www.ijarp.com ISSN 2456-9992 Page: 01-08

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 04

DEVELOPMENT AND STANDARDIZATION OF A POLYTOMOUS MATHEMATICS ACHIEVEMENT TEST USING ITEM RESPONSE THEORY IN THE FEDERAL CAPITAL TERRITORY, ABUJA, NIGERIA.

*[1]Ameh Friday Attai, [2]Dr. Agi C.I [3] Prof. Obinne A.D.E, [4]Prof. Emaikwu S. O.

Department of Guidance and counselling Joseph sarwuan Tarkaa University, Makurdi

Article Received: 21 February 2026, Article Revised: 12 March 2026, Published on: 01 April 2026

***Corresponding Author: Ameh Friday Attai**

Department of Guidance and counselling Joseph sarwuan Tarkaa university, Makurdi

DOI: https://doi-doi.org/101555/ijarp.6784

ABSTRACT

The increasing use of constructed-response and graded assessment formats in Mathematics necessitates robust psychometric frameworks capable of capturing varying levels of student performance. This study investigated the application of Item Response Theory in the development, calibration, and validation of a polytomous Mathematics Achievement Test administered to senior secondary school students in the Federal Capital Territory (FCT), Abuja, Nigeria. Guided by two research questions, the study adopted a descriptive survey research design. A total of 381 SSII students were selected from a population of 54,139 using the Krejcie and Morgan sample size determination procedure. Data were collected using a polytomously scored Mathematics Achievement Test and analyzed using appropriate polytomous IRT models implemented through the mirt package in R statistical software. Results indicated that the test measured multiple latent mathematical proficiencies, with category threshold parameters demonstrating orderly progression along the ability continuum. The majority of items exhibited satisfactory slope parameters, suggesting that the graded response categories effectively differentiated students across varying ability levels. The test also demonstrated adequate reliability and provided substantial information across a broad range of student abilities. These findings confirm the suitability of polytomous IRT models for evaluating Mathematics achievement tests that incorporate partial credit scoring. It was recommended that test developers pay close attention to category functioning, ensure well-defined scoring rubrics, and conduct routine psychometric evaluations to maintain the

validity, fairness, and interpretability of polytomous assessment instruments. Continuous professional development for Mathematics educators and test developers in polytomous item construction and IRT-based analysis was also advocated.

KEYWORDS: *Polytomous Item Response Theory, Mathematics Achievement Test, Partial Credit, Validation, Calibration, Category Thresholds.*

INTRODUCTION:

Background of the Study

Assessment remains a fundamental component of the school system, serving as a primary means of evaluating learners' understanding and mastery of academic content. In Mathematics education, effective assessment goes beyond determining whether an answer is right or wrong; it involves examining the quality, structure, and progression of students' reasoning. When assessment practices are properly designed and implemented, they yield accurate and fair judgments of learners' competencies, thereby supporting informed educational decisions.

In Nigeria, challenges persist in the accurate measurement of students' Mathematics achievement, particularly in high-stakes examinations that influence promotion, certification, and admission into higher institutions. These challenges have raised concerns about the adequacy of conventional scoring approaches in capturing the depth of students' understanding and problem-solving processes. As a result, greater attention has been directed toward assessment formats that allow for partial demonstration of knowledge and varying levels of performance (Alves, 2022).

Polytomous test items are designed to capture multiple levels of response quality by awarding scores across ordered categories rather than a single outcome. In Mathematics achievement testing, such items typically reflect stepwise problem-solving procedures, where students may earn partial credit for correct methods, intermediate steps, or logical reasoning even when the final answer is incomplete or incorrect. This approach provides richer information about students' cognitive processes and learning outcomes (Gracia & Rodriguese, 2016). The quality of polytomous test items is determined by their item parametric characteristics, which describe how well the items function in measuring the intended construct. These characteristics include item difficulty, discrimination, and category threshold parameters (Chen & Li, 2019). Item difficulty reflects the location of response categories along the ability continuum, while discrimination indicates how effectively an item differentiates

among students with varying levels of mathematical proficiency. Threshold parameters describe the points at which students are likely to transition from one score category to another. Proper estimation of these parameters is essential for ensuring that polytomous items yield meaningful and interpretable scores (Siqueira, 2022).

Item Response Theory (IRT) provides a robust framework for estimating the parameters of polytomous items (Zampetakis, 2019). Polytomous IRT models, such as the Partial Credit Model and the Graded Response Model, enable test developers to account for the ordered nature of response categories and to model the probability of students attaining different score levels based on their underlying abilities. Through item calibration, these models ensure that test scores accurately reflect students' mathematical competencies and that comparisons across examinees are fair and reliable (Chen & Li, 2019).

Mathematics achievement testing plays a critical role in Nigeria's educational system, particularly at the senior secondary school level where performance in certificate examinations determines access to higher education and career opportunities (Effiom, 2021). Despite the importance of Mathematics, evidence from public examination bodies indicates fluctuating achievement trends among students in the Federal Capital Territory, Abuja. These fluctuations have raised concerns among educators, policymakers, and stakeholders regarding the effectiveness of current assessment practices (Bello & Sanusi, 2023).

Although previous studies have explored factors influencing students' achievement in Mathematics, limited attention has been given to the role of polytomous item construction, scoring procedures, and calibration methods in improving the accuracy of achievement measurement. Given the capacity of polytomous assessment to capture varying levels of understanding and reasoning, there is a need to examine its application within a rigorous measurement framework.This study therefore assessed the use of Item Response Theory in the development and standardization of polytomous responses in Mathematics achievement tests in the Federal Capital Territory, Abuja, with a view to improving the validity, reliability, and interpretability of students' achievement scores.

Statement of the Problem

Assessment plays a pivotal role in secondary education, particularly in Mathematics, where test results inform critical decisions such as promotion, placement, certification, and scholarship awards. For these decisions to be defensible, Mathematics achievement tests must yield valid, reliable, and precise estimates of students' ability levels. Accurate measurement is especially important in Mathematics, a subject that demands multistep reasoning,

conceptual understanding, and partial problem-solving, all of which require assessment approaches capable of capturing varying degrees of proficiency. In the Federal Capital Territory (FCT), Abuja, analyses of recent mock examinations and school-based assessments consistently reveal low levels of student achievement in Mathematics. These persistent outcomes raise serious concerns about the adequacy of existing Mathematics achievement tests to accurately reflect students' true levels of mathematical competence. The limited variability and weak performance patterns observed over time suggest that current assessment practices may not sufficiently discriminate among students with differing levels of ability.

A major source of this problem may lie in the development and scoring of Mathematics achievement tests. Many teacher-constructed instruments rely on traditional test development methods that inadequately represent the complexity of mathematical tasks. When tests fail to incorporate systematic validation procedures and appropriate scoring frameworks, they may overlook partial knowledge, misrepresent item difficulty, and produce scores that are weak indicators of actual achievement. Consequently, the validity and reliability of test scores are compromised, limiting their usefulness for instructional and policy-related decision-making. Although numerous interventions—including curriculum reforms, improved instructional strategies, professional development for teachers, and technology-enhanced learning—have been implemented to address low Mathematics achievement, student performance remains unsatisfactory. This persistence suggests that instructional improvements alone may be insufficient without corresponding advances in the quality of assessment instruments used to measure learning outcomes.

Polytomous scoring models offer a more realistic representation of students' performance in Mathematics by allowing multiple score categories that reflect varying levels of understanding and partial solution processes. Item Response Theory (IRT), particularly polytomous IRT models, provides a robust framework for the development, calibration, and standardization of such assessment instruments by simultaneously accounting for item characteristics and examinee ability. However, the application of polytomous IRT models in the development of Mathematics achievement tests within the FCT remains limited. Against this backdrop, the problem addressed in this study is the extent to which the application of Item Response Theory can enhance the development, validation, and standardization of **polytomous Mathematics Achievement Test scores** among senior secondary school students in the Federal Capital Territory, Abuja, Nigeria.

Objectives of the Study

The objective of this study is the development and standardization of polytomous responses of Mathematics Achievement Test in the Federal Capital Territory Abuja. Specifically, the study used IRT to:

1. Determine the item parameters of the polytomous Mathematics Achievement Test in the Federal Capital Territory Abuja
2. Determine the reliability coefficient of the polytomous Mathematics Achievement Test in the Federal Capital Territory Abuja

Research Questions

The following research questions guided this study

1. What are the item parameters of the polytomous Mathematics Achievement Test in the Federal Capital Territory Abuja?
2. What is the reliability coefficient of the polytomous Mathematics Achievement Test in the Federal Capital Territory Abuja?

METHODS

The study adopted descriptive survey research design, 54139 students of Senior Secondary two made up of the population. While 381 Senior Secondary School two students were sampled using Krejcie and Morgan table for determining sample size from a given population. Polytomous Mathematics Achievement Test (PMAT) was used for data collection. The reliability coefficient obtained for the research instrument was 0.81 for the Polytomous Mathematics Achievement Test (PMAT). Cronbach Alpha was used to determine the internal consistency of the Polytomous Mathematics Achievement Test (PMAT). Research questions were answered using R statistical software with the mirt package of Item Response Theory

RESULTS

To answer research question 1, the analysis was conducted with the mirt package of R language and environment for statistical computing. The results are presented in Table 1.

Table 1 Item Parameter Estimate of Polytomous Mathematics Item.

Item	A	d0	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11
IT1	0.00	0.00	1.18	1.42	1.10	0.69	0.82	1.03	0.76	0.73	**-0.41**	-1.10	-1.61
IT2	**-0.08**	0.00	0.92	1.30	1.12	1.25	1.25	1.21	1.07	0.90	-0.98	NA	NA
IT3	0.03	0.00	0.70	2.13	2.50	2.70	2.68	2.91	2.44	2.50	2.58	-1.09	**-0.40**
IT4	0.01	0.00	1.39	2.71	3.18	3.20	3.18	2.95	2.95	2.57	2.44	NA	NA
IT5	0.05	0.00	0.98	1.55	1.83	1.90	1.88	1.74	1.51	1.39	1.27	1.15	1.02
IT6	0.06	0.00	1.12	1.62	1.87	2.03	2.11	2.03	1.93	1.84	1.76	1.60	1.41
IT7	**-0.01**	0.00	1.22	1.35	1.49	1.65	1.72	1.69	1.61	1.50	1.37	NA	NA
IT8	0.07	0.00	1.34	1.77	2.01	2.19	2.27	2.15	2.01	1.90	1.78	1.67	1.54
IT9	0.03	0.00	1.10	1.47	1.62	1.70	1.76	1.72	1.64	1.50	1.37	NA	NA
IT10	0.06	0.00	1.17	1.58	1.92	2.09	2.23	2.11	1.99	1.86	1.74	1.61	1.48

Table 2. Estimated Reliability Coefficients for the Mathematics Polytomous Test.

Reliability
0.5

FINDINGS

The findings in table 1 indicated that **IT1** has a discrimination parameter of a = 0.00, **IT2** has a discrimination parameter of a = **-0.08**, **IT3** has a very low discrimination parameter (a = 0.03), **IT4** has a discrimination parameter of a = 0.01, **IT5** has a low discrimination parameter (a = 0.05), **IT6** has a discrimination parameter of a = 0.06, **IT7** has a very slight negative discrimination parameter (a = **-0.01**), **IT8** has a discrimination parameter of a = 0.07, **IT9** has a very low discrimination parameter (a = 0.03), **IT10** has a discrimination parameter of a = 0.06. This result agreed with the results of Schmucker and Moore (2025)**,** they analyzed over 7,000 multiple-choice questions across STEM subjects, this study found that approximately 11.1% of items had low discrimination parameters. The presence of item-writing flaws, such as negative wording or implausible distractors, was significantly associated with reduced discrimination, particularly in life and physical sciences.

Result in table 2 shows that the reliability coefficient of the polytomous Mathematics Achievement Test in the Federal Capital Territory Abuja was 0.5. The result agreed with the

findings of Adebule **(2009)** who examined the reliability of multiple-choice and true/false test items in a Mathematics Achievement Test administered to 500 senior secondary students in Ondo State, Nigeria. The findings revealed reliability coefficients of 0.35 for multiple-choice and 0.25 for true/false formats, indicating moderate reliability levels. Effiom **(2021)** utilized Item Response Theory, this research assessed the Differential Item Functioning of a Mathematics Achievement Test among 1,751 SS2 students in Cross River State. The study found that 6% of the items exhibited differential functioning, suggesting potential reliability concerns in the test items.

CONCLUSION

Based on the findings of this study, the following conclusions were drawn: Most test items showed **negative or even positive discrimination parameters**, indicating they differentiate effectively between high- and low-performing students. This enhances the **validity and fairness** of the assessment. Furthermore, with a reliability coefficient of 0.5**,** the test demonstrates moderate consistency in measuring students' true abilities, suggesting some level of stability in the assessment results. To enhance the effectiveness of the test, items with poor discrimination should be revised or eliminated**,** and more reliable constructs should be developed. These steps are essential to improving the validity, reliability, and fairness of the Mathematics Achievement Test for future use.

RECOMMENDATIONS

Efforts should be made to organize workshops and training sessions for Mathematics educators and assessment designers on item writing, test construction, and psychometric evaluation. This will enhance their ability to create and evaluate high-quality, reliable polytomous questions

REFERENCES

1. Adebule, S. O. (2009). Reliability of senior secondary school certificate Mathematics examination in Nigeria. *Pakistan Journal of Social Sciences*, 6(1), 31–35.
2. Alves, J. O. (2022). Assessment of Polytomous and Dichotomous Mathematics Achievement test Among Junior Secondary School Students in Rivers State. Unpublished M. Ed Thesis University of Maiduguri.
3. Bello, A. A. & Sanusi, S. L. (2023). Determinants of Students' Academic Performance in Mathematics Among Senior Secondary School Students in Federal Capital Territory Abuja, Nigeria. 2(1)222-231.

4. Bulala, T. M. (2023). *An evaluation of test item discrimination in secondary school assessments using the Kuder-Richardson method*. Journal of Educational Research and Review, 11(2), 25–34.
5. Chen, Y., & Li, H. (2019). An Examination of Item Response Theory Models in Validating Mathematics Achievement Tests: Evidence from Taiwanese Schools. Unpublished Masters' thesis University of Taiwan
6. Effiom, B. E. (2021). Gender-based differential item functioning in Mathematics achievement tests: An application of item response theory. *Journal of Educational Measurement and Evaluation*, 13(2), 22–37.
7. Ejakpovi, S.U. & Uveruveh, F.O. (2014).The Role of Mathematical Graph Board in the Teaching and Learning of Mathematics in Senior Secondary Schools in Delta State.*Abacus, the Journal of the Mathematical Association of Nigeria 39*(1), 14-24.
8. García, M., & Rodríguez, P (2016). Validation of Mathematics Achievement Tests Using Item Response Theory: A Study in Spanish Schools. Unpublished PhD thesis. International University of Andalusia
9. Siqueira, M. D. (2022). Validation of dichotomous Mathematics Achievement Test based on many – faceted Rasch model. Unpublished PhD Thesis , University of Ghana
10. Schmucker, D. G., & Moore, E. R. (2025). Discrimination parameters in STEM assessments: A review of 7,000 multiple-choice questions. *Journal of STEM Education Research*, 4(1), 1–15.
11. Zampetakis, M.A. (2019) Using Item Response Theory to investigate Students' performance in Mathematics Achievement Test. M.Ed Unpublished thesis. University of Calabar.