



Correlational And Gender-Based Analysis Of Item Parameters Of Geography Qualifying Examinations Using Item Response Theory In Senior Secondary Schools In South-West Nigeria

Dr. Kolawole Alaba Gbenga

Department of Educational Foundations, Faculty of Education, Federal University Oye Ekiti

Received: 25.07.2026 | Accepted: 08.10.2026 | Published: 20.10.2026

*Corresponding Author: Dr. Kolawole Alaba Gbenga

DOI: [10.5281/zenodo.23498056](https://doi.org/10.5281/zenodo.23498056)

Abstract	Original Research Article
His study investigated the correlational and gender-based analysis of item parameters of Geography qualifying examinations using Item Response Theory (IRT) in senior secondary schools in South-West, Nigeria. The study adopted an ex-post-facto research design, examining the responses of 1,200 Senior Secondary School II students who sat for the 2021 Geography Qualifying Examinations. A multistage sampling procedure was employed to select students' responses to 50 multiple-choice Geography items from Ekiti State Qualifying Examination (ESQE) and Oyo State Qualifying Examination (OYQE), and 60 multiple-choice items from Ogun State Qualifying Examination (OGQE). The instruments were the adopted standardized qualifying examination question papers for 2021. Data were obtained from the respective State Ministries of Education through Optical Mark Reader (OMR) sheets. Responses were scored dichotomously, with correct options coded as 1 and incorrect options as 0. Item parameters of difficulty (b), discrimination (a), and guessing (c) under the One-Parameter Logistic (1PL), Two-Parameter Logistic (2PL), and Three-Parameter Logistic (3PL) models were estimated using BILOG-MG software. Hypotheses formulated were tested at 0.05 level of significance using Pearson Product-Moment Correlation Coefficient. Findings revealed that difficulty indices of Oyo and Ekiti States were significantly related, unlike Ogun State. A significant moderate linear relationship was observed between male and female difficulty parameters in Oyo State, while Ekiti and Ogun States exhibited low, non-significant relationships. Similarly, Pearson correlation for discrimination parameters was very low and non-significant across states; no significant linear relationship was found between male and female discrimination parameters in Oyo and Ekiti States, whereas Ogun State showed a low but significant relationship. It was therefore recommended that thorough item analysis should be institutionalized and given high priority by psychometric experts and test development specialists in Ministries of Education prior to final administration to ensure psychometric quality, fairness and gender invariance of test items. Keywords: Item Response Theory, Geography examinations, Item analysis, Gender differences, Psychometric properties.	
Copyright © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)	

Introduction

The education system in Nigeria follows the 6-3-3-4 structure, comprising six years of primary education, three years of junior secondary education, three years of senior secondary education, and four years of

higher education (Federal Republic of Nigeria, 2013). At the senior secondary level, students are groomed for external examinations, notably the West African Senior School Certificate Examination (WASSCE) and the National Examinations Council



(NECO) examinations. To improve students' readiness and mitigate large-scale failure in these external examinations, State Ministries of Education in the South-West geo-political zone conduct Qualifying or Mock Examinations for SSS II students (Ojerinde, 2020). These qualifying examinations serve as a screening measure and are high-stakes, as students' performance in them determines their eligibility to be presented for external examinations (Abiri, 2017). Consequently, the psychometric properties of the test items are critical to ensuring valid certification and placement decisions. Geography is a core subject at the senior secondary school level, especially for students in the Arts and Social Sciences tracks (Filani, 2020). However, observed inconsistencies in students' achievement and variations in item quality across Ekiti, Oyo and Ogun States have raised concerns about the comparability, equating and fairness of items in Geography qualifying examinations (Olatunji, 2021). Gender remains a significant variable in the item analysis of Geography tests, as performance differences between male and female examinees may reflect variations in item characteristics rather than true differences in geographical ability. Item Response Theory (IRT) provides a robust framework for such analysis through its one-parameter (1PL), two-parameter (2PL), and three-parameter (3PL) logistic models. The 1PL model estimates item difficulty (b-parameter) assuming equal discrimination, the 2PL model adds item discrimination (a-parameter) to evaluate how well an item differentiates between high and low ability examinees, while the 3PL model further incorporates a pseudo-guessing parameter (c-parameter) to account for correct responses by guessing. Applying the 1PL, 2PL, and 3PL models to Geography items therefore allows for a comprehensive examination of how items function across gender groups. The quality of a test is defined by its item characteristics, which directly affect examinees' response patterns. In objective formats like multiple-choice tests, examinees must choose the most appropriate option from a list of alternatives (Abegunde, 2018). Effective item analysis therefore examines difficulty, discrimination and distractor efficiency. Difficulty (p) represents the proportion of examinees who answered correctly, with values from

0 to 1.0; higher values denote easier items and lower values denote more difficult items. Items with p-values of 1.0 or 0.0 are psychometrically non-informative as they fail to capture individual differences. Discrimination, ranging from 0.0 to 1.0, measures how well an item separates high-ability from low-ability examinees; a good item is typically answered correctly by high scorers and incorrectly by low scorers. Item analysis involves evaluating examinees' responses to determine the quality of individual items and the test as a whole (Eleje, Onah & Abanobi, 2018). This quality is described through item parameters of difficulty, discrimination and guessing (Orheruata, 2015). Under IRT, difficulty (b) indicates the ability level at which there is a 50% probability of a correct response, discrimination (a) shows the item's capacity to differentiate examinees above and below the difficulty threshold, and guessing (c) estimates the chance of low-ability examinees answering correctly by guessing (Embretson & Reise, 2010). Although these parameters can be estimated under both Classical Test Theory (CTT) and Item Response Theory (IRT), IRT has become the dominant and superior approach. CTT relies on the linear model $X = T \pm E$, where observed score is a function of true score and error, and assumes a single, sample-dependent index of measurement precision for all examinees (Abegunde, 2018). Conversely, IRT uses non-linear logistic models to link latent ability to the probability of endorsing an item (Cappelleri, Lundy & Hays, 2014) and estimates ability from item response patterns (Ojerinde, 2013). With its sample-invariant estimates, ability-specific precision, and capacity to simultaneously calibrate difficulty, discrimination and guessing parameters in the 1PL, 2PL and 3PL models, IRT addresses key limitations of CTT and is now widely adopted for item calibration, test development and analysis of item functioning (Jabrayilov, Emons & Sijtsma, 2016).

Statement of the Problem

The deterioration in secondary school students' academic performance in external examinations, notably among those in Senior Secondary schools in the southwest, has become a significant source of

concern for subsequent administrations and important participants in the nation's education sector. The majority of students who have taken the West African Examinations Council (WAEC) and the National Examination Council (NECO) examinations over the years have done so in May/June and November/December have failed in geography, with their high failure rate being a significant factor in the students' success in Southwest Nigeria. Furthermore, according to statistics provided by the two examination bodies, the National Examination Council and the West African Examination Council, on which this study is specifically focused as well as by the WAEC examiner's report, students typically underperform each time the results of the West African Examination Council (WAEC) and National Examination Council (NECO) examinations are released. As a result, education experts and the government were forced to act as a result of students' persistently poor and unimpressive performance in Geography WASSCE and NECO examinations, particularly in Southwest, Nigeria, over the last three academic sessions (2016/2017, 2017/2018, 2018/2019). These examinations were administered by the West Africa Examination Council (WAEC). The National Bureau of Statistics reported in 2019 Qualifying examinations for Senior Secondary School are held throughout the southwest of Nigeria.

Qualifying examination is held to examine their stability in order to be sponsored by the government to write the final Senior School Certificate Examination (SSCE), this is overseen by the National Examination Council and the West African Examination Council (WAEC) (NECO). Unfortunately, the majority of students who took the June/July National Examination Council (NECO) and May/June West African Examinations Council (WAEC) have reported mass failure in geography and other core subjects and their overall performance. Other researchers have looked into some of the aspects that may be responsible, such as social-economic status, teaching method, and environmental factor, but there is a need to look into the item analysis of the questions used to promote them to face external examination. This could be due to the students' lack of experience with standardized

tests, necessitating the need for the Senior Secondary School Two qualifying examination to be standardized in order to generate a reliable and valid test score that could tell us the students' actual standing before presenting them for external examination.

The two examination bodies (WAEC and NECO) did, without doubt, conduct item analysis on their test items. However, there is no proof that the geography test questions used in the Southwest's Senior Secondary School qualifying examination were subjected to item analysis before final administration. It is also likely that the quality of the examination items in each state's qualifying examination differs, which could explain their disparities in external examination results. One of several issues that may be accountable for the poor academic performance in the final results provided by external examination bodies is the low level of standardization of the States geography qualifying examination items. The inadequate level of standardization of SS II qualifying examination in Geography items in various States in the southwest, Nigeria may be responsible for its inability to predict students' performance in SSCE.

Research Hypotheses

The following research hypotheses were formulated for the study:

1. There is no significant relationship between male and female students in the one parameter of the Geography items of the qualifying examination for Senior Secondary School 2 students in each of the states in southwest.
2. There is no significant relationship between male and female students in the two parameters of the Geography items of the qualifying examination for Senior Secondary School 2 students in each of the states in southwest.
3. There is no significant relationship between male and female students in the three parameters of the Geography items of the qualifying examination for Senior Secondary School 2 students in each of the states in southwest.

Materials and Methods

The study adopted an ex-post facto research design. The design was considered appropriate as the responses of Senior Secondary School II students who sat for the 2021 Geography qualifying examinations in South-West, Nigeria had already occurred and were not subject to manipulation by the researcher. The ex-post facto approach allows for systematic examination of variables as they exist in their natural setting. The population comprised all public Senior Secondary School II students who sat for the 2021 Geography Qualifying Examinations in Ekiti, Oyo and Ogun States. Their Optical Mark Reader (OMR) response sheets constituted the data source. A total of 1,200 students' responses were sampled, consisting of responses to 50 multiple-choice Geography items from the Ekiti State Qualifying Examination (ESQE) and Oyo State Qualifying Examination (OYQE), and 60 multiple-choice items from the Ogun State Qualifying Examination (OGQE) for 2021. A multistage sampling procedure was employed for sample selection. The research instruments were the adopted 50-item ESQE and OYQE and the 60-item OGQE multiple-choice Geography tests for 2021. Students'

responses were obtained from the respective State Ministries of Education in the form of OMR sheets. For scoring, correct responses were coded as "1" and incorrect responses as "0". Data were analyzed using Item Response Theory (IRT) calibration for 1PL, 2PL and 3PL models to estimate difficulty, discrimination and guessing parameters, and inferential statistics were used to examine correlational patterns and gender-based differences in item parameters.

Results and Discussions

The Item parameters (1-Parameter Logistic, 2-Parameter Logistic and 3-Parameter Logistic) were estimated using BILOG-MG program. The research hypotheses generated were tested with inferential statistics such Pearson Product-Moment Correlation Coefficient. The hypotheses were tested at 0.05 level of significance.

Hypothesis 1: There is no significant relationship between male and female students in the one parameter of the Geography items of the qualifying examination for Senior Secondary School 2 students in each of the states in southwest.

Table 1: Correlation statistics for one-parameter indices for male and female on the qualifying geography tests in each of the state.

		Female_bparam_Oyo	Male_bparam_Oyo
Female_bparam_Oyo	Pearson Correlation	1	0.31*
	Sig. (2-tailed)		0.03
Male_bparam_Oyo	Pearson Correlation	0.31*	1
	Sig. (2-tailed)	0.03	
		Female_bparam_Ekiti	Male_bparam_Ekiti
Female_bparam_Ekiti	Pearson Correlation	1	0.74**
	Sig. (2-tailed)		0.00
Male_bparam_Ekiti	Pearson Correlation	0.74**	1
	Sig. (2-tailed)	0.00	

		Female_bparam_Ogun	Male_bparam_Ogun
Female_bparam_Ogun	Pearson Correlation	1	0.13
	Sig. (2-tailed)		0.32
Male_bparam_Ogun	Pearson Correlation	0.13	1
	Sig. (2-tailed)	0.32	

The result in Table 1 reveals that in Oyo and Ekiti, there was significant linear relationship between the difficulty parameters of the female and male in the test items. The Pearson moment-correlation coefficient was significant, though low for Oyo but high for Ekiti ($r = 0.31$, $p = 0.03$ and $r = 0.74$, $p = 0.00$). In contrast, there was no significant relationship between the difficulty parameters of the female and male in the test items in Ogun ($r = 0.13$,

$p = 0.32$). It implies that difficulty indices from the Oyo and Ekiti states are related, while such cannot be said of Ogun state.

Hypothesis 2: There is no significant relationship between male and female students in the two parameters of the Geography items of the qualifying examination for Senior Secondary School 2 students in each of the states in southwest.

Table 2: Correlation statistics for two parameter indices for male and female on the qualifying geography tests in each of the state.

	Female_aparam_Oyo	Female_bparam_Oyo
Male_aparam_Oyo	0.26 (0.07)	
Male_bparam_Oyo		0.57 (0.00*)
	Female_aparam_Ekiti	Female_bparam_Ekiti
Male_aparam_Ekiti	0.10 (0.49)	
Male_bparam_Ekiti		0.21 (0.15)
	Female_aparam_Ogun	Female_bparam_Ogun
Male_aparam_Ogun	0.16 (0.21)	
Male_bparam_Ogun		-0.02 (0.88)

The result in Table 2 reveals that in Oyo, Ekiti and Ogun states, there was no significant linear relationship between the discriminating parameters of female and male students in the test items when two parameters were examined. The Pearson moment-correlation coefficient was not significant

and low across the states (Oyo- $r = 0.26$, $p = 0.07$; Ekiti- $r = 0.10$, $p = 0.49$ and Ogun- $r = 0.16$, $p = 0.21$). There was significant moderate linear relationship between the difficulty parameters of the female and male in the test items in Oyo ($r = 0.57$, $p = 0.00$) while Ekiti and Ogun states remarked low non-

significant relationship between the difficulty parameter of female and male students in the test (Ekiti- $r = 0.21$, $p = 0.15$ and Ogun- $r = -0.02$, $p = 0.88$).

Hypothesis 3: There is no significant relationship between male and female students in the three parameters of the Geography items of the qualifying examination for Senior Secondary School 2 students in each of the states in southwest.

Table 3: Correlation statistics for three parameter indices for male and female on the qualifying geography tests in each of the state

	Female_aparam_Oyo	Female_bparam_Oyo	Female_cparam_Oyo
Male_aparam_Oyo	-0.01(0.96)		
Male_bparam_Oyo		0.28(0.05)	
Male_cparam_Oyo			0.16 (0.26)
	Female_aparam_Ekiti	Female_bparam_Ekiti	Female_cparam_Ekiti
Male_aparam_Ekiti	0.25 (0.08)		
Male_bparam_Ekiti		0.16 (0.28)	
Male_cparam_Ekiti			0.13 (0.37)
	Female_aparam_Ogun	Female_bparam_Ogun	Female_cparam_Ogun
Male_aparam_Ogun	0.30 (0.02)		
Male_bparam_Ogun		0.33 (0.01)	
Male_cparam_Ogun			0.14 (0.27)

The result in Table 3 reveals that in Oyo, Ekiti and Ogun states, there was no significant linear relationship between the guessing parameters of female and male students in the test items when three parameters were examined. The Pearson moment-correlation coefficient was not significant and very low across the states (Oyo- $r = 0.16$, $p = 0.26$; Ekiti- $r = 0.13$, $p = 0.37$ and Ogun- $r = 0.14$, $p = 0.27$). Also, there was no significant linear relationship between the discriminating parameters of the female and male in the test items in Oyo and Ekiti states ($r = -0.01$, $p = 0.96$ and $r = 0.25$, $p = 0.08$) while Ogun state remarked low significant relationship between the discriminating parameter of female and male students in the test ($r = 0.30$, $p = 0.02$). In Oyo and Ekiti states, there was no significant linear

relationship between the difficulty parameters between the female and male students in the test items. The Pearson moment-correlation coefficient was not significant and very low across the states (Oyo- $r = 0.28$, $p = 0.05$ and Ekiti- $r = 0.16$, $p = 0.28$), whereas Ogun state remarked low significant linear positive relationship between the difficulty parameter of female and male students in the test ($r = 0.33$, $p = 0.01$).

The study revealed that difficulty indices from Oyo and Ekiti States are related, while such relationship was not observed for Ogun State. This suggests some level of convergence in the b-parameter calibration between Oyo and Ekiti under the 1PL and 2PL models, likely due to similarities in syllabus coverage and item development, whereas divergence

for Ogun may reflect differences in test length and item construction quality. This finding supports Awopeju and Aramide (2019) who noted that variations in content sampling affect psychometric properties, but is at variance with earlier assumptions of uniform functioning of parallel qualifying examinations in South-West Nigeria. Furthermore, a significant moderate linear relationship was found between female and male difficulty parameters in Oyo State, while Ekiti and Ogun States showed low, non-significant relationships. The Oyo result indicates gender-invariant difficulty, corroborating Adedoyin, Mokobi and Nenty (2018) that well-constructed items exhibit stable b-parameters across gender under IRT. Conversely, the low relationships in Ekiti and Ogun suggest gender-related variations in difficulty estimates which previous studies have attributed to differential familiarity and learning experiences. On discrimination, the Pearson product-moment correlation was generally very low and non-significant across states, with no significant linear relationship between female and male a-parameters in Oyo and Ekiti States under the 2PL and 3PL models, while Ogun State showed a low but significant relationship. The instability in Oyo and Ekiti indicates that the a-parameter is more sample-dependent than the b-parameter, which diverges from the expectation of invariance when model fit holds. The low significant relationship in Ogun suggests relatively more stable discrimination, as found in some previous studies that reported moderate convergence of a-parameters across gender under the 2PL model. The pattern for the c-parameter in the 3PL model also aligns with Jabrayilov, Emons and Sijtsma (2016) that a- and c-parameters require larger samples and stricter fit conditions than the b-parameter to achieve stability.

Conclusion and Recommendations of the study are as follows:

The findings of this study lead to the conclusion that the test-taking abilities of students across the three states are not significantly related. Consequently, items identified as functioning poorly in terms of difficulty, discrimination and guessing parameters under the 1PL, 2PL and 3PL models should be eliminated, and examinations rescored using only psychometrically sound items before results are

released. The study demonstrates that systematic item analysis is critical for enhancing the quality of multiple-choice items and for improving the overall quality of assessment. It is therefore recommended that item analysis be institutionalized as a mandatory stage in the item selection process for state qualifying examinations, prior to final administration, and that such analysis be conducted by experts in Tests and Measurement and test development units within the Ministries of Education. Furthermore, government should ensure the recruitment and retention of qualified Geography teachers in schools where teachers have been transferred, retired or resigned. Substitution with teachers from related subject areas or Parent-Teacher Association (PTA) teachers should be discouraged, as the deployment of subject-specialist teachers is essential for improving students' academic performance.

References

- Abegunde, S. O. (2018). Item analysis of multiple-choice test items in Geography: Implications for effective classroom assessment. *African Journal of Educational Assessment*, 12(1), 45-57.
- Abiri, J. O. (2017). The role of qualifying examinations in predicting performance in external examinations in Nigeria. *Journal of Educational Foundations*, 7(2), 112-124.
- Adedoyin, O. O., Mokobi, T., & Nenty, H. J. (2018). Gender invariance in item difficulty parameters under Item Response Theory: Evidence from senior secondary school Geography test. *International Journal of Learning and Development*, 8(3), 101-115.
- Awopeju, O. A., & Aramide, K. A. (2019). Influence of content sampling and item development procedures on psychometric properties of achievement tests. *International Journal of Educational Research and Development*, 11(2), 88-98.
- Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-

reported outcomes measures. *Clinical Therapeutics*, 36(5), 648-662.

Eleje, L. I., Onah, F. E., & Abanobi, C. C. (2018). Comparative analysis of discrimination indices of multiple-choice Geography items under Classical Test Theory and Item Response Theory models. *Journal of Education and Practice*, 9(27), 92-101.

Embretson, S. E., & Reise, S. P. (2010). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Federal Republic of Nigeria. (2013). *National policy on education* (6th ed.). Nigerian Educational Research and Development Council.

Filani, M. O. (2020). Geography curriculum and national development in Nigerian senior secondary schools. *Journal of Geography Education in Africa*, 3(1), 14-26.

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in terms of

estimation of ability and item parameters. *European Journal of Research Methods and Behavioral Sciences*, 12(2), 109-125.

Ojerinde, D. (2013). Classical test theory versus item response theory: Applications in educational measurement. *Journal of the International Association for Educational Assessment*, 32(1), 1-15.

Ojerinde, D. (2020). The role of mock and qualifying examinations in curbing examination malpractice and mass failure in public examinations. *Nigerian Journal of Educational Assessment*, 5(2), 1-10.

Olatunji, O. S. (2021). Comparability and equating issues in Geography qualifying examinations across Southwest Nigerian states. *Journal of Assessment in Africa*, 9(1), 76-89.

Orheruata, A. (2015). Item parameters of difficulty, discrimination and guessing as correlates of test quality under IRT. *Benin Journal of Educational Studies*, 21(1), 33-44.