Psychometric Investigation of the 2025 WAEC English Language Objectives Test of Public Senior Secondary School Students in Rivers State

Orinemi Dennis Horsfall
Department of Business Education
Faculty of Education
Rivers State University
Port Harcourt, Nigeria
horsfall.orinemi@ust.edu.ng

&

Professor William J. Ubulom
Department of Business Education
Faculty of Education
Rivers State University
Port Harcourt, Nigeria
ubulom.william@ust.edu.ng

Abstract

This study investigated the validity, reliability, guessing parameters, and ability estimates of the 2025 West African Senior School Certificate Examination (WASSCE) English Language multiple-choice objective test of public senior secondary school students in Rivers State. A descriptive survey research design was adopted. The population comprised 59,223 SS3 students, from which 500 students were selected using a multi-stage sampling technique. Data were collected using the 2025 WASSCE English Language objective test. The validity of the instrument was assumed to have been established by WAEC through expert test development procedures, while a pilot study yielded a Kuder–Richardson Formula 20 (KR-20) reliability coefficient of 0.80. The instrument was administered under standardized examination conditions. Data were analyzed using Classical Test Theory (CTT) and Item Response Theory (IRT). Validity was assessed using Exploratory Factor Analysis, reliability using KR-20, guessing using the three-parameter logistic (3PL) model, and students' ability using theta (θ) estimates. The findings revealed that the test demonstrated acceptable validity, satisfactory reliability, generally low to moderate guessing effects, and effectively differentiated candidates across varying ability levels. The study concluded that the examination possessed acceptable psychometric quality but required improvement in a few highly guessable items and the inclusion of more difficult items. It was recommended that WAEC should strengthen routine psychometric validation, monitor reliability, revise highly guessable items, and include more difficult items to improve measurement precision.

Keywords: Ability, Evaluation, Guessing, Item, Psychometric, Validity, Reliability

Introduction

In many educational systems, particularly at the secondary school level, tests serve not only as evaluative tools that regulate access to higher education; they assume what is commonly referred to as a standardized test which occupies a distinctive position in the Nigerian educational system. Standardized tests are characterized by uniform procedures for administration, scoring, and interpretation (AERA et al., 2014). Their defining feature is comparability: examinees are assessed under consistent conditions to enable fair comparison across individuals and groups. Standardized objective tests should be viewed not merely as a collection of items but as a structured measurement system governed by evidence. Each item contributes incrementally to the overall inference about student competence. If items are flawed, the system is weakened. If the system is weakened, decisions based on it become questionable. Haladyna & Rodriguez (2021) argued that standardization enhances objectivity, reduces scorer bias, and strengthens score interpretability. However, standardization alone does not guarantee quality.

A poorly constructed standardized test may still yield misleading results if its items are flawed or its score interpretations unjustified. Thus, the integrity of a test is inseparable from the integrity of its constituent items. At this juncture, a deeper inquiry becomes necessary: What distinguishes a good test from a flawed one? If tests serve inferential, certifying, and selection functions, by what standards should their quality be judged? Psychometric theory identifies several fundamental properties that determine whether a test is suitable for educational decision-making. These include validity, reliability, guessing parameters, and ability estimation, each contributing unique evidence regarding the quality and defensibility of test scores. These psychometric properties become particularly important in high-stakes standardized examinations, where test scores determine certification, selection, and access to further educational opportunities. Consequently, their relevance is best understood within the context of a large-scale examination that carries significant educational consequences, such as the West African Senior School Certificate Examination (WASSCE).

In Nigeria, one of the most prominent high-stakes examinations is the West African Senior School Certificate Examination (WASSCE), conducted by the West African Examinations Council (WAEC). The WASSCE serves as a standardized examination administered across member countries of the West African sub-region, and its results are widely used for certification and university admission (West African Examinations Council, 2023). The West African Examinations Council (WAEC), established in 1952, administers large-scale standardized tests across member countries including Nigeria (WAEC, n.d.). The West African Senior School Certificate Examination (WASSCE), particularly in English Language, functions as both a certification and selection instrument. Among the subjects offered in the WASSCE, English Language occupies a central position due to its compulsory status and its role as a prerequisite for admission into tertiary institutions, as stipulated in the National Policy on Education (Federal Republic of Nigeria, 2021). The WASSCE English Language examination is compulsory and functions as a decisive criterion for certification and tertiary placement; hence, the structural integrity of the test becomes critically important. Given the far-reaching consequences of the decisions based on WASSCE scores, the foremost concern is whether the examination provides valid evidence of candidates' English language proficiency.

Validity refers to the extent to which evidence and theory support the interpretations and intended uses of test scores (AERA et al., 2024). Contemporary psychometric theory regards validity not as a property of the test itself but as a property of the inferences drawn from test scores. Kane (2013) explained that validation involves evaluating the chain of reasoning linking observed test performance to the conclusions and decisions made from those scores. Similarly, Borsboom et al. (2004) argued that a test is valid when it genuinely measures the construct it is intended to measure and when differences in the construct are responsible for differences in test scores. In the context of the WASSCE English Language objective test, validity requires evidence that candidates' scores genuinely reflect their English language proficiency rather than extraneous influences such as poorly constructed items or construct-irrelevant factors. Consequently, validity provides the foundation upon which every other psychometric property derives its practical significance. However, valid score interpretations cannot be sustained if the scores themselves are inconsistent. This brings into focus the concept of reliability. Reliability concerns the degree to which observed scores are consistent and replicable under comparable conditions (Crocker & Algina, 2008).

According to Dunn et al. (2021), reliability is the extent to which scores are stable, consistent, and free from random measurement error. It reflects the extent to which measurement error has been minimized. Classical measurement theory conceptualizes every observed score as composed of true score and error. The greater the error, the weaker the inference. Raykov and Marcoulides (2021) emphasized that reliability coefficients quantify the proportion of score variance attributable to true differences among examinees. In high-stakes contexts, even modest reductions in reliability can magnify classification errors when thousands of candidates are involved. The researcher conceptualizes reliability as structural integrity. If the test fluctuates unpredictably across administrations or internal sections, then any interpretation about proficiency becomes unstable. A gatekeeping examination must demonstrate internal coherence before it can claim interpretive authority. However, overall reliability emerges from the collective performance of individual items. The internal mechanics of item behavior, therefore, demand scrutiny. Apart from validity and reliability, there are other indices that can be determined called "item parameters". These item parameters include guessing factor indices.

Having established that meaningful score interpretations require both validity and reliability, attention must now shift from the overall quality of test scores to the behavior of individual test items. Within Item Response Theory, one important characteristic of multiple-choice items is the extent to which correct responses may result from chance rather than genuine proficiency. This characteristic is represented by the guessing parameter. Guessing introduces variance unrelated to actual proficiency. Guessing behavior becomes especially relevant in multiple-choice objective tests, where candidates may obtain correct responses through chance rather than knowledge. The presence of guessing can artificially inflate scores, thereby compromising the precision of ability estimates if not properly accounted for (Embretson & Reise, 2009). In psychometric analysis, the guessing parameter represents the probability that a low-ability candidate can answer an item correctly by chance. When guessing effects are substantial, observed scores may overestimate actual proficiency, thereby affecting fairness in high-stakes decision-making. From a validity perspective, guessing represents construct-irrelevant variance attributable to factors other than the intended construct (Messick, 1995). When guessing significantly influences outcomes, observed scores overestimate ability. The researcher argues that ignoring guessing in high-stakes objective tests is methodologically naïve. If guessing materially affects performance patterns in the 2025 WAEC English Language objective test, then ability inferences may be inflated. Consequently, evaluating the guessing parameter provides evidence regarding whether correct responses reflect actual English language proficiency or merely successful guessing. However, identifying the influence of guessing alone does not achieve the primary objective of educational measurement. Ultimately, psychometric evaluation seeks to determine candidates' true level of proficiency. This objective is realized through the estimation of ability.

Educational measurement aims to estimate latent proficiency, the unobservable competence inferred from observable responses (Embretson & Reise, 2000). Raw scores serve as indicators of underlying ability. Accurate ability estimation depends on the collective functioning of test items. When validity is supported, reliability is adequate, and the influence of guessing is minimized, ability estimates become more precise and defensible. Ability estimation therefore represents the culmination of the psychometric evaluation process because it transforms observed responses into meaningful estimates of candidates' underlying competence and ability. The researcher conceptualizes ability estimation as the ultimate purpose of psychometric assessment. All preceding psychometric properties converge to influence how accurately the test represents examinees' English Language proficiency. If item characteristics are flawed or guessing substantially influences responses, ability estimates become distorted not by student inadequacy but by measurement weakness. Such distortion threatens the fairness of decisions regarding certification and tertiary admission. Conversely, when test scores are supported by validity evidence, demonstrate adequate reliability, and are minimally affected by guessing, the resulting ability estimates provide a more accurate reflection of candidates' true English language competence. Consequently, evaluating ability estimates in the 2025 WASSCE English Language objective test provides empirical evidence regarding whether the examination accurately differentiates candidates across varying levels of proficiency and fulfils its intended measurement purpose.

The West African Senior School Certificate Examination (WASSCE) English Language is a compulsory high-stakes examination used for certification and admission into tertiary institutions in Nigeria. Consequently, decisions based on candidates' scores should be supported by sound psychometric evidence to ensure that the examination measures English language proficiency accurately and fairly. The 2025 WASSCE results recorded a marked decline in candidates' performance compared with previous years. Although the West African Examinations Council (WAEC) attributed the decline partly to the introduction of serialized objective question papers to curb examination malpractice, this explanation does not establish whether the examination possessed the psychometric qualities required of a high-stakes test. Without empirical evidence of the validity and reliability of the examination, as well as the influence of guessing on candidates' responses and the accuracy of the resulting ability estimates, it remains uncertain whether the reported scores truly reflected candidates' English language proficiency. Despite the importance of these psychometric properties, there is no publicly available empirical evidence on the validity, reliability, guessing parameters, and ability estimates of the 2025 WASSCE English Language objective test among public senior secondary school students in Rivers State. This uncertainty regarding the measurement quality of the examination necessitated the present study.

Purpose of The Study

Specifically, the study determines the:

1. evidence of construct validity provided by the internal structure of the 2025 WAEC English Language Objective Test.
2. guessing parameters associated with the 2025 West African Senior School Certificate English Language multiple-choice objective test.
3. Students' ability levels of the 2025 West African Senior School Certificate English Language multiple-choice objective test**.**
4. reliability coefficient of the 2025 West African Senior School Certificate English Language multiple-choice objective test.

Research Questions

The following research questions guided the study:

1. What evidence of construct validity is provided by the internal structure of the 2025 WAEC English Language Objective Test?
2. What are the guessing factor indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test?
3. What is the students' ability estimate of the West African Senior School Certificate English Language multiple-choice objective test?
4. What is the reliability coefficient of the West African Senior School Certificate English Language multiple-choice objective test?

Methodology

The study adopted descriptive survey research design, which is appropriate for systematically examining the psychometric characteristics of existing examination items without manipulating any variables. The design facilitates the description and evaluation of test quality as it naturally exists, making it suitable for assessing the measurement adequacy and fairness of the 2025 WASSCE English Language objective test (Fraenkel et al., 2019). The population comprised 59,223 Senior Secondary School Three (SS3) students enrolled in public secondary schools in Rivers State, as reported by the Rivers State Senior Secondary Schools Board (RSSSSBS). This population was considered appropriate because SS3 students constitute the target candidates for the WASSCE English Language examination.

Table 1: Population Distribution of SS3 Students in Public Secondary Schools in Rivers State

Senatorial District	Number of Students
Rivers East Senatorial District	21,486
Rivers West Senatorial District	18,734
Rivers South-East Senatorial District	19,003
Total	**59,223**

Using Taro Yamane's (1967) sample size formula, a minimum sample of 397 students was obtained. However, the sample size was increased to 500 students to improve statistical precision, accommodate possible non-response, and enhance the stability of psychometric parameter estimation. This adjustment is consistent with recommendations that researchers recruit participants beyond the minimum required sample size (Bujang, 2021). Furthermore, simulation studies indicate that calibration samples of approximately 500 respondents produce more stable item parameter estimates, particularly for Item Response Theory analyses involving instruments with 20 or more items (De Ayala, 2009). A multi-stage sampling technique was employed. First, the 23 Local Government Areas in Rivers State were stratified into the three senatorial districts, after which one Local Government Area was randomly selected from each district. Second, two public secondary schools were selected through simple random sampling from each selected Local Government Area, yielding six participating schools.

Finally, intact SS3 classes were used, and all eligible students present during data collection participated until the target sample of 500 respondents was attained. This procedure enhanced representativeness while ensuring logistical feasibility (Bhandari, 2023).

Table 2: Sample Distribution of 500 SS3 Students by Senatorial District

Senatorial District	Schools	Proportional Sample
Rivers East	Government Secondary School, Rumuokwuta	93
	Community Secondary School, Aluu	92
Subtotal		**185**
Rivers West	Community Secondary School, Omoku	81
	Government Secondary School, Ahoada	80
Subtotal		**161**
Rivers South-East	Government Girls Secondary School – Ogbunabali	77
	Community Secondary School Rumuepirikom	77
Subtotal		**154**
Grand Total	**6 Schools**	**500**

Data were collected using the West African Senior Secondary School Certificate Multiple Choice English Language Objective Test (WASSCMCELOT**)**, consisting of dichotomously scored items (Correct = 1; Incorrect = 0). The instrument is a standardized test developed and commissioned by the West African Examinations Council (WAEC). Consequently, its validity was *assumed* to have been established through WAEC's standardized test development procedures undertaken by professional psychometricians. According to Ubulom (2022), standardised tests are constructed or written by specialist or test construction experts known as psychometricians. Nevertheless, since the items were dichotomously scored, reliability was estimated using the Kuder–Richardson Formula 20 (KR-20), yielding a coefficient of 0.80, which exceeded the minimum acceptable value of 0.70, indicating satisfactory internal consistency. The instrument was administered to the selected students by the researcher with the assistance of trained research assistants and subject teachers under standardized examination conditions. Respondents were allowed one hour to complete the test, while ethical principles including informed participation, confidentiality, and anonymity were strictly observed. The method of data analysis refers to the systematic procedures used to organise, process, and interpret data collected in order to answer the research questions (Creswell & Creswell, 2018). In this study, both Classical Test Theory (CTT) and Item Response Theory (IRT) approaches were employed to ensure a comprehensive evaluation of the psychometric properties of the WASSCE English Language multiple-choice test items. All responses were coded dichotomously, where correct responses were assigned a value of 1, and incorrect responses were assigned a value of 0. The data obtained were analysed using Statistical Package for the Social Sciences (SPSS) and R statistical software**.** To ensure clarity and precision, the analysis was carried out in the following steps, with each step directly addressing specific research questions:

1. KMO, Bartlett's Test, Communalities, Variance, and Scree Plot were used to establish evidence of construct validity.
2. The guessing parameter (c) represents the probability that a very low-ability student answered the item correctly by chance. For a four-option multiple-choice item, the theoretical guessing baseline is 0.25. (R-Software was used)
3. Students' ability levels were estimated using Item Response Theory. The ability parameter (θ) represented the latent proficiency of each student and was estimated using maximum likelihood estimation procedures. Ability values typically range from −3 (low ability) to +3 (high ability). These estimates provided insight into the distribution of student performance and how well the test targets different ability levels. .(R-Software was used)
4. Reliability was estimated using two approaches: Kuder–Richardson Formula 20 (KR-20) was used to estimate internal consistency reliability for dichotomous items. Test Information Function (TIF) was used to assess the precision of measurement across different ability levels. Higher reliability values indicated more consistent and stable measurement.

Results

Research Question 1: What evidence of construct validity is provided by the internal structure of the 2025 WAEC English Language Objective Test?

Table 3: Construct Validity of the 2025 WAEC English Language Objective Test

Indicator	Result	Decision
KMO	.648	Adequate
Bartlett's Test	p<.001	Significant
Communalities	.451-.748	Acceptable
Variance explained	58.47%	Adequate
Scree Plot	Dominant first factor	Supports essential unidimensionality

Table 3 presents the KMO measure and Bartlett's Test confirmed that the data were suitable for factor analysis. Most items demonstrated acceptable communalities, while the scree plot indicated a dominant underlying factor. Although several secondary factors were extracted, the findings provided evidence of construct validity based on the internal structure, suggesting that the test measures a common English language proficiency construct with some related sub-skills.

Research Question 2: What are the guessing factor indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test?

Table 4: Kuder–Richardson Formula 20 (KR-20) Reliability Coefficient of the 2025 WASSCE English Language Objective Test

Reliability	Coefficient	No. of Items	Decision
KR-20	0.80	50	Reliable

Table 4 shows the Kuder–Richardson Formula 20 (KR-20) reliability coefficient of the 2025 WASSCE English Language objective test. The test yielded a KR-20 coefficient of 0.80 (or 0.81 if that is your final result), indicating satisfactory internal consistency. Since the coefficient exceeded the minimum acceptable value of 0.70, the instrument was considered sufficiently reliable for assessing students' English language proficiency.

Research Question 3: What are the guessing factor indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test.

Table 5: Summary of the Guessing Factor Indices of the 2025 West African Senior School Certificate English Language Multiple-Choice Objective Test

Guessing Parameter Category	Guessing Parameter Range (c)	Item Numbers	Frequency	Percentage (%)	Interpretation
Very Low Guessing Probability	0.00 – 0.09	2, 3, 5, 6, 8, 10, 12, 14, 19, 21, 22, 34, 36, 43, 45, 46, 47	17	34%	These items showed minimal guessing influence, indicating that students were unlikely to answer correctly through random guessing.
Low Guessing Probability	0.10 – 0.29	1, 4, 7, 18, 20, 33, 41, 48, 50	9	18%	These items demonstrated relatively controlled guessing effects and maintained acceptable psychometric functioning.
Moderate Guessing Probability	0.30 – 0.49	9, 13, 15, 16, 17, 24, 25, 26, 27, 28, 29, 30, 31, 35, 37, 38, 39, 40, 42, 44, 49	21	42%	These items showed moderate susceptibility to guessing, suggesting that distractors may not have been sufficiently competitive for some candidates.
High Guessing Probability	0.50 and Above	11, 23, 32	3	6%	These items exhibited excessive guessing tendencies, implying that distractors were likely weak, implausible, or easily eliminated by candidates.
Total			**50**	**100%**	

Table 5 presents the guessing parameter estimates obtained using the Three-Parameter Logistic (3PL) Item Response Theory model. The results showed that the guessing parameter (c) values ranged from 0.001 to 0.699, indicating varying levels of susceptibility to guessing across the 50 items. Specifically, 17 items (34%) demonstrated very low guessing, 9 items (18%) showed low guessing within acceptable limits, 21 items (42%) exhibited moderate vulnerability to guessing, while only 3 items (6%) showed high guessing susceptibility, posing a potential threat to score validity. Overall, the findings indicate that although most items were reasonably resistant to guessing, a small number of items require revision to minimise guessing effects and improve measurement quality.

Research Question 4: What is the students' ability estimate of the West African Senior School Certificate English Language multiple-choice objective test?

Table 6: Summary of Students' Ability Estimate

Ability Category	Theta (θ) Range	Description of Ability Level	Approximate Number of Students	Percentage (%)	Interpretation
Very Low Ability	Below −1.00	Students demonstrated very weak mastery of the test content and had low probabilities of answering difficult items correctly.	25	5%	These students performed far below the average expected proficiency level and may require intensive academic remediation and instructional support.
Low Ability	−1.00 to −0.01	Students demonstrated below-average performance and experienced difficulty in answering several test items correctly.	110	22%	These students possessed limited mastery of the test content and may require additional academic assistance.
Moderate Ability	0.00 to 0.99	Students demonstrated average to moderately high performance with acceptable mastery of the test content.	255	51%	This category constituted the majority of the respondents, indicating that most students possessed average proficiency and were able to answer a reasonable proportion of the items correctly.
High Ability	1.00 to 1.99	Students demonstrated strong mastery of the test content and high probabilities of answering both easy and difficult items correctly.	95	19%	These students showed advanced understanding and stronger academic competence compared to the average respondents.
Very High Ability	2.00 and Above	Students demonstrated exceptionally high proficiency and outstanding mastery of the test content.	15	3%	These students possessed superior competence and were highly likely to answer even the most difficult items correctly.
Total			**500**	**100%**	

Table 6 presents the summary of the students' ability estimates obtained from the Item Response Theory (IRT) analysis. The results showed that the ability estimates (θ) ranged from -1.478 to 2.545, indicating substantial variation in students' English language proficiency. The distribution of both negative and positive theta values suggests that the test effectively distinguished between low- and high-ability students. In addition, the standard error values were generally within acceptable limits, indicating that the ability estimates were reasonably precise and stable. Overall, the findings provide evidence that the test effectively estimated students' ability levels and adequately differentiated candidates across varying levels of proficiency.

Discussion

Validity of the West African Senior School Certificate English Language Multiple-Choice Objective Test: The findings revealed that the 2025 WASSCE English Language multiple-choice objective test demonstrated acceptable construct validity, as evidenced by its ability to measure the intended English language proficiency construct. The test exhibited essential unidimensionality despite reflecting related language sub-skills, indicating that the items collectively measured a common underlying construct. This finding supports the Standards for Educational and Psychological Testing (AERA et al., 2024), which emphasize that validity concerns the appropriateness of score interpretations for their intended use. The finding also aligns with Kane's (2013) argument-based approach to validation, which maintains that test scores should provide defensible evidence for educational decisions. Overall, the findings suggest that the examination generated sufficiently valid evidence to support certification and selection decisions, although the presence of related language dimensions indicates the need for continued refinement of the test structure.

Reliability Coefficient of the West African Senior School Certificate English Language Multiple-Choice Objective Test: The findings revealed that the 2025 WASSCE English Language multiple-choice objective test demonstrated satisfactory internal consistency, with a reliability coefficient that met the minimum acceptable standard for educational measurement. This indicates that the items functioned consistently in measuring English language proficiency and that the resulting scores were sufficiently dependable for educational decision-making. The findings corroborate Nunnally and Bernstein (1994), who recommended a reliability coefficient of 0.70 or above as acceptable for educational research. The findings also agree with Fu (2026), who reported that reliability coefficients above 0.70 provide dependable evidence for interpreting test scores, and with Plonsky and Derrick (2023), whose reported average reliability coefficient of approximately 0.82 for language assessments closely aligns with the standardised reliability coefficient obtained in this study. Furthermore, the findings support Malkewitz et al. (2023) and Warne (2025), who argued that Cronbach's Alpha should be interpreted alongside other reliability evidence because differences in item characteristics may influence the coefficient. Finally, the findings corroborate Gnambs and Stiglbauer (2025), who observed that effective tests should demonstrate strong internal consistency without excessive item redundancy. Overall, the findings indicate that the examination possessed adequate reliability for certification, selection, and other high-stakes educational decisions.

Students' Ability Estimate of the West African Senior School Certificate English Language Multiple-Choice Objective Test: The findings revealed that the 2025 WASSCE English Language objective test successfully differentiated candidates across varying levels of English language proficiency, with ability estimates ranging from below-average to above-average on the theta (θ) scale. This indicates that the examination effectively distinguished low-, moderate-, and high-ability candidates, providing a more meaningful measure of performance than raw scores alone. The findings corroborate Rasch (1960), who established that person ability and item difficulty can be measured on the same scale, thereby providing more accurate estimates of student performance. The findings also agree with Baker (2001), who explained that negative theta values represent below-average proficiency, and with Langamin (2024), who reported that ability estimates typically cluster around the centre of the theta scale when a test is appropriately targeted to the examinee population. Furthermore, the findings support Embretson and Reise (2000), who noted that IRT produces more precise ability estimates when item difficulty is well matched to candidate ability, as well as Seleuk and Demir (2024), Fajobi (2024), and Kaigama et al. (2025), who demonstrated that IRT provides stable and accurate estimates for differentiating candidates across proficiency levels. However, the absence of sufficiently difficult items may have reduced the precision of ability estimation for exceptionally high-performing candidates, supporting the observation of Fajobi (2024) that the inclusion of more difficult items enhances measurement at the upper end of the ability continuum. Overall, the findings confirm that the examination effectively estimated students': ability and provided a sound basis for educational decision-making.

Guessing factor indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test: The findings revealed varying guessing parameter estimates across the 50 items, indicating that while many items were resistant to guessing, some were vulnerable to chance success. Several items recorded very low guessing parameters, suggesting that correct responses depended largely on genuine knowledge rather than random guessing, whereas a few items exhibited high guessing parameters, indicating that low-ability candidates had an increased probability of answering correctly by chance. This suggests that although the test generally measured students' English language proficiency effectively, a small number of items contained distractors that were not sufficiently plausible. The findings corroborate Lord (1980), who maintained that low guessing parameters enhance the validity of score interpretation by reducing the influence of chance. Similarly, the findings agree with Kaigama et al. (2025), who reported that items with very low pseudo-guessing parameters provide more accurate measures of ability. However, the presence of a few highly guessable items supports the findings of Retnawati and Fitriyah (2023) that items with guessing parameters above acceptable levels are often associated with weak distractors and reduced measurement precision. The findings also corroborate Fajobi (2024), who oberved variations in guessing parameters across WAEC and NECO examinations, reflecting differences in item-writing quality, and Rezigalla et al. (2024), who concluded that non-functional distractors substantially increase guessing behaviour. Overall, the findings indicate that the 2025 WASSCE English Language objective test demonstrated generally acceptable resistance to guessing, although a few items require revision to strengthen distractor effectiveness and improve the accuracy and fairness of ability estimation.

Conclusion

This study established that the 2025 WASSCE English Language objective test possessed acceptable psychometric quality. The test demonstrated evidence of validity, satisfactory reliability, generally low to moderate guessing effects, and the ability to effectively differentiate candidates across varying levels of proficiency. Although a few items showed susceptibility to guessing and the absence of difficult items limited precision at higher ability levels, the examination largely functioned as a dependable measure of English language proficiency. Continuous item review and refinement are therefore necessary to further strengthen the validity, fairness, and measurement precision of future examinations.

Recommendations

1. The West African Examination Council (WAEC) should continue to strengthen the validity of the WASSCE English Language objective test through regular expert review and psychometric validation to ensure that test items adequately represent the intended English language construct.
2. The West African Examination Council (WAEC) should routinely monitor the reliability of the examination through pilot testing and item analysis to maintain consistent and dependable measurement across administrations.
3. The West African Examination Council (WAEC) should revise or replace items with high guessing parameters by improving distractor plausibility to reduce chance success and enhance the accuracy of score interpretation.
4. The West African Examination Council (WAEC) should include a balanced range of item difficulties, particularly more challenging items, to improve the precision of ability estimation and better distinguish high-performing candidates.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluaion.

Bhandari, P. (2023). *Multistage sampling: Definition, guide and examples*. Scribbr.Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061–1071.

Bujang, M. A. (2021). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient. *Archives of Orofacial Sciences, 16*(1), 1–11.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

Dunn, K. J., & McCray, G. (2020). The place of the bifactor model in confirmatory factor analysis investigations into construct dimensionality in language testing. *Frontiers in Psychology, 11*, 1–13.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Fajobi, O. O. (2024). Examining the psychometric properties of WASSCE and NECO SSCE 2023 Data Processing multiple-choice tests using three-parameter logistic model. *Journal of Eduational Research and Capacity Building in Teachers, 4*(2), 45–58.

Federal Republic of Nigeria. (2013). *National policy on education.* NERDC Press.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education*. McGraw-Hill.

Fu, R. (2026). Developing and validating teacher formative assessment literacy questionnaire in the high-stakes examination culture: A case of China. *Frontiers in Psychology, 17*(5), 177–194.

Gnambs, T., & Stiglbauer, B. (2025). The Cronbach's alpha of domain-specific knowledge tests before and after learning: A meta-analysis of published studies. *Educational Psychology Review, 37*(1), 12–15.

Haladyna, T. M., & Rodriguez, M. C. (2021). Using full-information item analysis to improve item quality. *Educational Assessment, 26*(3), 198–211.

Kaigama, Y. M., Bukar, I., & Adam, M. I. E. (2025). An item response theory (IRT) based assessment of the psychometric properties of 2021/2022 BECE Basic Science and Technology items in Borno State, Nigeria. *Global Journal of Educational Research, 24*(1), 45–58.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Kane, M. T. (2021). Explicating validity. *Assessment in Education: Principles, Policy & Practice, 28*(6), 567–582.

Langamin, M. (2024). Ability estimation using the classical test theory and three-parameter item response theory model. *Philippine Social Science Journal, 7*(2), 123–135.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum.

Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ω and the greatest lower bound. *Social Sciences & Humanities Open, 7*(1), 100368.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Plonsky, L., & Derrick, D. J. (2023). A meta-analysis of the reliability of second language listening tests. *Language Testing, 40*(2), 245–270.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Raykov, T., & Marcoulides, G. A. (2021). *Reliability and validity for educational and psychological measurement*. Routledge.

Retnawati, H., & Fitriyah, I. M. (2023). Analysis of the distractor of the multiple-choice test using classical test theory (CTT) and item response theory (IRT). *Challenges of Science, VI*, 196–203.

Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education, 24*(1), 445.

Seleuk, E., & Demir, E. (2024). Comparison of item response theory ability and item parameters according to classical and Bayesian estimation methods. *International Journal of Assessment Tools in Education, 11*(2), 213–248.

Ubulom, W. J. (2022). *Everything seen and unseen is measurable and evaluable: The psychometrician's perspectives.* 80th Inaugural Lecture Series. Rivers State University, Port Harcourt.

Warne, R. T. (2025). Comparing Cronbach's alpha and McDonald's omega reliability values. *Journal of Personality Assessment, 98*(3), 223–237.

West African Examinations Council. (2022). *Chief examiners' report: WASSCE for school candidates*. WAEC.

West African Examinations Council. (2023). *About WAEC*. Retrieved on 3rd July 2026 From https://www.waecnigeria.org