

Analytical Comparison of Item Statistics Employing IRT 3PLM in WASSCE and NECO SSCE 2023 English Language Multiple Choice Tests

Rioborue Alexander Oghenerume^{1*}, Stella Inwuototso Uyi-Osaretin²

¹Educational Evaluation and Counselling Psychology, University of Benin, Benin City, Nigeria, 0009-0008-7808-4810

²Educational Foundations, University of Benin, Benin City, Nigeria

*Corresponding author: raoghenerume@gmail.com

Abstract

This assessment encompassed the analytical comparison of item statistics employing the Item Response Theory Three Parameter Logistic Model (IRT 3PLM) in WASSCE and NECO SSCE 2023 English Language multiple choice tests. Three hypothesised questions were articulated. In this study, a survey research design was utilised. This investigation's population comprised 72,400 English Language SS 3 students in 384 public senior secondary schools in Ogun State. A multistage sampling procedure was applied to pick a sample size of 8,000 English Language SS 3 students. The assessment instruments used were WASSCE and NECO SSCE 2023 English Language multiple choice tests. The data collected were subjected to IRT 3PLM using jMetrik psychometric software to generate the item parameters. All hypothesised questions were tested for significance utilising independent samples t-test statistics. This assessment's outcomes divulged a significant dissimilarity in the guessing parameter between WASSCE and NECO SSCE 2023 English Language multiple-choice items. However, no significant dissimilarities existed in the discrimination and difficulty parameters between the two. The pragmatic extrapolation of this analytical comparison accentuated in the findings alluded that employing IRT 3PLM to establish the item statistics before administering the tests could further enhance the quality of English Language multiple choice items developed by WAEC and NECO.

Keywords: English Language, Item Statistics, IRT 3PLM, WASSCE, NECO SSCE

1. Introduction

The English Language is a globally influential means of communication that has evolved over centuries, originating in England but now spoken and understood by millions worldwide, which serves as a bridge between people of different cultures, enabling them to exchange ideas, conduct business, and engage in various forms of social interaction (Crystal, 2003). English, with its diverse vocabulary, flexible grammar, and global usage across fields like science, technology, diplomacy, and entertainment, embodies significant cultural, historical, and social importance, reflecting the

dynamic nature of global communication and interaction (Rao, 2019; Crystal, 2003).

In Nigeria, English Language is a compulsory subject at all levels of education from primary school through secondary school (Foyewa, 2020; Federal Republic of Nigeria, 2013). To discern the magnitude to which the students have understood a wide range of English Language topics within the curricula, a test is usually conducted to ensure that their grammar, vocabulary, writing, and reading aptitudes align with curricula standards. The test undergoes item statistics computation to assess

the quality of each item and identify any that may need revision or elimination. Oghenerume (2022) expounded that item statistics are quantitative measures that capture the numerical properties and characteristics of test items. These quantitative measures encompass various aspects of test items, including their parameters. For an item to be considered a good item, it should satisfy the three parameters: difficulty, discrimination, and guessing. Difficulty is how challenging an item is for test-takers, discrimination indicates an item's adeptness to split amid test-takers with differing points of the construct being assessed, and guessing quantifies the possibility that a test-taker with minimal adeptness would reply to an item correctly purely by chance (Oghenerume, 2022). English Language tests usually consist of both objective and subjective items that assess the students' knowledge and understanding of the subject.

The objective items are usually in the form of multiple-choice items that test the students' proficiency in grammar, vocabulary, and reading understanding. The subjective items often require students to exhibit their aptitude to articulate ideas clearly and meritoriously through writing tasks. The two major bodies responsible for conducting and grading English Language tests for Senior Secondary 3 (SS 3) students in Nigeria are the West African Examinations Council (WAEC) and the National Examinations Council (NECO). Both WAEC and NECO are in charge of Senior Secondary Certificate Examinations (SSCE). While WAEC conducts the West African Senior School Certificate Examination (WASSCE) for students in Nigeria, Gambia, Ghana, Sierra Leone, and Liberia, NECO is responsible for the students in Nigeria writing the National Examinations Council Senior School Certificate Examination (NECO SSCE). The NECO SSCE is similar to the WASSCE, but it is primarily taken by students in Nigeria.

2. Literature Review

The theoretical framework for this study is hinged on Item Response Theory (IRT) by Thurstone (1925) and the Three-Parameter Logistic Model (3PLM) by Birnbaum (1968). In 1925, Louis Thurstone propounded Item Response Theory (IRT) and established its conceptual framework in his paper titled, "A Method of Scaling Psychological and Educational Tests." In this influential paper, Thurstone provided a procedure for accurately positioning the Binet scale's items (Oghenerume, 2022).

Cai and Huang (2022) noted that IRT can be viewed minimally as a set of psychometric models for categorical item-level response data. von Davier et al. (2021) described categorical item-level response data as data collected from individuals responding to items with multiple categories or response options which are typically discrete and qualitative, representing different levels or alternatives. As specified by Oghenerume (2022) and Carlson (2020), examples of categorical item-level response data are multiple-choice options with predefined categories or categorical reply choices including strongly disagree, disagree, agree, and strongly agree.

Birnbaum (1968) postulated that the Three-Parameter Logistic Model (3PLM) is rooted in IRT. As a result, the possibility of a correct reply to an item depends upon three parameters: slope 'a' (discrimination), threshold 'b' (difficulty), and lower asymptote 'c' (guessing). 3PLM allows items to vary in their adeptness to split amid test-takers of varying proficiency levels and in their difficulty, while also accounting for the possibility of test-takers with extremely minimal proficiency points guessing the correct replies to items. The psychometric formula of 3PLM describes the possibility that an unintentionally chosen test-taker with competence ' θ ' upon ' k ' measure will perfectly reply to the ' j ' item:

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

In which

x_j = reply to j item;

a_j = discrimination parameter of j item, characterising the slope;

b_j = difficulty of j item, characterising the threshold;

c_j = guessing of j item, accounting for the possibility of test-takers with extremely low proficiency points guessing the correct reply;

D = subjective measuring factor usually fixed to 1.7 to imprecise outcomes in a conventional ogive prototype.

Both IRT and 3PLM operate on certain underlying assumptions when applied to the data analytics of tests, particularly in WASSCE and NECO SSCE 2023 English Language multiple choice tests. de Ayala (2022) and Bock and Gibbons (2021) elucidated these assumptions. Firstly, it is assumed the modelled hidden trait accurately describes the entirety of the hidden competence gap and effectively accounts for test-taker accomplishment. In numerous instances, countless tests assume the requirement of a single latent ability (unidimensionality). Secondly, it is assumed that all items are independent of one another locally when conditioned on the hidden competence. This implies an examinee's reply to a particular item is not influenced by their reply to other items. Lastly, it is posited that a typical examinee's reply to a test item can be effectively modelled using the Item Response Function (IRF). Being contingent upon the specific kind of model in IRT, such as 3PLM, IRF expresses the possibility of attaining a certain grade 'X' allowing every level of the hidden ability, denoted as θ . Such possibility association remains neutral of the θ in the distribution within the populace.

de Ayala (2022) explained that slope is expressed as a value between .2 and +2, where a value below 1 may indicate weaker discrimination, while a value above 1.5 may suggest excessive discrimination.

The threshold is represented by values ranging from .3 to +3. Difficulty values below 1.5 indicate easier items, whereas values above 2.5 indicate more difficult items. Guessing is expressed as a value between 0 and 1. Bock and Gibbons (2021) affirmed that there is no universally agreed-upon specific numerical value for a '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 from 0 to 1. Essentially, lower asymptote values exceeding 0.35 are judged unacceptable. 5-option and 4-option items are bound to have lower asymptote rates of about 0.20 and 0.25 respectively (van der Linden & Hambleton, 2019; Carlson, 2020; Oghenerume & Egberha, 2024).

Oghenerume (2022) conducted a study titled, "Comparative Analysis of Item Statistics of WASSCE and NECO SSCE 2022 Data Processing Multiple Choice Tests Using Item Response Theory." The study analytically compared the item statistics of WASSCE and NECO SSCE 2022 Data Processing multiple-choice tests using IRT 2PLM. Six hypothesised questions were affirmed. In this study, a survey research design was utilised. This study's population comprised ten thousand, eight hundred Data Processing SS 3 students in twenty-one public senior secondary schools in AMAC. A multistage sampling procedure was utilised to pick a sample size of one thousand and eighty Data Processing SS 3 students for this study. The WASSCE and NECO SSCE 2022 Data Processing multiple-choice tests are the instruments utilised for the study. The data collected were subjected to IRT 2PLM using Item Response Theory Assistant for Excel (EIRT) to statistically generate the item parameters (slope and threshold). All hypothesised questions were tested for significance utilising independent samples t-test statistics. This study's findings revealed a significant dissimilarity in the girls' discrimination indices. Significantly, no

dissimilarities existed in the discrimination indices for boys, as well as in the difficulty levels for boys and girls. Additionally, no significant dissimilarities were found in the discrimination indices and difficulty levels between WASSCE and NECO SSCE 2022 Data Processing multiple-choice items. Due to these findings, this study recommends that WAEC and NECO harness the IRT 2PLM to further improve the quality of Data Processing multiple-choice items.

Aborisade and Fajobi (2020) conducted a study titled, “Comparative Analysis of Psychometric Properties of Mathematics Items Constructed by WAEC and NECO in Nigeria Using Item Response Theory Approach.” The study explored the comparability of psychometric properties of Mathematics items constructed by the West Africa Examinations Council (WAEC) and National Examinations Council (NECO). This study utilised the IRT approach and focused on the difficulty index, discrimination power, and distractor parameter. A survey form of descriptive research design was espoused in this study. The population consisted of senior students in secondary schools who took the WAEC and NECO tests in South Western Nigeria in 2019. A sample size comprised one thousand, two hundred senior students in secondary schools, chosen through a multistage sampling process were used. The objective test items structured by both WAEC and NECO are used as instruments. The findings indicated that the difficulty and guessing indicators of the mathematics items structured by WAEC and NECO are comparable. However, the discriminating power of the items is not comparable between the two examination bodies. Based on these findings, the study suggested that credentials released by WAEC and NECO can be implemented for the same intentions without sentiments, as the quality of the items developed by them is equivalent.

Ogunbamowo et al. (2019) conducted a study titled, “Psychometric Properties of 2017 West African Examination Council and National Examinations Council's Economic Senior School Certificate Examination Items.” The study assessed the differences between WAEC and NECO Economics items’ discrimination, difficulty, and guessing employing Item Response Theory and Classical Test Theory. The research design employed was descriptive. The study’s population entailed secondary school students in Osun State, with a sample size of 540 students. The study utilised instruments adopted from the 2017 Economics WAEC and NECO Senior School Certificate Examinations, specifically Economics Achievement Test 1 (EAT 1) and Test 2 (EAT 2). The effects signified there was no significant dissimilarity between the discrimination pointers of NECO and WAEC Economics’ items when CTT was used. Moreover, a significant dissimilarity was observed in the discrimination indices when IRT was applied. Similarly, while there was no significant dissimilarity in the difficulty pointers of NECO and WAEC Economics’ test items with the use of CTT, a significant dissimilarity emerged when IRT was employed. In contrast, the dissimilarity in the guessing indications of NECO and WAEC Economics items was not significant. In conclusion, this study found a comparability of both tests under CTT and a dissimilarity under IRT.

2.1. Statement of the Problem

Over the years, both WAEC and NECO have not been immune to criticism. Daniel (2005) criticised NECO for its inferior test item quality and the credibility of the grades assigned to candidates. These deficiencies led to some federal Universities rejecting NECO results between 2002 and 2012, as documented by Peter (2012). On the contrary, Ahmed (2014) asserted that NECO test items from 2011 to 2014 exhibited higher standards compared to those of WAEC. Additionally, Dibu-Ojerinde and

Faleye (2005) argued that there was no significant dissimilarity between NECO and WAEC, while Oghenerume (2022) affirmed that no significant dissimilarities were found in the discrimination indices and difficulty levels between WASSCE and NECO SSCE 2022 multiple choice items in Data Processing when a comparison was made.

Statistically, the results released by WAEC and NECO (2023) revealed that only 84.38% of candidates who sat for the WASSCE 2023 English Language test earned credit, while 61.60% of candidates who sat for the NECO SSCE 2023 English test earned credit. Notably, 16.29% of WAEC candidates' results are withheld due to examination malpractice, showing a 6.54% decrease compared to the 22.83% recorded in 2022.

Comparatively, English Language tests were conducted for the same group of candidates with the expectation that they had been effectively taught by qualified English Language teachers who covered the syllabi. Additionally, both WAEC and NECO have similar syllabi for English Language and issue senior secondary school certificates with the same goals. Consequently, their test items are expected to be of similar standards, resulting in similar outcomes. Perhaps the underperformance of certain candidates could be linked to the parameters within the multiple-choice items of English Language tests.

Given these issues, this study undertook a meticulous analytical comparison of item statistics for the 2023 WASSCE and NECO SSCE English Language multiple choice tests. This analytical comparison entailed evaluating the three parameters: discrimination, difficulty, and guessing in the multiple-choice items using IRT 3PLM.

2.1.1. Questions

1. Will there be a dissimilarity between the difficulty parameters of WASSCE and NECO SSCE 2023 English Language multiple choice items?

2. Will there be a dissimilarity between the discrimination parameters of WASSCE and NECO SSCE 2023 English Language multiple choice items?
3. Will there be a dissimilarity between the guessing parameters of WASSCE and NECO SSCE 2023 English Language multiple choice items?

2.1.2. Hypothesised Questions

1. No significant dissimilarity between the difficulty parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items.
2. No significant dissimilarity between the discrimination parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items.
3. No significant dissimilarity between the guessing parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items.

3. Methodology

The survey research design was utilised in this investigation because this investigation involved a group of students from which data were obtained, only from a few students considered as emissaries of the all-inclusive group needed. This design is suitable for this study because both WASSCE and NECO SSCE 2023 English Language multiple choice test items were employed to obtain data and analyse the difficulty, discrimination, and guessing parameters of the items using the IRT 3PLM.

According to the Ogun State Ministry of Education, Science, and Technology (2023), this study's population encompassed 72,400 English Language SS 3 students in three hundred and eighty-four public senior secondary schools in Ogun State. However, this study's statistical population are eighty English Language multiple choice test items of WASSCE and one hundred English Language multiple choice test items of NECO SSCE. This study's sample comprised 8,000 English Language SS3 students from 40 schools in Ogun

State, representing 11.05%. They were picked employing a multistage sampling process.

During the first stage, simple random sampling was utilised to pick ten Local Government Areas (LGAs) representing fifty per cent of the twenty LGAs through a balloting method. During the second, proportionate stratified random sampling was employed within the ten LGAs to pick forty schools, considering that some LGAs possess more schools than others. In the third stage, stratified random sampling was employed to stratify by sex within the forty schools. In the fourth stage, a simple random sampling through the method of balloting was implemented to choose 100 boys and 100 girls from each school. From each school, 200 students were sampled, and therefore, a total of 8,000 students were sampled from the 40 selected schools.

The instruments utilized in this investigation are the multiple-choice items in the WASSCE and NECO SSCE 2023 English Language tests. The English Language multiple choice items in WASSCE 2023 consist of 80 items, while the NECO SSCE 2023 contains 100 English Language multiple choice items. In the WASSCE, the English Language multiple-choice items are structured with four options 'A' to 'D,' from which English Language students are required to select the correct option. In the NECO SSCE, they are structured with five options 'A' to 'E,' and students are required to indicate the correct option.

These instruments were deemed to be very valid and highly reliable because WAEC and NECO are reputable for constructing standardised tests and items were still within the three-year limit required before re-establishment of reliability. However, the researchers re-established the reliability through test-retest by administering the instruments for the first time to twenty SS3 students in schools in Ogun State. After two weeks, the same instruments were

administered to the same students. The scores from the first and second administrations were analysed using Pearson Product Moment Correlation to determine the coefficient. A correlation of 0.95 was obtained from WASSCE 2023 English Language multiple-choice items, and 0.93 was obtained from NECO SSCE 2023 English Language multiple-choice items, further affirming that the instruments were highly reliable.

Subsequently, these instruments were dispensed directly to the English Language SS 3 students in the selected schools in Ogun State with the help of the research assistants who are also their English Language teachers. The WASSCE multiple-choice items were administered first. Upon completion, they were immediately collected. After a day, the NECO SSCE multiple-choice items were administered and collected through the same procedure. The collected data were subjected to IRT 3PLM utilising jMetrik psychometric software to analytically generate the item parameters. All hypothesised questions were tested for significance harnessing independent samples t-test statistics in IBM SPSS.

4. Presentation of Results

Hypothesis One: No significant dissimilarity between the difficulty parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items.

Table 1: Thresholds in WASSCE and NECO SSCE 2023 English Language Multiple Choice Items

English Language	N	m	SD	df	t	p
WASSCE 2023	80	2.1473	.38057			
				178	.180	.857
NECO SSCE 2023	100	2.1364	.41616			
$\alpha = .05$.						

Table 1 shows the number of items in the WASSCE 2023 English Language multiple choice test was N=80, while for the NECO SSCE 2023 English Language multiple choice test, it was N=100. Their mean values were 2.1473 and 2.1364, while their standard deviations were .38057 and .41616 respectively. The degree of freedom, t-value, and p-value were 178, .180, and .857 respectively.

As the p-value is more than the alpha's, the hypothesis which affirms "no significant dissimilarity between the difficulty parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items," was retained. This implies that neither of the English Language multiple choice tests developed by both examination bodies, WAEC and NECO, was more difficult than the other.

Hypothesis Two: No significant dissimilarity between the discrimination parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items.

Table 2: Slopes in WASSCE and NECO SSCE 2023 English Language Multiple Choice Items

English Language	N	m	SD	df	t	p
WASSCE 2023	80	1.1316	.26439	178	.489	.625
NECO SSCE 2023	100	1.1096	.32590			

$\alpha = .05$.

Table 2 indicates the number of items in the WASSCE 2023 English Language multiple choice test was N=80, while for the NECO SSCE 2023 English Language multiple choice test, it was N=100. Their mean values were 1.1316 and 1.1096, while their standard deviations were .26439 and .32590 respectively. The degree of freedom, t-value, and p-value are 178, .489, and .625.

As the p-value is more than the alpha's, the hypothesis which affirms "no significant dissimilarity between the discrimination parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items," was retained. This implies that neither of the English Language multiple choice tests developed by both examination bodies, WAEC and NECO, was more discriminating than the other.

Hypothesis Three: No significant dissimilarity between the guessing parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items.

Table 3: Lower Asymptotes in WASSCE and NECO SSCE 2023 English Language Multiple Choice Items

English Language	N	m	SD	df	t	p
WASSCE 2023	80	.2273	.02886	178	-12.627	.000
NECO SSCE 2023	100	.2827	.02961			

$\alpha = .05$. $p < .05$. (Significant).

Table 3 revealed the number of items in the WASSCE 2023 English Language multiple choice test was N=80, while for the NECO SSCE 2023 English Language multiple choice test, it was N=100. Their mean values were .2273 and .2827, while their standard deviations were .02886 and .02961 respectively. The degree of freedom, t-value, and p-value were 178, -12.627, and .000 respectively.

As the p-value is less than the alpha's, the hypothesis which affirms "no significant dissimilarity between the guessing parameters of WASSCE and NECO SSCE 2023 English Language multiple-choice items," was rejected. This implies that the English Language multiple choice test developed by NECO had more items that were more susceptible to guessing than the corresponding WAEC.

4.1. Discussion of Findings

Hypothesis one divulged neither of the English Language multiple choice tests developed by both examination bodies, WAEC and NECO, was more difficult than the other. This finding aligns with the finding of Oghenerume (2022), who conducted a comparative analysis of item statistics for the WASSCE and NECO SSCE 2022 Data Processing multiple-choice tests using item response theory and found no significant dissimilarity in the difficulty levels between the WASSCE and NECO SSCE 2022 Data Processing multiple-choice items. It also agrees with the findings of Aborisade and Fajobi (2020), who conducted a comparative analysis of the psychometric properties of Mathematics items constructed by WAEC and NECO in Nigeria using the IRT approach. They did not find any significant dissimilarity in the difficulty levels of WASSCE and NECO SSCE Mathematics items; instead, they are comparable. Additionally, this supports the findings of Ogunbamowo et al. (2019), who investigated the psychometric properties of the 2017 West African Examinations Council and National Examinations Council's Economics Senior School Certificate Examination items. They did not find any significant dissimilarity amid the difficulty indices of WASSCE and NECO SSCE Economics multiple-choice items using CTT. Conversely, it opposes their finding that a significant dissimilarity existed amid the difficulty parameter using IRT. This discrepancy may be due to the sample size of 540 they used, as IRT typically relies on a sample size closer to 1,000.

Hypothesis two unveiled neither of the English Language multiple choice tests developed by both examination bodies, WAEC and NECO, was more discriminating than the other. This finding aligns with the finding of Oghenerume (2022), who conducted a comparative analysis of item statistics for the WASSCE and NECO SSCE 2022 Data Processing multiple-choice tests using item

response theory and found no significant dissimilarity in the discrimination indices between WASSCE and NECO SSCE 2022 Data Processing multiple-choice items. This supports the findings of Ogunbamowo et al. (2019), who investigated the psychometric properties of the 2017 West African Examinations Council and National Examinations Council's Economics Senior School Certificate Examination items. They found no significant dissimilarity in the discrimination indices of WASSCE and NECO SSCE Economics multiple-choice tests using CTT. Contrariwise, it opposes their finding that a significant dissimilarity existed in the discrimination parameter of WASSCE and NECO SSCE Economics multiple-choice tests using IRT. This discrepancy may be attributed to the sample size of 540 they used, as IRT typically relies on a sample size closer to 1,000.

Additionally, it disagrees with the findings of Aborisade and Fajobi (2020), who conducted a comparative analysis of the psychometric properties of Mathematics items constructed by WAEC and NECO in Nigeria using the IRT approach. They found a significant dissimilarity in the discrimination powers of WASSCE and NECO SSCE Mathematics items. Perhaps the WAEC 2017 Mathematics objective test with 50 items they used as one of their instruments could be the reason for the disparity, whereas the WASSCE 2023 English Language multiple choice test has 80 items.

Hypothesis three exhibited that the English Language multiple choice test developed by NECO had more items that were more susceptible to guessing than the corresponding WAEC test. This finding is not in alignment with the findings of Aborisade and Fajobi (2020), who conducted a comparative analysis of the psychometric properties of Mathematics items constructed by WAEC and NECO in Nigeria using the IRT approach. They did not find any significant dissimilarity in the guessing parameters of WASSCE and NECO SSCE

Mathematics items; instead, they are comparable. Perhaps the WAEC 2017 Mathematics objective test with 50 items they used as one of their instruments could be the reason for the disparity, whereas the WASSCE 2023 English Language multiple choice test has 80 items. Additionally, this disagrees with the findings of Ogunbamowo et al. (2019), who investigated the psychometric properties of the 2017 West African Examinations Council (WAEC) and National Examinations Council (NECO) Economics Senior School Certificate Examination items. They did not find any significant dissimilarity in the guessing parameters of WASSCE and NECO SSCE multiple-choice items in Economics using IRT. Perhaps this discrepancy could be attributed to the sample size of 540 they used, as IRT typically relies on a sample size closer to 1,000.

5. Conclusion and Recommendations

This assessment encompassed the analytical comparison of item statistics of WASSCE and NECO SSCE 2023 English Language multiple choice tests using the Item Response Theory Three-Parameter Logistic Model (IRT 3PLM). In line with the outcomes in this analytical comparison, the conclusion was that no significant dissimilarities in the difficulty and discrimination parameters between WASSCE and NECO SSCE 2023 English Language multiple-choice items. However, a significant dissimilarity existed in the guessing parameters between these two. The pragmatic extrapolation of this analytical comparison accentuated in the findings alluded that employing IRT 3PLM to establish the item statistics before administering the tests could further enhance the quality of English Language multiple choice items developed by WAEC and NECO.

Consequently, due to the results and conclusion of this analytical comparison, it is recommended that WAEC and NECO employ IRT 3PLM to determine the discrimination, difficulty, and

guessing parameters of their English Language test items before administering them, if they are not presently using it. Through the utilisation of IRT 3PLM, WAEC and NECO can detect items that do not satisfy the three parameters: discrimination, difficulty, and guessing. This is because items should not excessively discriminate between higher and lower achievers and should not be too difficult for the students.

Prominently, employing IRT 3PLM would help WAEC and NECO identify English Language multiple-choice items that are susceptible to guessing, allowing for their removal or refinement to mitigate guessing by students in subsequent tests.

6. 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.
- 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.
- 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. (2020). Introduction to item response theory models and applications. Routledge.
- Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511486999>
- 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.
- Federal Republic of Nigeria. (2013). *National policy on education* (4th ed.). NERDC Press.
- Foyewa, R. A. (2020). A review of the status of English Language in the Nigerian school system. *International Journal of English Language Teaching*, 8(6), 63–69.
- National Examinations Council. (2023). *Statistical report*. <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>
- 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>
- Ogun State Ministry of Education, Science and Technology (2023). *Statistics of senior students offering English Language*.
- Ogunbamowo, A. O., Adediwura, A. A., & Diyan, R. O. (2019). Psychometric properties of 2017 West African Examinations Council and National Examinations Council's Economic Senior School Certificate Examination items. *Journal of Advances in Education and Philosophy*, 3(1), 9–18.
- 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.
- Rao, P. S. (2019). The role of English as a global language. *Research Journal of English*, 4(1), 65–79.
- 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., & Hambleton, R. K. (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*. <https://www.waeconline.org.ng>