_j = rejoinder to item j;

a_j = slope of item j, depicting the discrimination;

b_j = threshold of item j, typifying the difficulty;

c_j = lower asymptote of item j, typifying the odds of test-takers with awfully least adeptness gambling the accurate rejoinder;

D = interpretive quantifying factor chiefly set to 1.7 to vague aftermaths in a usual ogive simulation.

Three-Parameter Logistic Model (3PLM) within IRT framework is germane to this exploration as it functions on programmed indispensable assumptions once harnessed to the advanced statistical analytics of assessments data, chiefly between the context of Mathematics fixed-choice tests in 2023 WASSCE and NECO SSCE. As clarified by Baker and Kim (2023), de Ayala (2022), Bock and Gibbons (2021), Carlson (2020), Baker and Kim (2017), and Ojerinde et al. (2014), these assumptions are crucial for the accurate application of this model. Firstly, it is assumed that the modeled latent trait accurately describes the entirety of the underlying competence and effectively accounts for test-taker performance (Baker & Kim, 2017; Ojerinde et al., 2014; Carlson, 2020; Bock & Gibbons, 2021; de Ayala, 2022). In many cases, tests assume the necessity of a single latent ability, a concept known as unidimensionality (Carlson, 2020; Oghenerume, 2022; Reise & Moore, 2023; Oghenerume & Uyi-Osaretin, 2024; Baker & Kim, 2017; Oghenerume & Egberha, 2024; de Ayala, 2022). Secondly, it is assumed that all items are locally independent when conditioned on the latent trait, meaning an examinee's response to a specific item is not influenced by their responses to other items (Baker & Kim, 2017; Ojerinde et al., 2014; Carlson, 2020; Bock & Gibbons, 2021; de Ayala, 2022). Additionally, it is assumed that an examinee's response to a test item can be effectively modeled using the Item Response Function (IRF) (Bock & Gibbons, 2021; Baker & Kim, 2023; Ojerinde et al., 2014; de Ayala, 2022; Baker & Kim, 2017; Baker & Kim, 2023). Depending on the specific IRT model, for instance 3PLM, the IRF enunciates the probability of achieving a particular score 'X' at various levels of the latent ability, denoted as θ (Baker & Kim, 2023; de Ayala, 2022; Oghenerume & Uyi-



Osaretin, 2024; Baker & Kim, 2017; Bock and Gibbons, 2021; Oghenerume & Egberha, 2024). This probability relationship remains invariant across different distributions of θ within the populace (Carlson, 2020; Bock & Gibbons, 2021; Oghenerume, 2022; Oghenerume & Uyi-Osaretin, 2024; Oghenerume & Egberha, 2024; Baker & Kim, 2023; de Ayala, 2022; Baker & Kim, 2017).

de Ayala (2022) explained that slope (a) is typically expressed as a value between -2 and +2, where values below 1 may indicate weaker discrimination, and values above 1.5 may suggest excessive discrimination. The threshold (b), representing difficulty, ranges from -3 to +3, with values below 1.5 indicating easier items and values above 2.5 indicating more difficult items (Baker & Kim, 2023; Oghenerume & Uyi-Osaretin, 2024; van der Linden & Hambleton, 2019; Oghenerume & Egberha, 2024; Carlson, 2020; Baker & Kim, 2017). The lower asymptote (c) is expressed as a value between 0 and 1 (van der Linden & Hambleton, 2019; Oghenerume, 2022; de Ayala, 2022). Bock and Gibbons (2021) asserted that there is no universally agreed-upon specific numerical value for 'moderate' slope, threshold, and lower asymptote. However, a commonly referenced guideline suggests a slope value between 0 and 2, a threshold value between -3 and +3, and a lower asymptote value between 0 and 1 (Baker & Kim, 2017; Oghenerume & Uyi-Osaretin, 2024; Baker & Kim, 2023; van der Linden & Hambleton, 2019; Carlson, 2020). Essentially, lower asymptote values exceeding 0.35 are deemed unacceptable (Baker & Kim, 2023; de Ayala, 2022; Oghenerume & Uyi-Osaretin, 2024; Carlson, 2020; Oghenerume & Egberha, 2024; Bock & Gibbons, 2021). Typically, 5-option items have lower asymptote rates around 0.20, and 4-option items have lower asymptote rates around 0.25 (Baker & Kim, 2017; van der Linden & Hambleton, 2019; Carlson, 2020; Oghenerume & Egberha, 2024; de Ayala, 2022; Baker & Kim, 2023).

Aborisade and Fajobi (2020) did an exploration on the comparative analysis of psychometric properties of Mathematics items in South Western Nigeria. Their exploration dealt with the affiliation of Mathematics fixed-choice tests formed by NECO and WAEC. Their exploration applied IRT and concerted on the threshold, slope, and lower asymptote. Survey system of research was executed. The populace entailed senior secondary schools' learners in South Western Nigeria who enrolled for WAEC and NECO 2019 large-scale assessments. Through a multistage sampling tactic, one thousand, two hundred learners were sampled in their exploration. The WAEC and NECO's formed fixed-choice items in Mathematics were harnessed as their instruments. Their exploration's outcomes signified that the thresholds and lower asymptotes of NECO and WAEC's formed fixed-choice items in Mathematics had no significant disparity. Oppositely, a significant disparity emerged between the slopes of the items in these assessments.

Oghenerume and Egberha (2024) conducted a comparative analysis of item statistics for the 2023 WASSCE and NECO SSCE Data Processing fixed-choice tests using Item Response Theory (IRT). Their exploration sought to juxtapose the item statistics of the 2023 WASSCE and NECO SSCE Data Processing fixed-choice tests via the application of the IRT 3-Parameter Logistic Model (3PLM). This exploration was based on three postulations. As a form of design, survey research was engaged, entailing a populace of 10,800 SS 3 Data Processing learners from twenty one public senior high schools in Abuja Municipal Area Council. Multistage sampling tactic was applied to opt for 1,080 learners in Data Processing. The instruments applied were the 2023 Data Processing fixed-choice tests from both WASSCE and NECO SSCE. Data were probed via 3PLM applying jMetrik software, and independent samples t-test verified the postulations. The outcomes exhibited no significant disparities in the slopes and thresholds of the items between 2023 NECO SSCE and WASSCE Data Processing fixed-choice-tests. Oppositely, a significant disparity was

spotted between these large-scale assessments' lower asymptotes.

Oghenerume and Uyi-Osaretin (2024) performed an analytical comparison of item statistics utilising IRT in the 2023 WASSCE and NECO SSCE English Language fixed-choice tests. The intent of this analysis was to explore and compare the item statistics of these large-scale assessments through the application of the Item Response Theory Three-Parameter Logistic Model (IRT 3PLM). The exploration was based on three postulations. As a form of design, survey research was engaged, entailing a populace of 72,400 SS 3 English Language learners from Ogun State's 384 public senior high schools. Multistage sampling tactic was applied to opt for 8,000 learners. These assessment tools included the 2023 English Language fixed-choice tests from both assessment bodies. Data were probed via 3PLM using jMetrik software, and independent samples t-test verified the postulations. The aftermaths uncovered a significant disparity in the lower asymptote between the 2023 NECO SSCE and WASSCE English Language fixed-choice tests. Oppositely, no significant disparities were spotted between these large-scale assessments' slopes and thresholds.

Statement of the Problem

The ongoing decline in the feat of senior learners in Mathematics raises significant concerns about the worth of education and the efficacy of existing assessment approaches. Despite claims from the West African Examinations Council (WAEC) that its mid-year 2023 WASSCE Mathematics fixed-choice test maintained consistent standards, widespread criticism continues. Detractors argued that the test items are overly arduous and misaligned with the curriculum, putting learners at a detriment. Similarly, the National Examinations Council (NECO) encounters accusations of producing tests perceived as too lenient and insufficient in assessing learners' erudition comprehensively. These divergent views on the attributes of WAEC and NECO large-scale assessments have resulted in growing skepticism about the consistency of these assessments as measures of learners' erudition. Historical perspectives reinforce these concerns. Daniel (2005) slammed NECO for its substandard test item attribute, resulting in some federal universities refusing NECO results from 2002 to 2012, as verified by Peter (2012). Nonetheless, Ahmed (2014) later argued that NECO's assessment items between 2011 and 2014 divulged superior benchmarks than WAEC's. Dibu-Ojerinde and Faleye (2005) attested no significant disparities between the test items of the two assessment bodies, while Oghenerume (2022) confirmed similar outcomes regarding the slopes and thresholds of 2022 WASSCE and NECO SSCE Data Processing test items.

Further reports have divulged disparities that warranted deeper investigation. Aborisade and Fajobi (2020) reported significant disparities in the slopes of Mathematics fixed-choice items in the 2017 WAEC and NECO tests. Of late, Oghenerume and Egberha (2024) and Oghenerume and Uyi-Osaretin (2024) found substantial disparities in the lower asymptotes of items in the Data Processing and English Language fixed-choice tests for the 2023 WASSCE and NECO SSCE. These reports raised an imperative query: could similar disparities transpire in the item statistics of the 2023 Mathematics fixed-choice tests in WASSCE and NECO SSCE? Adding to the complexity, assessment malpractice continues to plague the integrity of assessments. Conspicuously, as unveiled by WAEC (2023), "16.29% of WAEC learners' results in Mathematics for the year 2023 were withheld owing to assessment malpractice." This raises further concerns about whether the threshold and alignment of test items may have contributed to dishonest practices. While malpractice was less reported in NECO, the disparity in feat between WAEC (84.38% of learners, merited credit) and NECO (61.60% of learners, merited credit), as reported by WAEC (2023) and NECO (2023), raises queries about whether the attribute of the test items—predominantly their slopes, thresholds, and lower asymptotes—might be a driving factor.



In light of persistent concerns, historical criticisms, recent statistical evidence, the issue of assessment malpractice, and the disparity in feat between WAEC and NECO learners' results, the need for a rigorous, data-driven probing of the item statistics disparity between 2023 WASSCE and NECO SSCE Mathematics fixed-choice tests via IRT Three-Parameter Logistic Model became apparent. Grasping these factors (slope, threshold, and lower asymptote) is imperative in verifying whether they add to the observed disparities in learners' feat and whether the items themselves mirrored standardised, large-scale assessments that accurately measure learners' erudition and adeptness. Additionally, this exploration is significant for tackling the long-standing anomalies that have triggered concerns among stakeholders, including learners, policymakers, and educators.

Queries

1. Could there be a significant disparity between the 2023 WASSCE and NECO SSCE slopes of Mathematics fixed-choice items?
2. Could there be a significant disparity between the 2023 WASSCE and NECO SSCE thresholds of Mathematics fixed-choice items?
3. Could there be a significant disparity between the 2023 WASSCE and NECO SSCE lower asymptotes of Mathematics fixed-choice items?

Postulations

1. No significant disparity emerges between the 2023 WASSCE and NECO SSCE slopes of Mathematics fixed-choice items.
2. No significant disparity transpires between the 2023 WASSCE and NECO SSCE thresholds of Mathematics fixed-choice items.
3. No significant disparity ensues between the 2023 WASSCE and NECO SSCE lower asymptotes of Mathematics fixed-choice items.

METHOD

Research Design

This exploration applied survey research as its design since this exploration engaged a set of learners from which experiential evidences were obtained, and no more than a few learners were deemed as delegates of the wide-ranging learners sought after. As a form of design, survey research is apt for this exploration since both 2023 WASSCE and NECO SSCE Mathematics fixed-choice tests were used to get experiential evidence and probe slopes, thresholds, and lower asymptotes of the items harnessing the IRT Three-Parameter Logistic Model.

Setting and Participants

Enugu State Post-Primary Schools Management Board (2023) divulged that Enugu State has three senatorial districts—Enugu West with 469 schools, Enugu North with 472 schools, and Enugu East with 459 schools—and twenty five Local Government Areas (LGAs), with a populace of 31,182 SS3 Mathematics learners, incorporating 13,679 learners in 550 public senior high schools and 17,503 learners in 850 private senior high schools. Conversely, this exploration's



statistical populace entailed fifty WASSCE Mathematics fixed-choice test items and sixty NECO SSCE Mathematics fixed-choice test items. This exploration's sample entailed 9,724 SS3 Mathematics learners within 442 schools in Enugu State, typifying approximately 31.18% of the populace. These learners were opted for, utilising multistage sampling tactic.

In phase one, the stratified random sampling tactic was used to opt for nine LGAs (three from each of the three Senatorial Districts in Enugu State), typifying 36% of the twenty five LGAs in the state. In phase two, a proportionate stratified random sampling tactic was utilised to opt for 442 schools based on their taxonomy as public or private inside the hitherto designated nine LGAs. This was done to account for the varying number of schools and learners in each LGA. Specifically, 221 public schools and 221 private schools were opted for. In phase three, the stratified random sampling tactic was applied to stratify in lieu of gender within the 442 schools (221 public and 221 private). In phase four, the simple random sampling tactic via balloting was used to opt for 11 lads and 11 lassies per school. Accordingly, 22 learners were sampled per school, resulting in a sum of 9,724 learners sampled inside the 442 designated schools.

Instruments

This exploration utilised the Mathematics fixed-choice items in 2023 WASSCE and NECO SSCE as its instruments. Mathematics fixed-choice test in the 2023 WASSCE entails 50 items, while the 2023 NECO SSCE entails 60 Mathematics fixed-choice items. While WASSCE Mathematics fixed-choice items are formed with four alternatives, 'A' - 'D,' whereby learners are necessitated to opt for the precise rejoinder, the corresponding NECO SSCE fixed-choice items are formed with five alternatives, 'A' - 'E,' whereby learners are necessitated to opt for the precise rejoinder. These instruments' validity and dependability were verified, given WAEC and NECO's reputation for producing rigorous standardised tests. Additionally, the items were within the necessitated three-year period for maintaining dependability before re-evaluation. To further corroborate this, the researcher applied a test-retest strategy. Initially, the instruments were dispensed to 20 SS3 learners from schools in Enugu. Following a two-week interim, the instruments were re-dispensed to these learners. The outcomes from these two dispensations were then analysed utilising Pearson's Product Moment Correlation to verify their dependability. The correlation coefficient was 0.94 for the 2023 WASSCE Mathematics fixed-choice items and 0.96 for the 2023 NECO SSCE Mathematics fixed-choice items, further corroborating the high dependability of these standardised instruments.

Procedure

Afterwards, the verified instruments were dispensed straightly to the Mathematics SS3 learners in the designated schools in Enugu, via research assistants, who were the experienced and qualified Mathematics educators at these schools. First, the WASSCE fixed-choice test was dispensed. After it was concluded, instant collection was done and marked. The following day, the NECO SSCE fixed-choice test was dispensed, concluded, and marked using the same tactic. Right through, unethical factors were assiduously eschewed, ensuring voluntary engagement, notified endorsement, and learners' data concealment was sternly upheld.

Data Analysis

Data amassed were probed via 3PLM within the IRT framework, utilising flexMIRT psychometric software to produce the slopes, thresholds, and lower asymptotes. According to Fu (2020), flexMIRT is highly recommended for calibrating large-scale assessment data due to its significant advantages: it is well-documented, user-friendly, it offers fast



processing time, and provides MLEs for ability parameters and incorporates many familiar fit statistics (Fu, 2020). Notably, there are no obvious disadvantages associated with using flexMIRT (Fu, 2020). Conversely, an independent samples t-test was harnessed to verify all postulations for significance.

RESULTS AND DISCUSSION

Results

Postulation 1: No significant disparity stems between the slopes of 2023 WASSCE and NECO SSCE Mathematics fixed-choice items.

Table 1: Slopes between 2023 WASSCE and NECO SSCE Mathematics Fixed-Choice Items

Mathematics	n	M	SD	DF	t	p	Verdict
2023 WASSCE	50	1.06	.47				
				108	1.512	.134	Insignificant
2023 NECO SSCE	60	.87	.78				

$\alpha = .05$

Table 1 divulged that the 2023 WASSCE Mathematics fixed-choice test comprised n=50 items, while the 2023 NECO SSCE Mathematics fixed-choice test contained n=60 items. The corresponding mean values were found to be 1.06 and 0.87, with standard deviations of .47 and .78, respectively. The associated degrees of freedom, t-statistic, and value of 'p' are 108, 1.512, and .134, respectively. Given the p-value surpasses the significance level, unverified postulation expressing "no significant disparity stems between slopes of 2023 WASSCE and NECO SSCE Mathematics fixed-choice items," was approved. This proved none of the Mathematics fixed-choice tests formed by both assessment bodies was evidently more distinguishing than the other.

Postulation 2: No significant disparity emerges between the thresholds of 2023 WASSCE and NECO SSCE Mathematics fixed-choice items.

Table 2: Thresholds between 2023 WASSCE and NECO SSCE Mathematics Fixed-Choice Items

Mathematics	n	M	SD	DF	t	p	Verdict
2023 WASSCE	50	2.07	.40				
				108	.486	.628	Insignificant
2023 NECO SSCE	60	2.03	.40				

$\alpha = .05$

Table 2 unveiled that the 2023 WASSCE Mathematics fixed-choice test comprised n=50 items, while the 2023

NECO SSCE Mathematics fixed-choice test contained $n=60$ items. The corresponding mean values were found to be 2.07 and 2.03, with standard deviations of .40 and .40, respectively. The associated degrees of freedom, t-statistic, and value of 'p' are 108, .486, and .628, respectively. Given the p-value surpasses the significance level, unverified postulation expressing, "no significant disparity emerges between the thresholds of 2023 WASSCE and NECO SSCE Mathematics fixed-choice items," was upheld. This proved none of the Mathematics fixed-choice tests formed by both assessment bodies was conspicuously more arduous than the other.

Postulation 3: No significant disparity ensues between the lower asymptotes of 2023 WASSCE and NECO SSCE Mathematics fixed-choice items.

Table 3: Lower Asymptotes between 2023 WASSCE and NECO SSCE Mathematics Fixed-Choice Items

Mathematics	n	M	SD	DF	t	p	Verdict
2023 WASSCE	50	.07	.13				
				108	3.051	.003	Significant
2023 NECO SSCE	60	.01	.07				

$\alpha = .05$

Table 3 substantiated that the 2023 WASSCE Mathematics fixed-choice test encompassed $n=50$ items, while the 2023 NECO SSCE Mathematics fixed-choice test comprised $n=60$ items. The corresponding mean values were found to be .07 and .01, with standard deviations of .13 and .07. The associated degrees of freedom, t-statistic, and value of 'p' are 108, 3.051, and .003, respectively. Given the p-value falls short of the significance level, unverified postulation expressing, "no significant disparity ensues between the lower asymptotes of 2023 WASSCE and NECO SSCE Mathematics fixed-choice items," was refuted. This proved WAEC's Mathematics fixed-choice test items were significantly more vulnerable to guessing than the equivalent NECO items.

Findings

Postulation one proved none of the Mathematics fixed-choice tests formed by both assessment bodies was evidently more distinguishing than the other. This report is in nexus with those of Oghenerume and Egberha (2024) and Oghenerume and Uyi-Osaretin (2024), who spotted no significant disparities in the slopes of items in the Data Processing and English Language fixed-choice tests for the 2023 WASSCE and NECO SSCE, respectively. Oppositely, this report contradicts that of Aborisade and Fajobi (2020), who harnessed IRT to explore the comparative analysis of psychometric properties of Mathematics items formed by the same assessment bodies and reported a significant disparity in the slopes of items in the fixed-choice Mathematics tests for the year 2017. This contradiction may be attributed to divergences in the methodologies used, as Aborisade and Fajobi (2020) focused on public schools, while this current exploration incorporated both public and private schools and utilised flexMIRT psychometric software. Fu (2020) emphasised flexMIRT as a top choice for calibrating large-scale assessment data because it is well-documented,



easy to use, processes data quickly, offers MLEs for competence parameters, and includes familiar fit statistics. Importantly, no significant drawbacks have been identified in employing flexMIRT (Fu, 2020).

Postulation two proved none of the Mathematics fixed-choice tests formed by both assessment bodies was conspicuously more arduous than the other. This report associates with that of Aborisade and Fajobi (2020), who previously conducted an exploration via IRT on the comparative analysis of psychometric properties of Mathematics items formed by the same assessment bodies and reported a significant disparity in the thresholds of items in the Mathematics fixed-choice tests for the year 2017. This current outcome is consistent with the verdicts of Oghenerume and Egberha (2024), who executed a relative exploration of item statistics for 2023 WASSCE and NECO SSCE Data Processing fixed-choice tests expending IRT, and spotted no significant disparity in the thresholds of items in Data Processing. Correspondingly, it correlates with the verdicts of Oghenerume and Uyi-Osaretin (2024), who employed IRT in their exploration, and spotted no significant disparity in the thresholds of items in the English Language fixed-choice tests for the year 2023.

Postulation three proved WAEC's Mathematics fixed-choice test items were significantly more vulnerable to guessing than the equivalent NECO items. This newest discovery is analogous to the verdicts of Oghenerume and Egberha (2024), who executed a relative exploration of item statistics for 2023 WASSCE and NECO SSCE Data Processing fixed-choice tests applying IRT, and spotted a significant disparity in the lower asymptotes of items in Data Processing fixed-choice tests. Congruently, this current observation is akin to the verdicts of Oghenerume and Uyi-Osaretin (2024), who applied IRT in their exploration, and spotted no significant disparity in the lower asymptotes of items in the English Language fixed-choice tests. Oppositely, this current observation refutes that of Aborisade and Fajobi (2020), who harnessed IRT to explore the comparative analysis of psychometric properties of Mathematics items formed by the same assessment bodies and reported no significant disparity in the lower asymptotes of items in the fixed-choice Mathematics tests for the year 2017. This refutation may be attributed to variations in the methodologies employed. Aborisade and Fajobi (2020) did their investigation exclusively on public schools, whereas this current exploration broadened its scope to include both public and private schools. Additionally, this newest exploration utilised flexMIRT psychometric software to analyse the data, which may have contributed to the observed variations. As noted by Fu (2020), flexMIRT stands out for calibrating large-scale assessment data due to its comprehensive documentation, user-friendly interface, fast data processing, provision of MLEs for adeptness parameters, and inclusion of common fit statistics. Additionally, Fu (2020) found no significant limitations associated with its use.

CONCLUSION AND RECOMMENDATIONS

Given the outcomes in this exploration, it was resolved that a significant disparity ensued in the lower asymptotes between 2023 WASSCE and NECO SSCE Mathematics fixed-choice items. Nonetheless, no significant disparities emerged in the thresholds and slopes between these items in the large-scale assessments. The practical inference of these outcomes suggests that harnessing the IRT Three-Parameter Logistic Model could more effectually spot Mathematics fixed-choice items created by WAEC and NECO that are predisposed to conjecturing. This would allow for their deletion or revision to mitigate guessing by learners in successive large-scale assessments, thereby refining the overall statistical properties of the items.

Accordingly, in line with the outcomes of this concluded exploration, this report recommends that these assessment bodies could applied the IRT Three-Parameter Logistic Model to substantiate their Mathematics fixed-



choice tests' slopes, thresholds, and lower asymptotes prior to official dispensations yearly, if not currently being harnessed. As a consequence of harnessing this model, these assessment bodies would spot unfitting items such as those that are disproportionately distinguishing between learners, extremely arduous, or are vulnerable to being rejoined precisely by conjecturing.

This quantitative researcher proposes that further explorations be executed by probing item statistics disparity between WASSCE and NECO SSCE fixed-choice tests in other disciplines such as Biology, Christian Religious Studies (CRS), Civic Education, and Commerce, utilising the IRT Three-Parameter Logistic Model. These explorations should explore the similitudes and disparities in item attributes, thereby offering insights into the magnitude to which WAEC and NECO affiliate with each other and divulging plausible areas for amendment for both assessment bodies, resulting to a better erudition of their efficacy in measuring learners' feat. Additionally, a replica of this exploration may be undertaken with a divergent populace in any of Nigeria's North East, North West, and North Central geopolitical zones.

REFERENCES

- Aborisade, O. J., & Fajobi, O. O. (2020). Comparative analysis of psychometric properties of Mathematics items constructed by WAEC and NECO in Nigeria using item response theory approach. *Educational Research and Reviews*, 15(1), 1–7.
- Adeniyi, C. O., & Akinoso, S. O. (2020). Difficult concepts in Nigerian senior secondary school curriculum as perceived by students. *Ilorin Journal of Education*, 40, 114–124
- Ahmed, M. F. (2014). Difficulty index of Mathematics multiple-choice items of West African Examinations Council and National Examinations Council Senior Secondary School Certificate Examinations from 2006–2010. *Journal of ATIP*, 13, 25–31.
- Baker, F. B., & Kim, S.-H. (2017). *The basics of item response theory using R*. Springer. <https://doi.org/10.1007/978-3-319-54205-8>
- Baker, F. B., & Kim, S.-H. (Eds.). (2023). *Item response theory: Parameter estimation techniques* (2nd ed.). CRC Press.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 397–472). Reading, MA: Addison – Wesley Publishing.
- Bock, R. D., & Gibbons, R. D. (2021). *Item response theory*. Wiley.
- Cai, L., & Huang, S. (2022). Multidimensional item response theory. In *International Encyclopedia of Education* (4th ed., pp. 72–85). Elsevier.
- Carlson, J. E., & von Davier, M. (2017). Item response theory. In R. Bennett & M. von Davier (Eds.), *Advancing human assessment* (pp. 133–178). Springer. https://doi.org/10.1007/978-3-319-58689-2_5
- Carlson, J. E. (2020). *Introduction to item response theory models and applications*. Routledge.
- Daniel, F. (2005). A survey of teachers' and students' opinions on WAEC and NECO SSCE examinations in Mathematics. *Abuja Journal of Education*, 6(1), 32–39.
- de Ayala, R. J. (2022). *The theory and practice of item response theory* (2nd ed.). The Guilford Press.
- Dibu-Ojerinde, O. O., & Faleye, B. A. (2005). Do they end at the same point? *European Journal of Scientific Research*, 11(3), 449–452.
- Enugu State Post-Primary Schools Management Board (2023). *Statistical data of registered Mathematics learners, public and private senior secondary schools in Enugu State*.
- Federal Republic of Nigeria. (2013). *National policy on education* (4th ed.). NERDC Press.
- Fu, J. (2020). *A preliminary comparison of five software applications to estimate unidimensional item response theory models* (Research Memorandum No. RM-20-02). Educational Testing Service.
- Gerber, N. L., & Price, J. K. (2018). Measures of function and health-related quality of life. In *Principles and practice of clinical research* (4th ed., pp. 303–315). Elsevier. <https://doi.org/10.1016/B978-0-12-849905-4.00021-6>
- Glas, C. A. W. (2019). Reliability issues in high-stakes educational tests. In B. Veldkamp & C. Sluiter (Eds.), *Theoretical and practical advances in computer-based educational measurement* (pp. 213–230). Springer Cham. https://doi.org/10.1007/978-3-030-18480-3_11
- Hambleton, R. K., Robin, F., & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 553–585). Academic Press.
- Hambleton, R. K., & Jodoin, M. (2003). Item response theory: Models and features. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 509–514). Sage.

- Hambleton, R. K. (2005). Applications of item response theory to improve health outcomes assessment: Developing item banks, linking instruments, and computer-adaptive testing. In J. Lipscomb, C. C. Gotay, & C. Snyder (Eds.), *Outcomes assessment in cancer: Measures, methods and applications* (pp. 445–464). Cambridge University Press.
- Linn, R. L. (2010). Educational measurement: Overview. In *International Encyclopedia of Education* (3rd ed., pp. 45–49). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00243-8>
- National Examinations Council. (2023). *Statistical report for Mathematics*. <https://neco.gov.ng>
- Oghenerume, R. A. (2022). *Comparative analysis of item statistics of WASSCE and NECO SSCE 2022 Data Processing multiple choice tests using item response theory* [Unpublished master's thesis]. University of Benin. <http://dx.doi.org/10.13140/RG.2.2.33718.31046/1>
- Oghenerume, R. A., & Egberha, F. (2024). Comparative analysis of item statistics of WASSCE and NECO SSCE 2023 Data Processing multiple choice tests using item response theory. *International Journal of Educational Researchers*, 15(1), 58–67. <https://doi.org/10.29329/ijer.2024.652.4>
- Oghenerume, R. A., & Uyi-Osaretin, S. I. (2024). Analytical comparison of item statistics employing IRT 3PLM in WASSCE and NECO SSCE 2023 English Language multiple choice tests. *KIU Journal of Education*, 4(1), 69–78. <https://doi.org/10.59568/kjed-2024-4-1-07>
- Ojerinde, D., Popoola, K., Ojo, F., & Ariyo, A. (2014). *Practical applications of item response theory in large-scale assessment*. Marvelous Mike Press Limited.
- Okafor, C. F., & Anaduaka, U. S. (2013). Nigerian school children and Mathematics phobia: How the Mathematics teacher can help. *American Journal of Educational Research*, 1(7), 247–251.
- Peter, K. (2012). A study of the attitude of some Nigerian science students towards NECO and WAEC. *Journal of Professional Science and Vocational Teachers Association of Nigeria*, 12(1), 15–18.
- Reise, S. P., & Moore, T. M. (2023). Item response theory. In H. Cooper, M. N. Coutanche, L. M. McMullen, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Foundations, planning, measures, and psychometrics* (2nd ed., pp. 809–835). American Psychological Association. <https://doi.org/10.1037/0000318-037>
- Rosen, K. H. (2019). *Discrete Mathematics and its applications* (8th ed.). McGraw-Hill Education.
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16(7), 433–451. <https://doi.org/10.1037/h0073357>
- van der Linden, W. J. (2010). Item response theory. In *International Encyclopedia of Education* (3rd ed., pp. 81–88). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00250-5>
- van der Linden, W. J., & Hambleton, R. K. (Eds.). (2019). *Handbook of modern item response theory*. Springer.
- von Davier, A. A., Mislevy, R. J., & Hao, J. (Eds.). (2021). *Computational psychometrics: New methodologies for a new generation of digital learning and assessment* (1st ed.). Springer Cham. <https://doi.org/10.1007/978-3-030-74394-9>
- West Africa Examinations Council. (2023). *Statistical report for Mathematics*. <https://www.waeconline.org.ng>
- Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, 29, Article 18. <https://doi.org/10.1186/s41155-016-0056-7>