



ANALYSIS OF ITEM DIFFICULTY AND DISCRIMINATION DRIFT IN THE 2023 AND 2024 MATHEMATICS WEST AFRICAN SENIOR SCHOOL CERTIFICATE EXAMINATIONS USING THE FOUR-PARAMETER LOGISTIC MODEL OF ITEM RESPONSE THEORY

Prof. N.P.M. Esomonu

npm.esomonu@unizik.edu.ng

Department of Educational Foundations, Faculty of Education Nnamdi Azikiwe University, Awka. Anambra State

Chinweude Clara Efobi

clara.efobi@fcetumunze.edu.ng

Psychology Department, School of Education, Federal College of Education (Technical), Umuze, Anambra State

ABSTRACT

Without empirical evidence-based information on drift analysis of examination items, high-stake examination bodies may continue to use or compare items across years without knowing whether their measurement properties have changed. So, there is need for Analysis of Item Difficulty and Discrimination drift in the 2023 and 2024 Mathematics West African Senior School Certificate Examination using the four-parameter logistic model (4PLM) of item response theory. The study determined whether the two examination years exhibit comparable distributions of difficulty and discrimination after accounting for guessing and the possibility that highly proficient candidates may answer an item incorrectly. Three research questions guided the study. The study adopted longitudinal research design repeated measure. The target population is comprised of 22500 SS3 Students in public secondary schools across the six educational zone of Anambra State. The sample comprised 1125 students obtained through multi-stage sampling procedure using simple random technique and purposive sampling. 2023 and 2024 May/June WASSCE Mathematics multiple-choice questions was adopted instrument used to generate data. The test was dichotomously scored 1 for correct answer and 0 for wrong answer using the scoring guide and the results were used to determine the extent of drift in the difficulty and discrimination, parameters across administrations, Data were analysed using Item Parameter Drift (IPD) analysis within the framework of IRT. Results showed that the majority of the 2023 and 2024 WASSCE Mathematics Multiple- Choice Questions exhibited moderate drift in item difficulty parameters across administrations, indicating that most items maintained relatively acceptable stability over time. But the 2024 examination recorded a higher proportion of high-drift items than the 2023 examination, indicating comparatively lower stability in item difficulty characteristics. The findings of the present study therefore suggest the need for careful psychometric review, recalibration, revision, or replacement of high-drift items before they are reused in future WASSCE Mathematics examinations. Such measures would help strengthen score comparability, preserve measurement accuracy, and improve the overall integrity of the examination system.

Keywords: Drift, Difficulty, Discrimination. Mathematics, Multiple-choice, WASSCE, 4PLM of IRT



Introduction

The Mathematics West African Senior School Certificate Examination (WASSCE) is used in high-stake decisions such as certification and admission into institutions of higher learning, therefore their items should measure mathematical proficiency fairly across examination years. Candidates' opportunity for certification or admission should not depend on which year they sat a West Africa Senior School Certificate Examination. WASSCE Item parameters should remain reasonably stable across examination years particularly item difficulty and discrimination of mathematics multiple-choice items. Stability in these parameters is essential because WASSCE results are used for certification, admission into tertiary institutions and other high-stakes educational decisions. However, differences in item construction, curriculum coverage, and exposure of instrument may cause Mathematics items to function differently from one year to another. Such variations may produce drift in the item parameters,

Item Parameter Drift (IPD) refers to systematic changes in item characteristics such as difficulty, discrimination, or guessing across test administrations (Wollack et al., 2021). When IPD occurs, the stability of the measurement scale is compromised, and the validity of longitudinal comparisons is threatened. Drift in item difficulty is when items become relatively easier or more difficult, and drift in item discrimination is where items change in their ability to distinguish effectively between candidates of different proficiency levels. If it is not empirically identified, candidates with similar levels of mathematical proficiency may receive different scores simply because they responded to examination years with different measurement properties. This threatens fairness of candidate's classification, comparability of examination years, validity of interpretations and confidence in decisions based on WASSCE results.

Existing analyses of public examinations frequently rely on classical test theory or less complex IRT models that may not sufficiently account for guessing and the possibility that highly proficient candidates may still answer some items incorrectly. Consequently, evidence concerning the stability of both difficulty and discrimination parameters in successive WASSCE Mathematics examinations remains insufficient. This study therefore applies the Four-Parameter Logistic Item Response Model and an appropriate cross-year linking procedure to analyse drift in the difficulty and discrimination parameters of the 2023 and 2024 WASSCE Mathematics multiple-choice examinations and determine the implications for test comparability, score interpretation and psychometric quality of the two examination years. Research on a WAEC Core Mathematics examination in Ghana, for example, found that most items discriminated among examinees while many were also susceptible to guessing (Cobbinah & Ntumi, 2022). This finding illustrates why a multiple-choice item's observed success rate alone cannot establish its measurement quality. For the 2023 and 2024 WASSCE Mathematics multiple-choice examinations, differences in content emphasis, item wording, cognitive demands, instructions and candidate populations could affect estimated difficulty and discrimination. Such differences do not, by themselves, prove that either examination was unfair; they do mean that a comparison of scores across years needs psychometric support. If one year contains a greater concentration of difficult items, or items that distinguish less effectively among candidates, similar observed scores may not convey precisely the same information about proficiency. The unresolved challenge is to establish whether the difficulty and discrimination characteristics of the two years are sufficiently comparable, and to identify the nature and extent of any differences.



The challenge has persisted because examining item quality within a single year is easier than establishing what changes across years. Researchers have made substantial progress in analysing Mathematics examination items, but their studies address different assessments, populations and measurement questions. Using IRT to compare WAEC and NECO Mathematics items, Aborisade and Fajobi (2020) found comparable difficulty and guessing indices but differences in discrimination. Their result shows that agreement on one item property cannot be taken as evidence of agreement on another; it also concerns two examination bodies rather than successive WASSCE years. Cobbinah and Ntumi's (2022) analysis likewise demonstrated the value of estimating difficulty, discrimination and guessing, although its focus was a single WAEC examination in Ghana. More directly relevant to change over time, Esomonu and Okoli (2023) investigated WAEC Mathematics items from earlier years in relation to school type and reported significant difficulty differences for some items. Their work indicates that item characteristics warrant continued scrutiny, but its school-type comparison does not answer the question concerning the 2023 and 2024 WASSCE years. Opesemowo et al. (2026) compared Mathematics items from the 2022 and 2023 Osun State Basic Education Certificate Examinations using a three-parameter IRT model. They identified items with weak difficulty characteristics, although their overall comparison found no statistically significant difference between years, that study provides a useful example of cross-year investigation at the state examination level; its findings cannot be assumed to apply to WASSCE. Taken together, this literature shows both the success and the limit of previous approaches. Researchers have identified weaknesses that a total score can conceal and have shown that difficulty and discrimination may tell different stories. The cited studies, however, do not establish the cross-year stability of both properties in the 2023 and 2024 WASSCE Mathematics multiple-choice examinations using a four-parameter logistic model (4PLM). That model also represents guessing and the possibility that even highly proficient candidates may answer an item incorrectly. The available analyses do not yet show how the two target years compare when these features of multiple-choice responding are considered together.

The value of drift analysis moves examination quality assurance from assuming that items remain stable to testing whether they actually do. Without drift analysis, an examination body may continue to use or compare items across years without knowing whether their measurement properties have changed. Identifying drift, provide empirical evidence on which items remain stable, which require further investigation, and whether changes in item parameters have implications for interpreting candidates' performance. Therefore, this study investigated whether the measurement characteristics of 2023 and 2024 WASSCE Mathematics items using 4PLM of Item Response Theory are sufficiently stable to support the WASSCE assessment decisions for which they are used.

Purpose of the Study

The purpose of this study is to analyse the cross-year stability of the item parameters of the 2023 and 2024 WASSCE Mathematics multiple-choice examinations using the Four-Parameter Logistic Item Response Model. The study specifically seeks to;

1. determine the magnitude of drift in the difficulty parameters of the 2023 and 2024 WASSCE Mathematics multiple-choice items.
2. determine the magnitude of drift in the discrimination parameters of the 2023 and 2024 WASSCE Mathematics multiple-choice items.



3. Compare the degree of cross-year instability between the difficulty and discrimination parameter distributions.

Research Questions

The following research questions guided the study;

1. What is the magnitude of drift in the difficulty parameters of the 2023 and 2024 WASSCE Mathematics multiple-choice items?
2. What is the magnitude of drift in the discrimination parameters of the 2023 and 2024 WASSCE Mathematics multiple-choice items?
3. To what extent does cross-year instability differ between the difficulty and discrimination parameter distributions of the Mathematics WASSCE multiple-choice items

Methodology

The research design that guided this study is longitudinal survey research design. A longitudinal survey is a research design in which the same participants are measured multiple times typically with the same instrument, to examine changes, trends or stability over time (Crosswell, 2014). This design was considered most appropriate for this present study because the central focus was the examination of item parameter drift of WASSCE Mathematics multiple-choice items across different examination years. The repeated measure component of the design involved the analysis of the same set of Mathematics multiple-choice items administered in different WASSCE examination years to the same group of senior secondary school three (3) students over time. It ensured that item characteristics were repeatedly estimated over time thereby making it possible to detect changes in item parameters attributable to temporal factors rather than to differences in test content or population structure. Three research questions guided the study. The target population is comprised of 22500 SS3 Students in public secondary schools across the six education zone of Anambra State. A multi-stage sampling procedure was used to select 1125 SS3 Students from the six educational zone representing 5% of the total population. First stage, stratified sampling was employed to divide Anambra state into its six educational zones: Aguata, Awka, Nnewi, Ogidi, Onitsha and Otuocha. At the second Stage, simple random sampling technique was used to sample three (3) educational zones in Anambra State. This sampling was achieved by writing names of the educational zone on pieces of paper. The pieces of paper were properly folded and three educational zone were randomly picked with replacement. This was to ensure that each education zone had equal chance of being selected into the sample. Then in the second stage, from each selected education zone, a list of all public secondary schools was obtained from Anambra state ministry of Education. Using simple random sampling eight public secondary schools were sampled from each education zone yielding a total of twenty-four public secondary schools. At the final stage, one intact class of SS3 students was selected purposively from the sampled school. This is primarily undertaken to ensure that participants possessed the required knowledge and experience relevant to WASSCE Mathematics and to facilitate data collection. The instrument for data collection was adopted May/June WASSCE Mathematics paper 111, it is a multiple –choice question containing 50 items based on WASSCE syllabus. The instruments were adopted because WAEC is a recognized examination body with a statutory mandate to develop, validate and administer standard tests across member countries. WAEC May/June mathematics multiple-choice test items of 2023 and 2024 was administered to SS3 students with the help of their mathematics teachers in each of the sampled schools. The



mathematics teachers' that were involved in the administration was guided by the researcher, to ensure uniformity in the administration of the instruments. The instruments were also collected on the spot to avoid missing some scripts. Then the scoring was done by the researcher. The test was dichotomously scored 1 for correct answer and 0 for wrong answer using the scoring guide and the results were used to determine the extent of drift in the difficulty, discrimination, guessing, and upper asymptote parameters across administrations, were analysed using Item Parameter Drift (IPD) analysis within the framework of IRT. The analyses focused on the stability of the item parameters across the first administration (A_1), second administration (A_2), and third administration (A_3). Pairwise drift indices represented by Δ_{12} , Δ_{23} , and Δ_{13} were computed to determine the magnitude of parameter changes across administrations. The highest absolute drift value among the three pairwise comparisons was used to determine the overall extent of drift for each item. For difficulty and discrimination parameters, drift values ranging from 0.00 to 0.20 were classified as low drift, values from 0.21 to 0.50 as moderate drift, while values above 0.50 were classified as high drift. Descriptive statistics including frequencies and percentages were computed to summarise the extent of parameter drift across administrations.

Results

The 2023 and 2024 WASSCE multiple-choice showed that several items demonstrated moderate stability across administration, but substantial number exhibited significant difficulty drift too. The presence of high-drift items suggests the need for psychometric review, recalibration, revision, or possible replacement of unstable items before reuse in future WASSCE Mathematics examination to ensure fairness, consistency, and comparability of examination scores across administrations.

Table 1: Summary levels of drift in the difficulty parameter (b) for the 2023 and 2024 WASSCE Mathematics Multiple Choice Questions.

Drift Level	2023 Frequency	2023 Percentage (%)	2024 Frequency	2024 Percentage (%)
Low Drift	2	4.0	3	6.0
Moderate Drift	37	74.0	32	64.0
High Drift	11	22.0	15	30.0
Total	50	100.0	50	100.0

The summary results in Table 3 for the 2023 examination revealed that 2 items, representing 4.0% of the total items, exhibited low drift, indicating strong stability in item difficulty across administrations. Furthermore, 37 items, representing 74.0%, recorded moderate drift, while 11 items, representing 22.0%, and demonstrated high drift.

For the 2024 examination, 3 items, representing 6.0%, exhibited low drift, while 32 items, representing 64.0%, recorded moderate drift. Additionally, 15 items, representing 30.0%, demonstrated high drift. The larger proportion of high-drift items in 2024 compared with 2023 indicates greater instability in the difficulty characteristics of the 2024 examination items. Moderate drift was the dominant category in both years, suggesting that most items maintained acceptable stability across administrations. However, the comparatively higher proportion of high-drift items in 2024 indicates lower psychometric stability relative to the 2023 examination. These findings imply the need for continuous item



monitoring, recalibration, revision, and replacement of unstable items to ensure fairness, validity, reliability, and comparability of examination scores across administrations

The 2023 WASSCE Mathematics items maintained stable discrimination parameters across administrations, as reflected in the large number of low-drift items. but, the presence of a few high-drift items indicates that some items did not function consistently over time and therefore require psychometric review, recalibration, or revision before future use. The 2024 examination recorded a slightly greater number of high-drift items, suggesting comparatively lower stability in item discrimination. The observed instability may have resulted from repeated item exposure, curriculum changes, instructional differences, or variations in candidate preparedness. Consequently, items with high discrimination drift should be carefully reviewed, recalibrated, revised, or replaced before future administrations to ensure fairness, validity, and reliability of the examination

Table 2; Summary of Drift in Discrimination Parameter (a) – WASSCE Mathematics Multiple Choice Questions (2023 and 2024)

Drift Level	2023 2023 Percentage		2024 2024 Percentage	
	Frequency	(%)	Frequency	(%)
– 0.20)	34	68.0	33	66.0
Moderate Drift (0.21–0.50)	13	26.0	11	22.0
High Drift (Above 0.50)	3	6.0	6	12.0
Total	50	100.0	50	100.0

Low Drift (0.00

The results in Table 2 for the 2023 examination revealed that 34 items, representing 68.0% of the total items, exhibited low drift in the discrimination parameter. This indicates that the majority of the items maintained relatively stable discrimination power across the three administrations. Such stability implies that these items consistently distinguished between high-ability and low-ability candidates over time. Furthermore, 13 items, representing 26.0%, recorded moderate drift, suggesting noticeable but manageable changes in item discrimination across administrations. In addition, only 3 items, representing 6.0%, demonstrated high drift, indicating substantial instability in the discrimination parameter for a relatively small proportion of the items. For the 2024 examination, 33 items, representing 66.0%, recorded low drift, while 11 items, representing 22.0%, exhibited moderate drift in the discrimination parameter. Additionally, 6 items, representing 12.0%, showed high drift. The increase in the number and percentage of high-drift items in 2024 compared with 2023 suggests greater instability in the discrimination characteristics of the 2024 examination items.

The summary Table 2 indicates that low drift was the dominant category in both years, suggesting that most items maintained relatively stable discrimination power across administrations. However, the 2024 examination recorded a higher proportion of high-drift items than the 2023 examination, indicating comparatively lower stability in some items. Such instability may have resulted from repeated item exposure, ambiguous wording,



curriculum emphasis, ineffective distractors, or variations in candidate preparedness. Consequently, items exhibiting high discrimination drift should be carefully reviewed, recalibrated, revised, or replaced before future administrations to ensure fairness, validity, reliability, and measurement consistency of the examination.

A comparison of the drift levels for the discrimination parameter (a) in the 2023 and 2024 WASSCE Mathematics Multiple Choice Questions revealed noticeable difference in 2024 were classified under low drift. This suggests that most items consistently differentiated between high-ability and low-ability candidates across administrations in both examination years. However, the comparison further showed that the 2024 examination recorded a higher proportion of items with high drift than the 2023 examination. In 2023, only 6.0% of the items exhibited high drift, whereas 12.0% of the items in 2024 fell into the high drift category. This indicates that more items in the 2024 examination experienced substantial changes in discrimination power across administrations. Such changes imply greater instability in the functioning of the 2024 items compared with those of 2023.

Furthermore, the proportion of items with moderate drift was slightly higher in the 2023 examination than in the 2024 examination. Specifically, 26.0% of the items in 2023 exhibited moderate drift compared with 22.0% in 2024. Although these items still functioned reasonably well, the observed moderate drift suggests that some items experienced noticeable fluctuations in their ability to discriminate effectively between candidates of different ability levels. The higher proportion of high-drift items in 2024 may be attributed to factors such as repeated item exposure, ambiguous item wording, curriculum changes, instructional differences, ineffective distractors, coaching effects, or inconsistencies in candidate preparedness across administrations. These findings suggest that the 2024 WASSCE Mathematics Multiple Choice Questions were relatively less stable psychometrically than those of 2023 in terms of discrimination parameter stability. Overall, the comparison indicates that although both examination years demonstrated relatively good stability in most items, the 2023 examination exhibited better discrimination parameter stability than the 2024 examination. Consequently, the 2024 items with high drift require closer psychometric evaluation, recalibration, revision, or possible replacement to improve the consistency, fairness, validity, and comparability of future WASSCE Mathematics examinations. Between 0.10 and 0.19 were regarded as Moderate Drift, while values of 0.20 and above were considered High Drift

Discussion

The present analysis showed that the majority of the 2023 and 2024 WASSCE Mathematics Multiple Choice Questions exhibited moderate drift in item difficulty parameters across administrations, indicating that most items maintained relatively acceptable stability over time. But the 2024 examination recorded a higher proportion of high-drift items than the 2023 examination, indicating comparatively lower stability in item difficulty characteristics. In Item Response Theory (IRT), item parameter drift refers to changes in item characteristics across administrations that may affect score comparability and measurement consistency. The predominance of moderate-drift items in the present study therefore suggests that most WASSCE Mathematics items retained reasonably stable difficulty characteristics across administrations. Nevertheless, the presence of several high-drift items indicates that some items experienced substantial changes in difficulty over time, thereby raising concerns about the consistency of item functioning. As regards drift in discrimination parameter, the findings of the study revealed that the majority of the 2023



and 2024 WASSCE Mathematics Multiple Choice Questions exhibited low drift in discrimination parameters across administrations, indicating that most items maintained relatively stable discriminatory power over time and consistently differentiated between high-ability and low-ability candidates. In Item Response Theory (IRT), discrimination parameter drift occurs when the ability of an item to distinguish among candidates changes across test administrations. The predominance of low-drift items observed in the present study therefore suggests that most WASSCE Mathematics items retained stable psychometric functioning and continued to discriminate effectively between candidates of different ability levels over time.

This finding corroborates the work of Esomonu and Okoli (2023), who reported evidence of item parameter drift in WAEC Mathematics examinations across school types. Comparable findings were also obtained by Ighodaro and Oribhabor (2023), who observed moderate and high parameter drift among Chemistry multiple-choice items in NABTEB examinations. Similarly, Oguguo and Nwani (2024) found instability in item parameter estimates among Basic Science examination items due to differences in instructional exposure and candidate preparedness. The similarity between these studies and the present findings likely reflects the influence of repeated item usage, differences in instructional delivery, and variations in candidate preparation across administrations. Large-scale public examinations conducted repeatedly over time are particularly vulnerable to item drift because increased exposure to test items may alter candidates' familiarity with the questions and consequently affect their difficulty estimates. The findings of the present study therefore suggest the need for careful psychometric review, recalibration, revision, or replacement of high-drift items before they are reused in future WASSCE Mathematics examinations. Such measures would help strengthen score comparability, preserve measurement accuracy, and improve the overall integrity of the examination system.

Conclusion

In conclusion, the drift analyses revealed that the 2024 WASSCE Mathematics Multiple Choice Questions recorded higher proportions of high drift items across difficulty and discrimination, indicating comparatively lower item stability than the 2023 examination. These findings suggest that the 2023 WASSCE Mathematics Multiple Choice Questions demonstrated greater overall item stability and consistency, while the 2024 examination demonstrated stronger measurement precision and broader content representation.

Recommendations

Based on the findings of this study, the following recommendations were made:

1. Examination bodies and Mathematics teachers should place greater emphasis on the development of candidates' conceptual understanding, analytical reasoning, and problem-solving skills in order to adequately prepare students for increasingly demanding Mathematics examinations, particularly considering the higher proportion of difficult items observed in the 2024 WASSCE Mathematics Multiple Choice Questions.
2. Test developers and item writers should construct more highly discriminating items capable of effectively distinguishing between high-performing and low performing candidates. This can be achieved through regular training in psychometric principles, item writing techniques, and quality assurance procedures.



REFERENCES

- Aborisade, O., & Fajobi, 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*. <https://doi.org/10.5897/err2019.3850>
- Cobbinah, A., & Ntumi, S. (2022). Difficulty, Discrimination and Pseudo - Guessing Indices of West African Examinations Council Core Mathematics Multiple Choice Items: Theoretical and Practical Implications of Using Item Response Theory. *Journal of Research in Educational Sciences*. <https://doi.org/10.14505/jres.v13.15.06>
- Esomonu, N. P.-M., and Okoli, I. N. (2023). Item parameter drifts across school type in WAEC Mathematics examination from 2018–2019. *International Journal of Advanced Education and Research*, 8(3), 35–41.
- Ighodaro, A. E., and Oribhabor, B. C. (2023). Comparison of methods to detect item parameter drift in NABTEB SSCE 2012–2015 chemistry multiple-choice tests. *Al-Hikmah Journal of Arts and Social Sciences Education*, 5, 2715–2559
- Ighodaro, A. E., and Oribhabor, B. C. (2023). Comparison of methods to detect item parameter drift in NABTEB SSCE 2012–2015 chemistry multiple-choice tests. *Al-Hikmah Journal of Arts and Social Sciences Education*, 5, 2715–2559
- Oguguo, B. E. C., and Nwani, S. U. (2024). Effect of item drift on item parameter estimation of multiple-choice items of Basic Education Certificate Examination in Basic Science. *African Journal of Science, Technology, Mathematics and Education (AJSTME)*, 10(3), 433–439
- Opesemowo, O., Opatunji, K. O., Babatimehin, T., & Opesemowo, T. (2026). Analysis of 2022 and 2023 Osun State basic education certificate examination mathematics items using item response theory: Implications for large scale assessment. *Social Sciences & Humanities Open*. <https://doi.org/10.1016/j.ssaho.2025.102381>
- Wollack, J. A., Cohen, A. S., and Wells, C. S. (2021). Item parameter drift in educational assessments: Detection and implications. *Applied Psychological Measurement*, 45(2), 44-54.