Psychometric Properties Determination of 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 determined the psychometric properties of the 2025 West African Senior School Certificate Examination (WASSCE) English Language multiple-choice objective test administered to public senior secondary school students in Rivers State. A descriptive evaluation research design was adopted. The population comprised 59,223 SS3 students, while a sample of 500 students was selected using a multi-stage sampling technique. Data were collected using the standardized 2025 WASSCE English Language Objective Test. The validity and reliability of the instrument were assumed to have been established by WAEC experts; however, a pilot study conducted using SPSS produced a Kuder–Richardson Formula 20 (KR-20) reliability coefficient of 0.80. Data were analysed using SPSS. Dimensionality was assessed using Exploratory Factor Analysis, item difficulty using the difficulty index (p-value), item discrimination using the discrimination index (D-index), and distractor effectiveness through the combined interpretation of item difficulty and discrimination statistics. The findings showed that the test demonstrated essential unidimensionality, appropriate item difficulty, acceptable item discrimination, and generally functional distractors, although a few items required improvement. The study concluded that the examination met acceptable psychometric standards but requires periodic item refinement. It was recommended that WAEC should routinely evaluate dimensionality, maintain balanced item difficulty, revise weakly discriminating items, and improve distractor quality.

Keywords: Evaluation, Difficulty, Discrimination, Distractor, Item, Psychometric

Introduction

Education is widely recognized as a fundamental instrument for individual empowerment and national development. Through education, societies transmit knowledge, skills, values, and competencies necessary for social cohesion and economic advancement. At the individual level, education enhances intellectual capacity, critical reasoning, and problem-solving abilities required for meaningful participation in contemporary society, further reiterating that education is a transformative force that builds resilient societies, promotes social cohesion, and drives sustainable economic development in an increasingly complex world (United Nations Educational, Scientific and Cultural Organization, 2015). Although education transcends measurement, it inevitably requires structured evaluation to ensure accountability and credibility (World Bank, 2020; UNESCO, 2015). Such evaluation is operationalized through various assessment methods, of which tests remain the most widely used and objective means of determining whether educational goals have been achieved.

According to Obilor et al (2026), tests served as a measure of academic progress, a gateway to opportunities, and a tool for accountability. e to the educational gatekeeping duties, the researcher takes the position that education can achieved its societal purpose only when learning is authentically cultivated and accurately evaluated. If **e**valuation mechanisms are weak, misaligned, or psychometrically flawed, the entire educational **e**nterprise risks distortion. Students may be certified without competence, or competent students can be unjustly filtered out. In high-stakes environments, this distortion has emotional, economic, d social consequences. Therefore, if tests serve inferential, certifying, and selection functions, by what standards should their quality be judged? The answer lies in the essential properties of a go test that determine whether test scores are dependable, meaningful, and fair, especially in high takes standardized test like 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 standardised examination administered across member countries of the West African sub-region, and its results are widely used for certification and university admission (West African Education Council, 2023). The West African Examinations Council (WAEC), established in 1952, administers large-scale standardised 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. If the examination is intended to measure overall English language proficiency, then its items must collectively reflect a coherent underlying construct and evidence of appropriate dimensionality.

Dimensionality refers to the extent to which a set of test items measures a single dominant underlying construct or multiple constructs. In many standardised achievement tests, it is assumed that items are sufficiently homogeneous to reflect one primary latent trait, such as English language proficiency. When this assumption is violated, parameter estimates and ability interpretations may become distorted, thereby weakening the credibility of score-based decisions (Embretson & Reise, 2009; Demars, 2023). Establishing the dimensional structure of a test is therefore a critical preliminary step in psychometric evaluation, because meaningful estimation of other item characteristics depends on the structural coherence of the test, as evidence of appropriate dimensionality strengthens validity. Apart from dimensionality, there are other psychometric properties the test must meet, such as item difficulty indices. Item difficulty represents the proportion of examinees answering an item correctly (Haladyna & Rodriguez, 2013). Item difficuty refers to the proportion of candidates who answer an item correctly or within modern measurement frameworks, the location of the item along the latent ability continuum (DeMars, 2010). Yet beyond its computational simplicity lies an informational function. Items that are extremely easy or extremely difficult contribute minimal differentiation power and may reduce measurement precision. Embretson and Reise (2000) explained that optimal measurement occurs when item difficulty is aligned with examinee ability levels. When alignment is poor, score interpretations become less precise. Difficulty must be evaluated not only statistically but substantively. If a large proportion of candidates fail a particular English item, one must interrogate whether the difficulty reflects legitimate linguistic complexity or ambiguous item construction. Appropriate distribution of item difficulty ensures that a test neither clusters excessively at very easy nor very difficult levels, thereby allowing for effective discrimination.

Discrimination indices reflect how effectively an item distinguishes high-performing candidates from low-performing ones (Ebel & Frisbie, 1991). Item discrimination describes the extent to which an item distinguishes between candidates of higher and lower ability (Miller et al., 2013). High-discrimination items contribute positively to the accuracy of ability estimation, whereas poorly discriminating items may introduce measurement error and obscure genuine differences in competence. High discrimination strengthens the relationship between item performance and overall test performance. Items that fail to discriminate raise interpretive concerns. They may be mis-keyed, ambiguous, or measuring peripheral skills. As Haladyna and Rodriguez (2013) observed, well-constructed items should align with the construct hierarchy such that students with greater mastery are more likely to respond correctly. The researcher views discrimination as internal alignment.

If the WAEC English objective test claims to measure proficiency, then high-proficiency candidates should consistently outperform lower-proficiency peers at the item level. Absence of discrimination signals measurement incoherence. Yet in selected-response formats, discrimination can be distorted by superficial distractors or by strategic guessing. This necessitates examination of distractor efficiency.

Multiple-choice items rely heavily on the quality of distractors. Plausible distractors attract examinees lacking mastery; implausible ones fail to function (Rodriguez, 2016). Non-functional distractors reduce item complexity and may artificially inflate scores. Gierl et al (2017) highlighted that distractor analysis can reveal cognitive misconceptions and improve item quality. Effective distractors are not random alternatives; they are diagnostically informed options grounded in common learner errors. Distractor efficiency is particularly critical in English Language testing, where subtle grammatical or semantic distinctions must be represented convincingly. Weak distractors convert assessment into recognition rather than reasoning. Collectively, dimensionality, item difficulty, item discrimination, and distractor efficiency constitute essential psychometric indicators of test quality. Evaluating these properties provides empirical evidence on whether a high-stakes examination functions as intended and produces valid, reliable, and defensible scores for educational decision-making.

Despite the importance of these psychometric indicators, irrespective of whether examination performance trends demonstrate improvement or fluctuation, the absence of localized empirical evidence regarding the internal functioning of a specific administration of the WASSCE English Language objective test represents a measurement gap. In a context where examination outcomes determine certification, tertiary placement, and future opportunities, systematic psychometric verification becomes not only a methodological necessity but an ethical imperative. It is this need for empirical substantiation of measurement quality in the 2025 WASSCE English Language objective test among senior secondary school students in Rivers State that necessitated the present study.

Statement of the Problem

The West African Senior School Certificate Examination (WASSCE) English Language is a compulsory high-stakes subject that determines certification and university admission in Nigeria. However, the 2025 WASSCE results revealed a severe performance decline, with only 38.32% of 1,969,313 candidates obtaining credits in five subjects, including English Language, which represents a sharp drop of performance from 72.12% in 2024. This unprecedented decline raises serious concerns about the quality and stability of the examination instruments. However, WAEC attributed the decline partly to the introduction of serialized objective test papers aimed at curbing malpractice. However, this explanation raises a critical psychometric concern: whether changes in the structure of the objective test affected its measurement quality. Despite the magnitude of the performance drop, there is no publicly available item-level psychometric evidence to verify whether the 2025 English Language objective test possessed acceptable difficulty levels, discrimination power, dimensionality, and distractor efficiency.

Existing studies on WAEC examinations have focused mainly on Mathematics and Physics, reporting issues such as inconsistent item difficulty, weak discrimination, and dysfunctional distractors, but no equivalent empirical evaluation exists for the English Language objective test. Consequently, the 2025 WASSCE English Language examination operates without independent verification of its technical quality, leaving its measurement validity largely untested. Therefore, the problem of this study is the combination of a drastic decline in English Language performance and the absence of publicly verifiable psychometric evidence to confirm whether the 2025 WASSCE objective test functioned as a valid and reliable measurement instrument. This creates uncertainty about whether observed failures reflect true student ability or weaknesses in the test itself, necessitating an empirical evaluation of its psychometric properties.

Purpose of the Study

The purpose of this study is to assess the psychometric properties of the 2025 West African Senior School Certificate English Language objective test administered to senior secondary school students in Rivers State. Specifically, the study determines the:

1. dimensionality of the 2025 West African Senior School Certificate English Language multiple-choice objective test.
2. item difficulty indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test.
3. discrimination indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test.

4. effectiveness of distractors in the 2025 West African Senior School Certificate Examination English Language multiple-choice objective test.

Research Questions

The study following research questions guided the study:

1. What is the dimensionality of the 2025 West African Senior School Certificate English Language multiple-choice objective test?
2. What are the item difficulty indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test?
3. What are the discrimination indices of the items in the 2025 West African Senior School Certificate English Language multiple-choice objective test?
4. How effective are the distractors in the 2025 West African Senior School Certificate English Language multiple-choice objective test?

Methodology

This study adopted a descriptive evaluation 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 was considered appropriate because standardized examination items provide an objective basis for psychometric evaluation. 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 (Ubulom, 2022; Kane, 2021). Nevertheless, a pilot study involving respondents outside the study sample was conducted to determine its reliability within the present context. 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. Data were analysed using both Classical Test Theory (CTT**)** and Item Response Theory (IRT) approaches with the aid of Statistical Package for the Social Sciences (SPSS). Responses were coded dichotomously (Correct = 1; Incorrect = 0). Dimensionality was assessed using Exploratory Factor Analysis, with decisions based on eigenvalues greater than one, scree plot examination, and percentage of variance explained. Item difficulty was estimated using the difficulty index (p-value**)** and interpreted as difficult (0.00–0.29), moderate (0.30–0.70), or easy (0.71–1.00). Students were also classified into upper and lower 25% criterion groups to further examine item functioning. Item discrimination was evaluated using the discrimination index (D-index) and point-biserial correlation coefficient, with values of 0.40–1.00 regarded as good, 0.30–0.39 as marginal, and 0.20–0.29 as poor, while negatively discriminating items were considered defective. Distractor effectiveness was evaluated indirectly through the combined interpretation of item difficulty and discrimination statistics derived from the dichotomously scored item responses.

Results

Research Question 1: What is the dimensionality of the 2025 West African Senior School Certificate English Language Multiple-choice Objective Test?

Table 3: Summary of the Dimensionality Analysis of the 2025 WASSCE English Language Objective Test.

Statistical Indicator	Result Obtained	Benchmark/Decision Rule	Decision
KMO Measure of Sampling Adequacy	0.648	≥ 0.50 acceptable	Adequate for factor analysis
Bartlett's Test of Sphericity	$\chi^2 = 1939.741$, df = 1275, $p < .001$	$p < .05$	Significant
Communality Range	0.451 – 0.748	≥ 0.40 acceptable	Adequate communalities
Number of Components Extracted	22	Lower number preferred for unidimensionality	Indicates multidimensionality
Variance Explained by First Factor	7.14%	Higher percentage preferred	Weak dominant factor
Total Variance Explained	58.47%	≥ 50% acceptable	Adequate cumulative variance
Component Loadings	Several cross-loadings observed	Clear loading preferred	Structural inconsistency present
Scree Plot Pattern	Sharp drop after first factor	Dominant first factor expected	Supports essential unidimensionality
Overall Dimensionality Decision	Essential unidimensionality with multidimensional characteristics	One dominant latent trait expected	Partially satisfies unidimensionality assumption

Table 3 presents the dimensionality analysis of the 2025 WASSCE English Language objective test. The KMO value (0.648) and Bartlett's Test of Sphericity ($\chi^2(1275) = 1939.741$, $p < .001$) confirmed that the data were suitable for factor analysis. Most items exhibited acceptable communalities, indicating adequate representation by the extracted factors. Although the scree plot suggested a dominant general factor, the extraction of 22 components, the low variance explained by the first factor (7.14%), and the presence of cross-loading items indicated that the test demonstrated essential unidimensionality with notable multidimensional characteristics, suggesting that the items measured a general English language proficiency alongside several related sub-skills.

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

Table 4: Summary of Item Difficulty Indices of the 2025 WASSCE English Language Objective Test

Difficulty Category	Difficulty Index Range (p-value)	Number of Items	Percentage (%)	Interpretation
Difficult Items	Below 0.30	0	0%	No item was classified as difficult
Moderately Difficult Items	0.30 – 0.70	35	70%	Majority of the items possessed acceptable difficulty levels and were effective for differentiating students based on ability
Easy Items	Above 0.70	15	30%	These items were relatively easy for most students
Total	—	50	100%	—

Table 4 presents the item difficulty indices of the 2025 WASSCE English Language objective test. The p-values ranged from 0.36 to 0.87, with 70% of the items classified as moderately difficult and 30% as easy, while no difficult items were identified. This indicates that the test was generally well-targeted for the average examinee, although the absence of difficult items suggests reduced precision in distinguishing high-ability candidates and a potential ceiling effect.

Research Question 3: What are the discrimination indices of the items in the 2025 West African Senior School Certificate English Language multiple-choice objective test?

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

Discrimination Category	**Discrimination Range (D)**	**Item Numbers**	**Frequency**	**Percentage (%)**	**Interpretation**
Very Good Discrimination	0.40 – 0.52	4, 5, 26, 27, 29, 35, 48	7	14%	These items possessed strong discriminatory power and effectively differentiated high-ability students from low-ability students.
Good Discrimination	0.30 – 0.39	1, 2, 16, 17, 18, 22, 25, 31, 36, 38, 39, 40, 42, 49, 50	15	30%	These items demonstrated acceptable discriminatory power and adequately distinguished academically stronger students from weaker students.
Marginal Discrimination	0.20 – 0.29	3, 6, 8, 9, 10, 11, 13, 14, 15, 20, 21, 23, 28, 30, 32, 37, 41, 43, 44, 45, 46, 47	22	44%	These items showed only moderate ability to discriminate between students and may require revision or improvement.

Poor Discrimination	0.00 – 0.19	7, 12, 19, 24, 33, 34	6	12%	These items failed to adequately distinguish between high-performing and low-performing students and are considered weak items psychometrically.
Negative Discrimination	Below 0.00	None	0	0%	No item exhibited negative discrimination, indicating that none of the items functioned in a misleading direction.
Total			**50**	**100%**	

Table 5 presents the item discrimination indices of the 2025 WASSCE English Language objective test. The results showed that the majority of the items demonstrated acceptable to good discrimination, with 22 items exhibiting good discriminatory power ($D \geq 0.30$), 21 items showing marginal discrimination ($D = 0.20–0.29$), and 7 items displaying poor discrimination ($D < 0.20$). Notably, no item had a negative discrimination index, indicating that all items functioned in the intended direction. Overall, the test demonstrated moderate to good psychometric quality, although items with poor discrimination require revision to improve the test's measurement precision and reliability.

Research Question 4: How effective are the distractors in the 2025 West African Senior School Certificate English Language multiple-choice objective test?

Table 6: Summary of Indirect Distractor Effectiveness Analysis of the 2025 WASSCE English Language Objective Test

Distractor Effectiveness Category	Psychometric Characteristics	Item Numbers	Frequency	Percentage (%)	Interpretation
Highly Effective Distractors	Moderate item difficulty with strong discrimination indices	6, 7, 15, 27, 32, 33, 35, 43, 47	9	18%	Distractors were highly plausible and successfully attracted weaker students while allowing stronger students to identify the correct answers.
Moderately Effective Distractors	Acceptable item difficulty with moderate or marginal discrimination indices	13, 18, 25, 30, 39, 40	6	12%	Distractors functioned reasonably well but may require minor improvement to enhance their competitiveness and psychometric precision.

Ineffective Distractors	Excessively easy items with weak discrimination indices	14, 19, 21, 22, 23, 44	6	12%	Distractors appeared implausible or easily eliminated, thereby reducing the ability of the items to distinguish between students of varying ability levels.
Functional Distractors	Moderate difficulty and acceptable discrimination indices	1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 16, 17, 20, 24, 26, 28, 29, 31, 34, 36, 37, 38, 41, 42, 45, 46, 48, 49, 50	29	58%	Distractors generally functioned adequately and contributed positively to the psychometric quality of the examination.
Total			**50**	**100%**	

Table 6 presents the distractor effectiveness of the 2025 WASSCE English Language objective test, assessed indirectly using item difficulty and discrimination indices. The results showed that most items had functional distractors, indicating that the incorrect options effectively distinguished between high- and low-ability examinees. However, a few items exhibited suboptimal or poor distractor functioning, characterized by high facility and weak discrimination, suggesting implausible distractors that require revision. Overall, the test demonstrated generally effective distractor performance, with only a limited number of items needing improvement to enhance measurement precision and validity.

Discussion

The findings of the study are presented under the following headings:

Dimensionality of the 2025 West African Senior School Certificate English Language multiple-choice objective test.

The findings revealed that the 2025 WASSCE English Language objective test exhibited essential unidimensionality, with a dominant general English language proficiency factor alongside smaller dimensions relating to vocabulary, grammar, comprehension, and reasoning. This finding supports Stout's (1987) concept of essential unidimensionality and is further strengthened by the KMO and Bartlett's Test results, acceptable communalities, and the scree plot, despite the extraction of 22 components. Consistent with Golino et al. (2020) and Goretzko & Bühner (2020), the large number of extracted components is interpreted as statistical over-extraction rather than evidence of multiple independent constructs. The findings also corroborate Liu et al. (2022) and Dunn & McCray (2020) that language tests naturally exhibit some multidimensionality because they assess interconnected language skills, while the communalities support the recommendations of Tavakol and Wetzel (2020) that the items were adequately represented within the factor structure. Overall, the results indicate that the test provides a valid measure of overall English language proficiency despite reflecting the inherent complexity of language assessment.

Item difficulty indices of the 2025 West African Senior School Certificate English Language multiple-choice objective test.

The findings revealed that the 2025 WASSCE English Language objective test was dominated by moderately difficult items (70%), with the remaining 30% classified as easy, while no difficult items were identified. This indicates that the test was well targeted at measuring average language proficiency and was capable of effectively distinguishing between candidates of different ability levels. The findings corroborate Kim et al. (2024), Anwer et al. (2024), Cotner & Wassenberg (2024), and Schneider et al. (2020), who reported that predominantly moderate item difficulty enhances reliability, discrimination, and accessibility. However, the absence of difficult items supports the concerns of Evidence-based Education (2025), Hujar (2022), and Haladyna & Rodriguez (2021) that tests lacking challenging items may experience ceiling effects and have limited ability to differentiate high-performing candidates. Overall, the examination demonstrated good psychometric quality but would benefit from the inclusion of a modest number of difficult items to improve measurement precision at the upper end of the ability continuum.

Discrimination Indices of the Items in the 2025 West African Senior School Certificate Examination (WASSCE) English Language Multiple-Choice Objective Test

The findings revealed that the 2025 WASSCE English Language objective test was dominated by moderately difficult items (70%), with the remaining 30% classified as easy, while no difficult items were identified. This indicates that the test was well targeted at measuring average language proficiency and was capable of effectively distinguishing between candidates of different ability levels. The findings corroborate Kim et al. (2024), Anwer et al. (2024), Cotner and Wassenberg (2024), and Schneider et al. (2020), who reported that predominantly moderate item difficulty enhances reliability, discrimination, and accessibility. However, the absence of difficult items supports the concerns of Evidence Based Education (2025), Hujar (2022), and Haladyna and Rodriguez (2021) that tests lacking challenging items may experience ceiling effects and have limited ability to differentiate high-performing candidates. Overall, the examination demonstrated good psychometric quality but would benefit from the inclusion of a modest number of difficult items to improve measurement precision at the upper end of the ability continuum.

Effectiveness of Distractors in the 2025 WASSCE English Language Multiple-Choice Objective Test

The findings revealed that the distractors in the 2025 WASSCE English Language objective test were generally functional and contributed positively to the psychometric quality of the examination, as evidenced by the predominance of items with moderate difficulty and positive discrimination. These findings corroborate Haladyna and Rodriguez (2021) and Rezigalla et al. (2024), who emphasized that plausible distractors improve item discrimination and reduce random guessing. However, some overly easy and weakly discriminating items suggested the presence of implausible or ambiguous distractors, consistent with the findings of Al-Hadi et al. (2023) and Ansari (2022). Although the indirect method used could not identify specific non-functional distractors, it remained appropriate for evaluating distractor performance when option-level data were unavailable, as noted by Haladyna and Rodriguez (2021). Overall, the results indicate that while the distractors were largely effective, routine distractor review and improved item-writer training, as recommended by Garba et al. (2024) and Rezigalla et al. (2024), would further strengthen the validity and quality of future examinations.

Conclusion

This study evaluated the psychometric properties of the 2025 WASSCE English Language multiple-choice objective test using both Classical Test Theory (CTT) and Item Response Theory (IRT). The findings showed that the examination met acceptable psychometric standards for large-scale educational assessment. The test demonstrated appropriate item difficulty, with most items falling within the desirable moderate difficulty range, effective item discrimination, and generally functional distractors, indicating that it was suitable for assessing English language proficiency. However, the complete absence of difficult items suggests that the examination may not adequately challenge or differentiate candidates at the highest ability levels. In addition, a few items exhibited weak discrimination and less effective distractors, indicating the need for item refinement. Overall, although the examination satisfied the essential psychometric requirements, it was not entirely free from measurement limitations. Consequently, periodic item review and revision, particularly the inclusion of appropriately difficult items and improvement of weaker items and distractors, are recommended to further enhance the validity, fairness, precision, and overall quality of future WASSCE English Language examinations.

Recommendations

Based on the findings of the study, it is recommended that:

1. The West African Examinations Council (WAEC) should routinely evaluate the dimensionality of WASSCE English Language objective tests to ensure that items measure the intended language proficiency construct.
2. The West African Examinations Council (WAEC) should maintain a balanced distribution of easy, moderate, and difficult items to improve measurement across all ability levels.
3. The West African Examinations Council (WAEC) should revise or replace items with weak discrimination indices to enhance the test's ability to distinguish between high- and low-performing candidates.
4. The West African Examinations Council (WAEC) should improve distractor quality through regular item review and pilot testing to minimise guessing and strengthen the validity of the examination.

References

Anwer, L. A., Alghamdi, A. A., Alkattan, A. K., Aljadoa, O. M., Alkattan, K. A., Alghamdi, F. A., & Alotaibi, N. M. (2024). *Effective use of item analysis to improve the reliability and validity of undergraduate medical examinations: Evaluating the same exam over many years*. BMC Medical Education, 24(3), 13–17.

Cotner, S., & Wassenberg, D. (2024). *Question format is the best predictor of item discrimination*. BMC Medical Education, 24(2), 8–24.

Demars, C. E. (2023). *Item analysis for the modern educational assessment: Updated frameworks for 2020*. Journal of Educational Measurement, 60(1), 25–48.

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(7), 13–57.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement*. Prentice Hall.
Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Evidence-Based Education. (2025). *When tests hit the ceiling: Why ceiling effects matter*. Journal of Assessment Academy Resources, 14(2), 45–58.

Federal Republic of Nigeria. (2013). *National Policy on Education* (6th ed.). NERDC Press.

Golino, H. F., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). *Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial*. Psychological Methods, 25(3), 292–320.

Goretzko, D., & Bühner, M. (2020). *One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis*. Psychological Methods, 25(6), 776–786.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating multiple-choice test items*. Routledge.

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

Hujar, J. (2022). *Combatting ceiling effects: Modeling high-ability student growth* (Doctoral dissertation). University of North Carolina at Charlotte.

Kim, G., Yim, M., Bae, H., & Hur, J.-W. (2024). *Factor structures and psychometric properties of brief emotion regulation scales*. BMC Psychology, 12(7), 759–812.

Kim, S., Park, J., & Lee, S. (2024). *Item difficulty index, discrimination index, and reliability of the 26 health professions licensing examinations in 2024, Korea: A psychometric study*. Journal of Educational Evaluation for Health Professions, 21(2), 40–45.

Liu, Y., Liang, X., & Zhang, Y. (2022). *A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability*. Studies in Educational Evaluation, 75(2), 101–114.

Obilor E.I., Orinemi D.H., Goodluck U.M (2026) Test scores as indicators of academic excellence in Nigeria. *International Journal of Spectrum Research in Education & Humanities, 1*(3) 85-93

Rodriguez, M. C. (2016). *Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research*. Educational Measurement: Issues and Practice, 35(2), 5–13.

Schneider, M., Preckel, F., & Retelsdorf, J. (2020). *Person misfit, test anxiety, and test-taking motivation in a large-scale mathematics proficiency test for self-evaluation*. Studies in Educational Evaluation, 67(2), 100–918.

Stout, W. F. (1987). *A nonparametric approach for assessing latent trait dimensionality*. Psychometrika, 52(4), 589–617.

Tavakol, M., & Wetzel, A. P. (2020). *Factor analysis: A means to verify theoretical constructs*. International Journal of Research in Medical Sciences & Technology, 6(1), 1–8.

UNESCO. (2015). *Rethinking education: Towards a global common good?* UNESCO Publishing.

UNESCO. (2021). *PISA 2018 results (Volume VI): Are students ready to thrive in an interconnected world?* OECD Publishing.

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

West African Examinations Council. (2023). *About WAEC*. Retrieved February 23, 2026, from https://www.waecnigeria.org

World Bank. (2020). *Realizing the future of learning: From learning poverty to learning for everyone, everywhere.* World Bank.