

DIFFERENTIAL ITEM FUNCTIONING ANALYSIS OF 2022 AND 2023 ALGEBRA AND ELEMENTARY TRIGONOMETRY EXAMINATIONS AT FEDERAL POLYTECHNIC DAURA, KATSINA STATE, NIGERIA.

Shu'aibu Bello Kaseem

Directorate, Academic Planning, Research & Linkages
Federal Polytechnic Daura, Katsina State, Nigeria
08036868599

Corresponding Author Email: shuaibubellokaseem@fedpolydaura.edu.ng

Abstract

DOI: <https://doi.org/10.57233/gujed.v3i2.03>

This study examined Differential Item Functioning (DIF) in the Algebra and Elementary Trigonometry examinations administered in 2022 and 2023 at Federal Polytechnic Daura, Katsina State, Nigeria. The investigation was motivated by consistently poor student performance in these subjects, characterized by high failure rates and low grades. Using the Mantel-Haenszel procedure and logistic regression, item responses from 110 students were analyzed to identify DIF across gender and socio-economic groups. The results indicated that several items exhibited significant DIF, favoring male students or those from higher socio-economic backgrounds. These findings suggest the influence of construct-irrelevant factors on performance. Recommendations include revising biased items, enhancing instructional methods, and strengthening academic support to ensure equitable assessment practices.

Keywords: Differential Item Functioning, Mathematics Assessment, Socio-Economic Status

1.1 Introduction

Educational assessments are critical for measuring learning outcomes, informing instructional decisions, and promoting accountability in education (OECD, 2021). The validity of these assessments depends on the assumption that test items measure intended constructs consistently across all student groups (AERA, APA, & NCME, 2018). Differential Item Functioning (DIF) occurs when students of equivalent ability, but from different demographic groups, have differing probabilities of answering a test item correctly (Zumbo, 2007).

At Federal Polytechnic Daura, consistent underperformance in Algebra and Trigonometry during the 2022 and 2023 academic sessions prompted an examination of whether DIF contributed to inequitable outcomes. While instructional limitations, poor preparation, and resource constraints have been identified as factors (Yusuf & Lawal, 2023), the role of item-level bias has not been systematically investigated.

1.2 Statement of the Problem

Despite the implementation of various instructional interventions aimed at improving student performance, a substantial proportion of students continue to fail Algebra and Trigonometry courses at Federal Polytechnic Daura. Among those who manage to pass, a majority achieve only

low grades, typically D or E, suggesting limited mastery of the subject matter. This persistent underperformance raises questions not only about instructional quality but also about the nature of the examinations themselves. Specifically, it remains uncertain whether the test items operate consistently and fairly for all students, irrespective of demographic characteristics. Gender and socio-economic status, in particular, may influence how students respond to individual items, potentially advantaging some groups while disadvantaging others. If such disparities exist, they could compromise the validity of the examinations and misrepresent students' true mathematical abilities. Moreover, inequitable items may exacerbate existing performance gaps, further disadvantaging already marginalized groups. The uncertainty surrounding the fairness of the assessment process underscores the need for a systematic evaluation of the test items. Understanding whether differential item functioning occurs is crucial for ensuring that examination scores accurately reflect students' competencies. Addressing potential item bias is therefore essential for promoting equity, enhancing the credibility of assessment outcomes, and supporting all students in achieving their academic potential.

1.3 Objectives of the Study

The study aimed to:

1. Identify the presence and extent of DIF in the 2022 and 2023 Algebra and Elementary Trigonometry examinations.
2. Determine whether gender influenced performance on specific items.
3. Determine whether socio-economic status influenced performance on specific items.

1.4 Significance of the Study

The study offers empirical evidence highlighting the potential presence of item bias within core mathematics assessments, particularly in Algebra and Trigonometry. The findings indicate that certain test items may advantage or disadvantage students based on demographic characteristics such as gender and socio-economic status, rather than their actual mathematical ability. By identifying these disparities, the research provides examination panels with a basis for reviewing, revising, or removing biased items to ensure that assessments more accurately reflect students' true competencies. Furthermore, the results promote the adoption of fair and equitable assessment practices across the institution, supporting the development of policies that uphold academic integrity. The study also underscores the importance of aligning instructional and assessment strategies to mitigate inequities, thereby enhancing overall student performance. In addition, the findings contribute to ongoing scholarly discourse on fairness and validity in educational measurement, offering insights that can inform future research, policy decisions, and professional practice. By addressing item-level biases, institutions can create a more inclusive and supportive learning environment that fosters equal opportunities for success among all students. Overall, the research emphasizes the critical role of rigorous assessment evaluation in maintaining both the credibility and equity of academic outcomes.

1.5 Scope and Delimitation

The study focused on DIF analysis of the 2022 and 2023 Algebra and Trigonometry examinations among 110 students sampled across gender and socio-economic categories. Analysis was restricted to gender and SES variables, using quantitative techniques (Mantel–

Haenszel procedure and logistic regression). Other demographic variables were not considered, and qualitative factors such as teaching style were outside the scope.

2.1 Review of Related Literature

The conceptual framework of the study was anchored in Classical Test Theory (CTT) and Item Response Theory (IRT), which collectively emphasize that observed test scores are influenced both by student ability and item characteristics. CTT posits that a student's test score comprises a true score and random error, highlighting the importance of evaluating individual test items to ensure reliability and fairness. IRT complements this by modeling the probability of a correct response as a function of latent ability, allowing for detailed analysis of item difficulty and discrimination across diverse student groups. In addition, Messick's framework of validity was applied to underscore that fairness is an essential dimension of assessment validity, asserting that tests should measure intended constructs without introducing irrelevant sources of bias.

Differential Item Functioning (DIF) forms the core of the framework, providing a systematic approach to evaluate whether individual test items operate equivalently for students of equal ability but from different demographic backgrounds. DIF analysis identifies items that may favor certain groups, such as males over females or students from higher socio-economic status (SES) over those from lower SES, thereby detecting potential inequities embedded within assessments. It is important to note that the presence of DIF does not automatically imply intentional bias; rather, it signals the need for careful review and possible item revision to enhance fairness and validity.

The study also accounted for key student demographic characteristics, particularly gender and socio-economic status, as these factors have been shown to influence mathematics performance in the Nigerian context. Gender differences may arise due to socialization patterns, classroom dynamics, or differential familiarity with problem-solving approaches, while SES disparities often reflect differences in access to learning resources, technology, and academic support. These demographic variables interact with item characteristics to influence student performance, forming the basis for investigating DIF within Algebra and Trigonometry examinations.

Academic performance was operationalized as the scores and grades obtained in the Algebra and Elementary Trigonometry examinations. Performance outcomes were analyzed in relation to gender and SES to determine whether observed differences were attributable to item bias or other construct-irrelevant factors. Prior studies in Nigeria indicate that mathematics achievement is affected by instructional quality, student preparation, resource availability, and socio-cultural influences, providing context for interpreting DIF results.

The theoretical assumptions guiding the study further extended to the analysis of test items themselves. Items from the 2022 and 2023 examinations were evaluated for functionality across student groups, with the aim of identifying questions that may inadvertently disadvantage particular students. Statistical techniques such as the Mantel-Haenszel procedure and logistic regression were employed to detect uniform and non-uniform DIF, ensuring a robust assessment of item-level performance. By integrating CTT, IRT, and Messick's validity framework, the conceptual framework provided a comprehensive lens through which to examine the fairness, reliability, and equity of mathematics assessments, supporting the identification of factors that contribute to systemic disparities in student outcomes.

3.0 Methodology

A quantitative ex post facto design was employed. Stratified random sampling ensured representative inclusion of male and female students across SES categories. Data included item-

level responses from 110 students on the 2022 and 2023 Algebra and Trigonometry examinations. DIF detection used the Mantel–Haenszel procedure for uniform DIF and logistic regression for uniform and non-uniform DIF. Analyses were conducted using R (difR package) and SPSS, with $\alpha = 0.05$ and multiple testing correction applied via the Benjamini–Hochberg procedure.

4.0 Results

4.1 Student Performance

The analysis of student performance revealed that overall achievement in both Algebra and Elementary Trigonometry was suboptimal. The mean score for Algebra was 45.6 with a standard deviation of 14.3, while the mean score for Trigonometry was slightly higher at 48.1 with a standard deviation of 13.7. These results indicate that the majority of students scored below the conventional pass mark of 50%, suggesting widespread difficulties in mastering core mathematical concepts.

Further examination of the grade distributions showed that scores were predominantly clustered around grades C, D, and E. Specifically, only a small proportion of students achieved grades A or B, indicating high-level mastery, whereas the bulk of students obtained average or below-average grades. This clustering pattern underscores the prevalence of low achievement and highlights that many students were unable to demonstrate consistent proficiency in the subject matter.

More than half of the sampled students failed Algebra and Trigonometry, reflecting persistent challenges in these foundational courses. The high failure rates are indicative of systemic difficulties that may include instructional gaps, limited access to learning resources, and possible biases in the examination items themselves. The performance trends also suggest that students struggled with both conceptual understanding and problem-solving skills, which are critical for success in higher-level mathematics.

These findings provide essential context for the subsequent DIF analysis, as the low and clustered performance may exacerbate the impact of item-level disparities. Understanding the distribution of scores is crucial for interpreting whether specific test items functioned differently across demographic groups and for identifying potential inequities that could influence academic outcomes.

4.2 DIF Analysis

The Differential Item Functioning (DIF) analysis revealed notable disparities in how students responded to specific examination items. Using the Mantel–Haenszel procedure, several items were identified as exhibiting statistically significant DIF. In particular, items A5, A12, and T7 were found to favor male students, suggesting that these questions may have required skills or familiarity that advantaged male learners. Conversely, items T14 and A18 were observed to disadvantage students from lower socio-economic status (SES) backgrounds, indicating that access to resources, prior preparation, or other contextual factors may have influenced performance independently of mathematical ability.

To further validate these findings, logistic regression analyses were conducted, modeling the probability of correctly answering each item as a function of student gender, SES, and overall ability. The results confirmed that gender was a significant predictor for items A5, A12, and T7, with male students demonstrating higher odds of answering correctly compared to female students. Similarly, socio-economic status significantly predicted responses for items T14 and A18, with students from higher SES backgrounds exhibiting greater likelihoods of success. The

regression results reinforce the Mantel-Haenszel findings and highlight that DIF in these items reflects meaningful differences in performance that are linked to demographic characteristics rather than solely to students' underlying ability.

Overall, the DIF analysis demonstrates that certain examination items functioned inequitably across demographic groups, potentially contributing to the observed patterns of low achievement. The presence of DIF underscores the importance of reviewing and refining test items to ensure that assessments measure intended competencies fairly and accurately across all students, thereby promoting equity in academic evaluation.

4.3 Discussion

The findings of this study indicate that Differential Item Functioning (DIF) was present in several Algebra and Elementary Trigonometry examination items, suggesting that the assessments did not consistently measure student ability across demographic groups. The presence of DIF highlights a critical concern for assessment fairness, as certain items appear to advantage or disadvantage students based on gender or socio-economic status rather than their actual mathematical competency. Specifically, male students were consistently advantaged on items requiring complex reasoning and higher-order problem-solving skills, whereas students from lower socio-economic backgrounds were disproportionately disadvantaged on items emphasizing conceptual understanding or application of knowledge.

These disparities suggest that performance outcomes were influenced by factors unrelated to students' intrinsic ability, including prior access to educational resources, familiarity with problem formats, and differential exposure to supportive learning environments. When coupled with documented instructional limitations, such as reliance on traditional teaching methods and limited student engagement, as well as resource constraints affecting lower SES students, these item-level biases likely exacerbated the observed patterns of poor performance. The clustering of student scores around low to average grades further indicates that DIF may have amplified existing achievement gaps, particularly disadvantaging students already facing systemic challenges.

The study's findings reinforce the notion that equitable assessment requires careful item review and validation to ensure that test questions accurately reflect intended learning outcomes for all students. Addressing DIF is not only a technical imperative for improving test validity but also a critical component of promoting educational equity. Failure to account for such biases risks misclassifying student ability, limiting academic progression, and perpetuating disparities in higher education performance. Consequently, the results underscore the need for institutions to integrate fairness considerations into assessment design, adopt inclusive teaching strategies, and provide targeted support to mitigate the combined effects of instructional, resource, and item-level inequities.

5.1 Conclusion

The analysis of the 2022 and 2023 Algebra and Trigonometry examinations revealed that several items favored male students or those from higher socio-economic backgrounds, contributing to high failure rates and low grades. These biases, combined with instructional limitations and resource disparities, created an inequitable assessment environment. Addressing these issues requires revising biased items, enhancing inclusive teaching strategies, and providing targeted academic support for disadvantaged students. Implementing these measures can improve fairness and ensure that assessments more accurately reflect students' true mathematical abilities.

Ultimately, such interventions promote equity and enhance overall academic outcomes in higher education.

5.2 Recommendations

The study recommends reviewing and revising examination items flagged for Differential Item Functioning to ensure fairness across all student groups. Inclusive, student-centered instructional strategies should be adopted to enhance engagement and understanding. Academic support programs should be expanded, particularly for students from disadvantaged backgrounds. Socio-economic barriers can be mitigated through scholarships and provision of learning resources. Finally, ongoing monitoring and evaluation of item fairness should be institutionalized to maintain equitable assessment practices.

References

Adegbite, A. T., Yusuf, A., & Ibrahim, M. (2022). Assessment practices and student performance in Nigerian tertiary institutions: A case study of mathematics courses. *Nigerian Journal of Educational Measurement*, 30(1), 55–70.

Aina, J. K., & Ayodele, M. O. (2022). Gender differences in mathematics performance: A case study of selected Nigerian secondary schools. *African Journal of Educational Studies*, 15(2), 35–47.

Awofala, A. O. A., & Odogwu, H. N. (2022). Trends in mathematics performance in Nigerian institutions: Challenges and prospects. *Nigerian Journal of Curriculum Studies*, 29 (3), 49–65.

Agwagah, U. N. V., & Okafor, F. C. (2023). Innovations in mathematics teaching: A solution to students' poor performance in Nigerian tertiary institutions. *Journal of Educational Innovations and Research*, 18(1), 22–29.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57*(1), 289–300.

Boone, W. J. (2021). *Rasch analysis for instrument development: Why, when, and how?* SAGE Publications.

Camilli, G., & Shepard, L. A. (2013). *Methods for identifying biased test items*. SAGE Publications.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2020). *Applied multiple regression/correlation analysis for the behavioral sciences*. (3rd ed.). Routledge.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. *Educational Testing Service*.

Ekpo, C. M., & Etim, M. O. (2022). Curriculum overload and students' performance in mathematics in Nigerian polytechnics. *International Journal of Educational Policy and Practice*, 11(1), 77–89.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. *ETS Research Report Series*.

Ibrahim, A. M., & Lawal, B. A. (2023). Instructional strategies and their effect on mathematics performance among polytechnic students in Northern Nigeria. *Journal of Pedagogical Research*, 8(1), 93–106.

Li, M., & Zumbo, B. D. (2021). A review of differential item functioning methods and an agenda for future research. **Educational Measurement: Issues and Practice*, 40*(3), 36–45.

National Bureau of Statistics. (2023). *Education statistics in Nigeria: Annual report*. Abuja, Nigeria.

Obioma, G., & Udu, D. A. (2023). Study habits and academic achievement: Evidence from Nigerian higher education. *African Journal of Research in Education*, 12(4), 55–70.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370.

Yusuf, A., & Lawal, M. (2023). Challenges of teaching mathematics in Nigerian polytechnics: Implications for curriculum reform. *Journal of Technical Education Studies*, 14(2), 22–34.

Zumbo, B. D. (2007). Three generations of DIF analyses: A historical and methodological perspective. *Canadian Journal of Experimental Psychology*, 61(1), 16–26.

Zwick, R. (2012). *Introduction to test fairness*. Routledge.