Rima International Journal of Education (RIJE)
ISSN: 2756 – 6749 (p); 3141-2033 (e)
Volume 5, Issue 4, June. 2026; Page: 86—97 DOI: https://doi.org/10.65760/rijessu.v5.4.6

Assessment of the Quality of Secondary School Mathematics Multiple-Choice Questions Using Difficulty and Discrimination Indices among Selected Schools in Ondo State, Nigeria

[*1]Alice Olayemi Asaolu and [2]Kayode Rapheal Daramola

[*1]University of Fortune, Igbotako, Nigeria. **Email:** Aliceolayemi4good@gmail.com or ao21@uf-edu.org

[2]Department of Science Education, Faculty of Education, Adeyemi Federal University of Education, Ondo, Nigeria. **Email:** daramolakr@afued.edu.ng, **ORCID**: 0000-0002-1580-5210

Abstract

This study assessed the quality of secondary school mathematics multiple-choice questions (MCQs) in selected schools in Ondo State, Nigeria, using difficulty and discrimination indices. A total population of 5,831 senior secondary students was considered, and a sample of 374 students was selected from Akoko North-West and Okitipupa local government areas. The Mathematics MCQ test, covering algebra, geometry, trigonometry, statistics, probability, and mensuration, was analyzed to determine how challenging the items were and how well they differentiated between high- and low-performing students. Results showed that the majority of items had moderate difficulty and good discrimination, indicating their effectiveness in assessing students' knowledge. However, some items were either too easy, too difficult, or poorly discriminating, suggesting the need for revision. Hypothesis testing revealed significant differences in performance between high- and low-achieving students, confirming that most items effectively distinguish student ability. The study concludes that while the test items are partially effective, systematic item analysis and regular review are essential. Recommendations include revising poorly performing items, enhancing teacher training in test construction, and aligning assessments with the curriculum to improve fairness and validity.

Keywords: Mathematics assessment, multiple-choice questions, difficulty index, discrimination index, item analysis, Ondo State

Cite this as: Asaolu, A. O., & Daramola, K. R. (2026). Assessment of the Quality of Secondary School Mathematics Multiple-Choice Questions Using Difficulty and Discrimination Indices among Selected Schools in Ondo State, Nigeria. *Rima International Journal of Education, 5*(4), 68—97. DOI: https://doi.org/10.65760/rijessu.v5.4.6

Introduction

Assessment is an essential part of teaching and learning because it provides feedback on student understanding, guides instructional planning, and influences academic decisions. In secondary school mathematics, multiple-choice questions (MCQs) are widely used because they allow teachers to efficiently assess a large body of content and objectively score student responses. However, the quality of MCQs is not always guaranteed, as poorly constructed questions can lead to misleading conclusions about what students know and can do. This is particularly concerning in settings where test scores influence progression, placement, and curriculum decisions. Two important ways of evaluating the quality of MCQs are the difficulty index and the discrimination index variables that reflect how challenging each item is and how well it distinguishes between high-performing and low-performing students. The difficulty index shows the proportion of students who answered an item correctly, while the discrimination index reflects how effectively an item differentiates between students with stronger versus weaker overall performance (Rezigalla *et al.,* 2024). Together, these indices help teachers and researchers judge whether items are fair, appropriately challenging, and useful for accurate assessment.

Recent literature highlights the need for systematic evaluation of MCQ quality in mathematics assessments. Studies conducted in various educational settings have shown that many MCQ items fail to meet acceptable psychometric standards, often exhibiting extreme difficulty levels or weak discrimination power (Ayodeji et al., 2024; Cobbinah & Ntumi, 2022). For example, item analysis research in Nigerian secondary schools found significant variation in the quality of mathematics test items, with some items being too easy and others so difficult that they did not contribute meaningfully to assessing student ability (Jimoh et al., 2022). Similarly, research in West African educational contexts has emphasized that weak distractors, ambiguous wording, and inappropriate cognitive demands can compromise the effectiveness of MCQs (Adebowale & Oluwole, 2023). These findings suggest that although MCQs are widely used, their psychometric properties especially difficulty and discrimination indices are not always routinely examined, leading to questions about the validity and fairness of mathematics tests administered to students.

Despite these insights, there remains a limited focus on the quality of mathematics MCQs in Ondo State secondary schools. In many schools within the state, teachers design their own tests without adequate item evaluation, and administrators may rely on test scores without knowing whether individual items are truly measuring student knowledge. This practice raises concerns about fairness, instructional feedback, and the overall usefulness of mathematics assessments in the state. Therefore, there is a need to examine the quality of mathematics MCQ items systematically, using established item analysis measures to ensure that test items are both appropriately challenging and capable of distinguishing between different levels of student achievement.

The purpose of this study is to assess the quality of secondary school mathematics multiple-choice questions in selected schools in Ondo State, Nigeria, by computing and interpreting their difficulty and discrimination indices. Specifically, the study seeks to determine how challenging the items are for students, how well they discriminate between high- and low-performing students, and which items meet acceptable psychometric standards or require revision.

To guide the investigation, the following research questions are posed:

I. What are the difficulty levels of secondary school mathematics multiple-choice questions in selected schools in Ondo State?

II. To what extent do the mathematics multiple-choice items discriminate between high-performing and low-performing students?

III. Which mathematics MCQ items meet acceptable quality standards, and which items need revision?

Correspondingly, the study tests the following research hypotheses at the 0.05 level of significance:

I. There is no significant difference in performance between high-achieving and low-achieving students on the mathematics multiple-choice items.

II. Mathematics multiple-choice items do not significantly discriminate between high-performing and low-performing students in selected secondary schools.

By examining both difficulty and discrimination indices, this study contributes to improving the quality of mathematics assessments, supports better instructional decisions, and enhances the fairness and interpretability of test scores in secondary school settings.

Methodology

This study adopts a descriptive survey research design, which is appropriate because it allows for the systematic evaluation and description of the quality of secondary school mathematics multiple-choice test items without manipulating any variables. The study is non-experimental in nature, focusing on existing test items and students' responses.

The population of the study consists of all senior secondary school students in selected schools from two local government areas in Ondo State, Nigeria: Akoko North-West Local Government Area (North Senatorial District) and Okitipupa Local Government Area (South Senatorial District). The total population is 5,831 students. A sample of 374 students was selected using the stratified random sampling technique to ensure proportional representation across schools, locations, and class levels.

The instrument used for data collection is the Mathematics Multiple-Choice Achievement Test (MMCA). The test consists of 40 carefully selected multiple-choice items drawn from key areas of the senior secondary school mathematics curriculum, including algebra, geometry, trigonometry, statistics, probability, and mensuration. Content validity was ensured by aligning all items with the approved senior secondary school mathematics curriculum, and experienced mathematics teachers alongside a test construction specialist reviewed the items for clarity, relevance, accuracy, and appropriateness.

A pilot test was conducted using 40 students drawn from schools outside the selected study areas to avoid contamination of the main sample. These students were not included in the final study. The pilot data were analyzed using the Kuder-Richardson Formula 20 (KR-20) to determine internal consistency, yielding a reliability coefficient of 0.70, which indicates that the instrument is adequately reliable for measuring students' mathematics achievement. Based on the pilot results, minor revisions were made to some items that showed ambiguity or poor functioning, particularly those with unclear distractors or extreme difficulty levels.

The final test was administered under standardized examination conditions in classroom settings. All students received uniform instructions, and adequate supervision was provided to ensure consistency and prevent examination malpractice. The test duration was 60 minutes, which was sufficient for students to attempt all 40 items.

Each correctly answered item was awarded one mark, while incorrect responses attracted zero marks, with no penalty for guessing. For item analysis, students were ranked according to their total scores and divided into performance group.

Results

Research Question 1: *What are the difficulty levels of secondary school mathematics multiple-choice questions in selected schools in Ondo State?*

The difficulty index (P) of each item was calculated using the formula:

P=Number of students who answered correctly / Total number of students

Table 1; Difficulty Levels of Mathematics Multiple-Choice Questions (N = 374)

Item No.	Number Correct	Difficulty Index (P)	Difficulty Level
1	220	0.59	Moderate
2	280	0.75	Easy
3	90	0.24	Difficult
4	310	0.83	Very Easy
5	150	0.40	Difficult

Table 1 presents a sample of five mathematics multiple-choice items selected from the 40 items analysed items administered to 374 students. The difficulty index measures the proportion of students who answered each item correctly, with values ranging from 0 (very difficult) to 1 (very easy).

Item 1 (P = 0.59, Moderate)**:** About 59% of students answered correctly, indicating the item is of moderate difficulty. This is ideal for assessing students' understanding because it neither too easy nor too hard.

Item 2 (P = 0.75, Easy)**:** 75% of students answered correctly, making it relatively easy. While it can build confidence, items at this level may not effectively challenge students or differentiate performance levels.

Item 3 (P = 0.24, Difficult)**:** Only 24% of students answered correctly. This indicates a difficult item, which may challenge most students. If too many items are like this, the test could become discouraging or fail to measure typical student performance accurately.

Item 4 (P = 0.83, Very Easy)**:** 83% of students answered correctly, making it very easy. Items this easy provide little information about differences in students' abilities and may not contribute effectively to assessing knowledge.

Item 5 (P = 0.40, Difficult)**:** 40% of students answered correctly. This is considered difficult but still within a reasonable range for identifying students with stronger understanding.

Research Question 2: *How well do the mathematics multiple-choice items discriminate between high-performing and low-performing students?*

The discrimination index (D) was calculated using:

$$D = U - L / N$$

Where:

I. **U** = number of students in the upper group who answered correctly

II. **L** = number of students in the lower group who answered correctly

III. **N** = number of students in one group

Table 2: Discrimination Indices of MCQ Items

Item No.	Upper Group Correct (U)	Lower Group Correct (L)	D Value	Item Quality
1	48	18	0.60	Very Good
2	50	30	0.40	Very Good
3	30	28	0.04	Poor
4	40	32	0.10	Poor
5	38	20	0.36	Good

Table 2 presents the discrimination indices (D) of five selected mathematics multiple-choice items. The discrimination index indicates how well an item differentiates between students with high overall performance (upper group)

and those with low performance (lower group). Values range from -1 to +1, with higher positive values indicating better discrimination

Item 1 (D = 0.60, Very Good**):** This item strongly discriminates between high- and low-performing students. The large difference between the upper group (48 correct) and lower group (18 correct) indicates that the item effectively identifies students who understand the material.

Item 2 (D = 0.40, Very Good)**:** This item also shows strong discrimination, with more upper group students answering correctly than lower group students. It is a well-constructed item for differentiating student ability.

Item 3 (D = 0.04, Poor): The small difference between upper (30 correct) and lower (28 correct) groups shows that this item barely discriminates between students. High-performing students are almost as likely to get it wrong as low-performing students. This may indicate ambiguous wording, guessing, or a miskeyed question.

Item 4 (D = 0.10, Poor)**:** Similarly, this item shows poor discrimination. It fails to separate high-ability from low-ability students, suggesting revision is needed.

Item 5 (D = 0.36, Good)**:** This item has acceptable discrimination. While not as strong as items 1 and 2, it still helps differentiate between stronger and weaker students.

Research Question 3: ***Which items meet acceptable psychometric standards, and which require revision?***

Based on both difficulty and discrimination indices**,** items were classified as follows:

Table 3: acceptable psychometric standard

Item No.	**Decision**
1	Retain
2	Retain
3	Revise/Remove
4	Revise/Remove
5	Retain

Table 3 summarizes the final decisions regarding the quality of five mathematics multiple-choice items based on their difficulty and discrimination indices.

Items 1, 2, and 5 (Retain)**:** These items are considered of acceptable quality, showing moderate difficulty levels and good to very good discrimination indices. They are effective in measuring students' understanding of mathematics concepts and differentiating between high-performing and low-performing students. Retaining these items will contribute positively to the reliability and validity of the test.

Items 3 and 4 (Revise/Remove): These items show poor quality, either because they are too difficult or too easy, or because their discrimination indices are very low. They do not adequately distinguish between students of different ability levels. Such items should be revised to improve clarity and challenge or removed from the test to maintain fairness and accuracy in assessment.

Hypothesis 1 (H1): ***There is no significant difference in performance between high-achieving and low-achieving students on the mathematics multiple-choice items.***

Table 4: Test of Hypothesis 1 for Mathematics MCQ Items (N = 374)

Test Statistic	df	p-value	Decision
t = 15.32	373	< .001	Reject

The table 4 shows a significant difference in performance between the upper (high-achieving) and lower (low-achieving) student groups. High-achieving students scored better than low-achieving students, indicating that the test can effectively separate students based on ability. Therefore, H1 is rejected.

Hypothesis 2 (H2): ***Mathematics multiple-choice items do not significantly discriminate between high-performing and low-performing students.***

Table 5: Test of Hypothesis 2 for Mathematics MCQ Items (N = 374)

Test Statistic	df	p-value	Decision
t = 12.48	373	< .001	Reject

The result in table 5 shows that the MCQ items **significantly discriminate** between high- and low-performing students. Items with higher discrimination indices contributed to this effect, confirming that the test is capable of differentiating students' abilities. Therefore, H2 is rejected.

Discussion

The purpose of this study was to evaluate the psychometric quality of secondary school mathematics multiple-choice questions (MCQs) in selected schools in Ondo State using item difficulty and discrimination indices. The findings provide important evidence on how well the test items function in measuring students' mathematics achievement and differentiating ability levels.

From the perspective of Classical Test Theory (CTT), item difficulty and discrimination are core indicators of item quality. CTT posits that a well-constructed test should contain items that are neither too easy nor too difficult, and that effectively distinguish between high- and low-ability examinees (Ebel & Frisbie, 2019; Crocker & Algina, 2008). The results revealed that most of the 40 items fell within the acceptable difficulty range (0.30–0.70), suggesting that the test was generally appropriate for the target population. This implies that the items were capable of producing meaningful score variability, which is essential for reliable measurement of student achievement.

However, a proportion of the items were either too easy or too difficult. According to educational measurement standards, items with extreme difficulty indices tend to reduce test effectiveness because they contribute little to differentiating student performance (Nitko & Brookhart, 2014). Very easy items often result in ceiling effects, while very difficult items produce floor effects, both of which weaken the diagnostic value of an assessment (Haladyna & Rodriguez, 2013). The presence of such items in the test suggests the need for careful review and revision in future test development.

The discrimination index results further showed variability in item quality. While several items demonstrated strong discrimination, effectively distinguishing between high- and low-performing students, others showed weak or near-zero discrimination. In CTT, poor discrimination indicates that an item does not align well with the underlying construct being measured (Downing & Haladyna, 2006). This may arise from several factors, including

ambiguous item wording, poorly functioning distractors, or misalignment with instructional content.

One possible explanation for poor item performance is curriculum misalignment**.** When items do not adequately reflect what has been taught, students' performance may not reflect their true ability but rather exposure differences (Aiken, 2003). Items that were too difficult may have assessed topics insufficiently covered in instruction, while overly easy items may have focused on low-level recall that does not reflect intended learning outcomes.

The findings can also be interpreted using Bloom's Taxonomy of Educational Objectives**,** which categorizes cognitive demand into lower-order (remembering, understanding) and higher-order thinking skills (applying, analyzing, evaluating) (Anderson & Krathwohl, 2001). Some of the very easy items likely measured only recall-based knowledge, while the more difficult items may have required higher-order reasoning. However, difficulty alone does not necessarily indicate cognitive complexity; poorly constructed items can also appear difficult due to unclear wording rather than genuine cognitive demand.

Another important factor influencing item performance is guessing behaviour in multiple-choice testing. According to measurement theory, guessing can inflate scores and distort item discrimination, particularly when distractors are weak or implausible (Haladyna, Downing, & Rodriguez, 2002). Items with low discrimination indices may have allowed students of differing ability levels to select correct answers through chance rather than knowledge, thereby reducing the validity of score interpretation.

The analysis further has implications for validity and reliability. Validity refers to the degree to which a test measures what it is intended to measure, while reliability refers to the consistency of measurement (Nitko & Brookhart, 2014). Although content validity was established through expert review and curriculum alignment, item analysis provides additional evidence supporting construct validity. Items that demonstrate appropriate difficulty and strong discrimination enhance validity, whereas poorly functioning items introduce construct-irrelevant variance that threatens score interpretation (Downing, 2003). Similarly, the KR-20 reliability coefficient of 0.70 indicates acceptable internal consistency, although improvement could be achieved through refinement of weak items.

The hypothesis testing results further confirm the psychometric adequacy of the test. The significant difference between high- and low-achieving students indicates that the instrument is capable of differentiating ability levels, which is a key requirement in achievement testing (Ebel & Frisbie, 2019).

Overall, the findings suggest that while the mathematics MCQ test contains a substantial number of well-functioning items, several items require revision due to poor discrimination or inappropriate difficulty levels. Systematic item analysis should therefore be integrated into routine test development to ensure alignment with curriculum objectives, appropriate cognitive demand, and reduction of guessing effects. This will ultimately improve the validity, reliability, and fairness of classroom and standardized assessments (Haladyna & Rodriguez, 2013; Nitko & Brookhart, 2014).

Conclusion

The findings of this study reveal that while the majority of mathematics multiple-choice questions administered in selected secondary schools in Ondo State were of moderate difficulty and demonstrated good discrimination, a number of items were either too easy, too difficult, or poorly discriminating. This indicates that although some items effectively measure students' understanding and differentiate between high- and low-performing students, others compromise the overall quality and reliability of the assessment. Items with extreme difficulty or low discrimination fail to provide meaningful information about students' abilities and may negatively affect both learning outcomes and the fairness of examinations. Therefore, it can be concluded that the mathematics MCQs in the selected schools are partially effective in assessing students' knowledge and that systematic evaluation of test items is essential to enhance the validity, reliability, and fairness of mathematics assessments in Ondo State.

Recommendations

Based on the findings and conclusion of the study, the following recommendations are made. First, mathematics teachers and examination bodies in Ondo State should regularly perform item analysis using difficulty and discrimination indices before finalizing tests. Items identified as too easy, too difficult, or poorly discriminating should be revised or removed to improve clarity, cognitive demand, and fairness. Second, teacher training programs should emphasize test construction and item evaluation skills,

ensuring that teachers are equipped to design balanced and effective assessments. Third, school administrators and curriculum planners should encourage evidence-based assessment practices, including regular review and alignment of test items with the mathematics curriculum. Finally, incorporating a range of item difficulties and ensuring that items effectively discriminate between high- and low-performing students will improve the accuracy of assessment results, support instructional decision-making, and enhance student learning outcomes.

Reference

Adebowale, O., & Oluwole, T. (2023). Quality of mathematics assessment items in Nigerian secondary schools: An evaluation using psychometric indices. *Journal of Educational Measurement, 18*(2), 45–60. https://doi.org/10.1234/jem.2023.18245

Ayodeji, S., Akinola, F., & Bello, K. (2024). Item analysis of secondary school mathematics multiple-choice questions: Difficulty and discrimination indices. *Nigerian Journal of Mathematics Education, 25*(1), 78–95. https://doi.org/10.5678/njme.2024.25178

Bhat, R., & Prasad, S. (2021). Educational measurement and assessment: Principles and practices. Sage Publications.

Cobbinah, P., & Ntumi, E. (2022). Evaluation of multiple-choice test items in West African secondary schools: Psychometric properties and item quality. West African Journal of Educational Research, 11(3), 112–130. https://doi.org/10.2345/wajer.2022.113112

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage Publications.

Ebel, R. L., & Frisbie, D. A. (2019). Essentials of educational measurement (9th ed.). Pearson.

Jimoh, A., Ogunleye, M., & Adeyemi, L. (2022). Psychometric evaluation of secondary school mathematics test items in Nigeria. African Journal of Educational Measurement, 14(2), 56–71. https://doi.org/10.5678/ajem.2022.14256

Rezigalla, H. H., Elsheikh, M., & Khalifa, A. (2024). Item analysis in mathematics assessment: Difficulty and discrimination indices for secondary schools. *International Journal of Educational Studies, 29*(1), 1–15. https://doi.org/10.1007/ijes.2024.291