

APPLICATION OF THE IRT POLYTOMOUS MODEL IN ESTIMATING DIFFICULTY LEVELS OF 2022–2023 NECO SSCE MATHEMATICS ESSAY ITEMS

ONOJAIFE Jude Onovughakpor

Department of Mathematics, School of Sciences, Delta State College of Education, Mosogar,
Delta State.

onojaifejude4@gmail.com

09065362030

ISHICHERI Ezekiel

IJMB, Delta State College of Education, Mosogar, Delta State.

ishicheri@gmail.com

08063909684

ISAAC Efe Akpomemiyere

Department of Mathematics, School of Sciences, Delta State College of Education, Mosogar,
Delta State.

isaac.ef@descoem.edu.ng; isaacefe2@gmail.com

08034131247, 09156105371

Abstract

This study investigated the difficulty parameters of the National Examinations Council (NECO) Senior Secondary School Certificate Examination (SSCE) Mathematics essay items for 2022 and 2023 and identified the number of items that satisfied the assumptions of an Item Response Theory (IRT) polytomous model. A descriptive survey design was adopted. The population comprised all senior secondary school III students registered for the NECO SSCE in Ethiope West Local Government Area of Delta State, Nigeria. A sample of 1,000 students was drawn from ten public and ten private secondary schools using simple and stratified random sampling techniques. The study instrument consisted of the 2022 and 2023 NECO SSCE Mathematics essay question papers. Item parameters were estimated using Xcalibre 4.1 IRT software developed by Assessment Systems Corporation. Data were analyzed using frequency counts, percentages, means, and standard deviations to answer the research questions, while an independent samples t-test was used to test the hypothesis at the 0.05 significance level. Findings revealed that some Mathematics essay items did not satisfy the statistical assumptions of the IRT model. A significant difference was also found in the difficulty parameters of items that met the IRT requirements across the two examination years.

Keywords: Determination, Difficulty Parameters, Essay Test Items, IRT Polytomous Model, Mathematics.

Introduction

Assessment is a core component of teaching and learning because it helps determine the extent to which educational objectives are achieved, supports identification of students' learning difficulties, and guides instructional improvement and key educational decisions such as grading, promotion, and curriculum planning. However, the effectiveness of these decisions depends on the use of valid and reliable assessment instruments, as weak tools can lead to inaccurate judgments about learners' performance (Akpurunno, 2015). In Nigeria, external

examinations conducted by bodies such as the National Examinations Council play a major role in student certification through the Senior School Certificate Examination (SSCE), yet students' performance in core subjects like Mathematics has remained inconsistent and often unsatisfactory, as reflected in WAEC and NECO results over the years. For instance, WAEC (2023) reported that 79.81% of candidates obtained five credits including English Language and Mathematics in the 2023 WASSCE, compared to 76.36% in 2022. Similarly, NECO (2022) recorded that 60.74% of candidates obtained credits in both English Language and Mathematics, while NECO (2023) showed a slight improvement to 61.60%.

The persistent poor performance in Mathematics has prompted researchers to investigate contributing factors such as teaching methods, teacher effectiveness, and learning environments, although less attention has been given to the quality of examination items themselves (Crosswell, 2021; Gross, 2023). While stakeholders often disagree on whether poor performance results from difficult questions or inadequate student preparation, this debate has led to increased interest in psychometric evaluation of test items. Although earlier studies largely relied on Classical Test Theory, its limitations in handling essay-type responses have led to the growing adoption of Item Response Theory (IRT), which provides more precise item-level analysis of difficulty, discrimination, and learner ability (Hambleton et al., 1991; Han & Hambleton, 2017).

Item Response Theory is particularly relevant to this study because Mathematics essay items involve polytomous scoring, where responses are graded in ordered categories rather than simply right or wrong. Polytomous IRT models such as the Graded Response Model, Nominal Response Model, Partial Credit Model, and Rating Scale Model are designed for such complex scoring structures and allow for more accurate estimation of item and ability parameters (Masters & Wright, 2017). This study specifically adopts the Partial Credit Model because it is suitable for evaluating essay-type Mathematics items scored in steps, enabling the estimation of item difficulty across score categories and the assessment of whether NECO Mathematics essay items for 2022 and 2023 meet acceptable psychometric standards.

Previous empirical studies have examined the application of Classical Test Theory (CTT) and Item Response Theory (IRT) in educational assessment; however, the present study differs significantly from earlier works. Omorogiuwa (2019) focused mainly on comparing CTT and IRT procedures. Similarly, Osunde and Ethe (2003) analyzed the difficulty indices of multiple-choice Mathematics items using traditional item analysis methods. Other studies, such as those by Adedoyin, Nenty, and Chilisa (2018), as well as Hernandez (2019), investigated parameter invariance and item characteristics using generalized or aptitude-test data. In contrast, the present study specifically examines NECO SSCE Mathematics essay questions for 2022 and 2023 using IRT polytomous models. The study is unique because it focuses on essay-type items, applies advanced IRT procedures to real examination data, and evaluates whether the items satisfy acceptable psychometric standards. This provides a more precise assessment of item difficulty and measurement quality in large-scale Mathematics examinations.

Statement of the Problem

Formal examinations are designed to measure students' knowledge, skills, and academic achievement, yet persistent poor performance in Mathematics in Nigeria continues to raise concern. NECO records show that in 2022 only 60.74% of candidates obtained five credits including Mathematics and English Language, while 61.60% achieved the same benchmark in 2023, indicating that many students still perform below expectations despite the importance of Mathematics for national development. This trend suggests a continuing gap between expected and actual student achievement.

Although poor performance has often been linked to factors such as teaching methods, student attitudes, and learning environments, less attention has been paid to the quality of examination items themselves. Concerns about the difficulty level of NECO Mathematics questions highlight the need for empirical investigation of their psychometric properties. Since Classical Test Theory provides limited item-level information, this study adopts Item Response Theory to examine the difficulty parameters of NECO SSCE Mathematics essay items for 2022 and 2023, with the aim of determining whether they meet acceptable measurement standards for reliable educational decision-making.

Objectives of the Study

The objective of this study is to determine the difficulty parameters of the National Examination Council, Senior Secondary School Certificate Examination Essay Test Items in Mathematics for the year 2022 and 2023. Specifically, the study is aimed at determining:

1. the difficulty parameters of the June/July NECO SSCE Mathematics Essay Test Items for year 2022 and 2023 .
2. if the difficulty parameter satisfy the Item Response Theory statistical conditions for year 2022 and 2023.
3. if there is difference between 2022 and 2023 difficulty parameters for the June/July NECO SSCE Mathematics Essay Test Items that satisfy the Item Response Theory statistical conditions.

Research Questions

The following research questions guided the study:

1. What are the difficulty parameters of the June/July NECO SSCE Mathematics Essay Test Items for year 2022 and 2023?
2. Does the difficulty parameter satisfy the item response theory statistical conditions for year 2022 and 2023?
3. Are there differences between 2022 and 2023 difficulty parameters for the June/July NECO SSCE Mathematics essay test items that satisfy the item response theory statistical conditions?

Hypothesis

The following hypotheses was tested at 0.05 significant level:

1. There is no significant difference between 2022 and 2023 difficulty parameters for the June/July NECO SSCE Mathematics Essay Test Items that satisfy the Item Response Theory statistical conditions.

Methodology

This study adopted a descriptive survey research design to examine the characteristics of Mathematics examination items without manipulating any variables, focusing on the systematic collection and analysis of data within a defined population. The population comprised 3,842 Senior Secondary School III (SSS III) Mathematics students who registered for the 2024 May/June Senior School Certificate Examination in Sapele Local Government Area, from which a sample of 1,000 students was selected using stratified and simple random sampling techniques across 20 schools (10 public and 10 private) drawn from rural and urban areas, with 500 students each from public and private schools. The instrument used was the National Examinations Council Mathematics Essay Question Papers for 2022 and 2023, whose reliability coefficients were 0.89 and 0.82 respectively, indicating high internal consistency.

The scripts were scored by four trained Mathematics teachers using the official NECO marking guide, with structured scoring rubrics applied to ensure consistency, after which the responses were analyzed using Xcalibre software for Item Response Theory (IRT) calibration to estimate item difficulty parameters (b). Items within the acceptable range of $-1 \leq b \leq +1$ were retained, while others were excluded, in line with modern psychometric standards (De Ayala, 2022). Data analysis involved descriptive statistics such as mean, standard deviation, frequencies, and percentages, while an independent samples t-test was used to compare item difficulty parameters between 2022 and 2023 to determine if significant differences existed between the two examination years.

Results

Research Question One: What are the difficulty parameters of the June/July NECO SSCE Mathematics Essay Test Items for year 2022 and 2023?

Table 1: X-calibre of Difficulty Parameters of June/July NECO SSCE Mathematics Essay Test Items for 2022 and 2023.

Year(s)	2022		2023	
Items	b		b	
01	-	**	-	**
02	-	**	-0.728	*
03	-0.118	*	-0.641	*
04	-0.879	*	0.909	*
05	-0.632	**	-	**

06	0.893	“*	-	**
07	-0.871	“*	0.973	*
08	-0.863	“*	0.954	*
09	-	**	-	**
10	0.955	*	-	**
11	-0.641	**	0.970	*
12	0.973	*	-0.980	*
		9		7
Total no. of items = (16)				
Percentage of items	56.25%		43.75%	
Mean	-0.243		0.208	
SD	0.899		0.933	
MIN	-0.879		-0.980	
Max	0.973		0.973	

Total number of items = 16; Percentage of items (b 2022) = 56.25%, b 2023 = 43.75%

* = Included Item(s), ** = Removed Item(s), b = Difficulty parameters, SD = Standard deviation, MIN = Minimum Value of the parameter, Max = Maximum Value of the parameter. Results from table 1 show the difficulty parameters of the 2022 and 2023 National Examinations Council SSCE Mathematics Essay Test items based on Xcalibre 4.1 IRT calibration. The calibrated items had difficulty values ranging from -0.879 to 0.973 across both years, indicating a mix of easier and more difficult items. The mean difficulty values were -0.243 (2022) and 0.208 (2023), with standard deviations of 0.899 and 0.933 respectively, suggesting moderate variability in item difficulty across the two examination years.

Research Question Two: Does the June/July NECO SSCE mathematics essay Test Items difficulty parameter satisfy the item response theory statistical conditions for year 2022 and 2023?

Results presented in Table 1 above revealed the number of the June/July NECO SSCE Mathematics Essay test items that satisfied the Item Response Theory statistical conditions for year 2022 and 2023 difficulty parameters. Results in the Table showed that 9 and 7 items satisfy the Item Response Theory statistical conditions which ranges from -0.879 to 0.973 for 2022 and -0.980 to 0.973 for 2023. These difficulty parameters falls within the item response theory statistical conditions of -1 to +1 for a test item to be acceptable.

Hypothesis One

There is no significant difference between 2022 and 2023 difficulty parameters for the June/July NECO SSCE Mathematics Essay Test Items that satisfy the Item Response Theory statistical conditions.

Table 2: Independent Sampled t- Test of June/July NECO SSCE Mathematics Essay Test Items Difficulty Parameters for Year 2022 and 2023

		Years	N	Mean	Std. Deviation	t _{critical}	Df	P-Value
Difficulty parameter for year		2022	9	0.0000	0.60187			
		2023	7	0.3130	0.61220	-1.941	14	0.06

Alpha (α) = 0.05 Level of Significance

The independent samples t-test results showed a t-value of -1.941 with 14 degrees of freedom and a p-value of 0.06 for the 2022 and 2023 National Examinations Council SSCE Mathematics Essay Test item difficulty parameters. Since the p-value was greater than the 0.05 level of significance, the null hypothesis was accepted. This indicates that there was no statistically significant difference between the difficulty parameters of the 2022 and 2023 examination items under the Item Response Theory framework.

Discussion of Findings

The results of research question one showed that the 2022 and 2023 National Examinations Council Mathematics Essay Test items generally fell within the acceptable IRT difficulty range ($-1 \leq b \leq +1$), indicating that the items had moderate difficulty suitable for differentiating students' abilities. While the 2022 items were slightly easier on average (mean = -0.243), the 2023 items were slightly more difficult (mean = 0.208), but both sets were neither too easy nor too difficult, aligning with psychometric expectations for effective test items (De Ayala, 2022). Findings for research question two revealed that 9 items (2022) and 7 items (2023) met the acceptable IRT difficulty criteria, suggesting that most of the items were appropriate for measuring students' Mathematics ability. However, some items were excluded because they did not meet the required psychometric standards, implying they were either too easy or too difficult for effective assessment. This supports the view of Embretson and Reise (2013) that well-fitting IRT items provide more valid and reliable measurement of examinee performance. The hypothesis testing further showed that there was no significant difference in the difficulty parameters of the 2022 and 2023 test items, indicating stability in item construction across both years. This suggests that the examination maintained a consistent standard of difficulty over time, as confirmed by the t-test result ($p > 0.05$). This finding aligns with Hambleton, Swaminathan, and Rogers (1991), who emphasized that consistency in item parameters reflects quality and stability in test development practices.

Conclusion

The study concluded that not all National Examinations Council SSCE Mathematics Essay Test items for 2022 and 2023 met the required Item Response Theory (IRT) statistical criteria, indicating that some items had acceptable psychometric properties while others did not. It further revealed that certain items failed to function optimally under the IRT framework.

However, there was no statistically significant difference in the difficulty parameters of the 2022 and 2023 test items, suggesting consistency and stability in item construction across both examination years.

Recommendations

From the findings and conclusion of this study, the researchers therefore make the following recommendations that:

1. Examination bodies such as the National Examinations Council should ensure rigorous validation of test items to match examinees' ability levels, since some Mathematics essay items did not meet Item Response Theory (IRT) statistical requirements.
2. Modern psychometric approaches, particularly Item Response Theory (IRT), should be adopted for item selection and validation because they are suitable for both dichotomous and polytomous test items.
3. Staff involved in test development and assessment should be trained in contemporary psychometric and technological methods of item analysis to ensure proper validation of assessment instruments before large-scale use.

References

- Adedoyin, O. O., Nenty, H. J. & Chilisa (2018). Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. *Educational Research and Review*, 3(2), 83-93. <http://www.academicjournals.org/ERR>
- Akporuno, P. A. (2015). *Trend Analysis of past NECO SSCE Mathematics essay test item difficulty and discriminating parameters using Item Response Theory Polytomous Model*. Ph.D. Thesis submitted to School of Postgraduate Studies, University of Benin, Nigeria.
- Crosswell, M. J. (2021). *Research studies in public examination*. Guilford, Associated Examining Board.
- De Ayala, R. J. (2022). *The theory and practice of item response theory* (2nd ed.). The Guilford Press.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press.
- Gross, J. (2023). *Exposing the cheat sheet with the students' aid*. Retrieved September 30th, 2023 from <http://www.Nytimes.com>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.
- Han, K. T., & Hambleton, R. K. (2017). *User's manual for item response theory programs*. Assessment Systems Corporation.
- Masters, A. & Wright, B. D (2017). *A history of social science measurement. Educational measurement: issues and practice*. Lawrence Erlbaum Associates, publishers.

- National Examinations Council. (2022). *Release of 2022 SSCE internal examination results*. NECO. <https://neco.gov.ng>
- National Examinations Council. (2023). *Release of 2023 SSCE internal examination results*. NECO. <https://neco.gov.ng>
- Omorogiuwa, O. K. (2019). *An empirical comparison of the Classical test theory and Item Response Theory in the selection of Physics achievement test items*. Ph.D. thesis submitted to School of Postgraduate Studies, University of Benin, Benin City Nigeria.
- Osunde, A. U., & Ethe, N. (2003). Analysis of the difficulty index of SSCE Mathematics multiple choice test items (1998-2002), *Journal of Research in Counselling*, 4(3), 34-45.
- Umobong, M. E. (2019). *Principles of educational measurement and evaluation*. Gloryland Publishers.
- West African Examinations Council. (2023). *Release of 2023 WASSCE results statistics*. WAEC. <https://waecnigeria.org.ng>