Kashere Journal of Education 2024, 7(2):216-226 ISSN: 2756-6021 (print) 2756-6013 (online)
Abubakar, M., Mayowa O. O., Dauda M., & Abubakar, H. (2024)

DIFFERENTIAL ITEM FUNCTIONING METHOD AS AN ITEM BIAS INDICATOR FOR ENGLISH LANGUAGE MOCK SSCE EXAMINATION IN GOMBE STATE

Abubakar, Mohammed
Department of Educational Foundations,
Faculty of Education, Federal University of Kashere
abuamirandamira@gmail.com

And
Mayowa O. Ogunjimi
Department of Adult and Primary Education Studies,
Faculty of Education, University of Ilorin, Ilorin, Kwara State

And
Dauda Muhammad
Science Education Department,
Faculty of Education, Federal University of Kashere, Gombe State.

And
Abubakar, Hauwau
Department of Science Education,
Faculty of Education Gombe State University, Gombe State.

Abstract
This study examined Differential Item Functioning (DIF) in the 2023 English Language Multiple-choice Mock Examination in Gombe State, Nigeria. The Objectives of this study was to detect differential Item Functioning of 2023 English Language Multiple-choice Mock Senior School Certificate Examination in Gombe state through the following objectives: to find out the number of items functioned differentially between Boys' and Girls' students of Mega schools in the 2023 English Language Multiple-choice Mock Examination; to find out the number of items functioned differentially between Boys' and Girls' students of urban schools in the 2023 English Language Multiple-choice Mock Examination. Using a sample of 756 students from 16 public senior secondary schools, the study employed binary logistic regression analysis to detect DIF. The results revealed that some items exhibited DIF across gender, school type, and location. Specifically, items exhibited DIF in mega schools, urban schools, and science-based schools. The findings highlight the importance of examining DIF in educational tests to ensure fairness and validity. The study recommends that test developers should regularly examine DIF in educational tests and review or replace items exhibiting DIF to prevent bias against certain groups of students.

Introduction

Assessment plays a critical role in the Nigerian educational system, particularly in measuring students' academic achievement. Among the core subjects, English Language holds a unique position as both a subject of study and a medium of instruction. Success in Senior School Certificate Examinations (SSCE) such as WAEC, NECO, and NABTEB is a prerequisite for tertiary

education admission in Nigeria. However, concerns have been raised regarding the fairness of test items, particularly their potential bias against different subgroups of students based on gender, location, school type, and subject combination.

English Language serves as a critical tool for communication, education, and social mobility in Nigeria. It is the official language and the primary medium of instruction across all levels of education, from primary school to tertiary institutions. English is also a compulsory subject in secondary schools and a prerequisite for admission into higher institutions, as stipulated in the National Policy on Education (NPE, 2013). Given its significance, students' achievement in English Language is an essential indicator of academic success and future educational opportunities. Thus, English Language is a medium of interaction in schools and even outside and is a compulsory subject offered in all formal institutions in Nigeria that must be passed at SSCE (WAEC, NECO, NABTEB, ETC) before being given admission into higher levels of education. It is also a medium of instruction for all the school subjects from the primary school level to the tertiary institutions as well as the official language of the country. Therefore, higher institution students need effective English Language to function properly or to perform better (Umar et al, 2024).

As such, to achieve the goals and objectives of teaching English language at the Senior Secondary School level, the assessment of student's academic achievement demands attention. Academic achievement is defined as the degree or level of success attained at the end of an academic endeavour. Chowdhury and Pati (2011) opined that it is a test or examination marks and grades given by teachers in academic subjects. Ricarda (2015) viewed academic achievement within any context as performance relative to some standard and further stated that academic achievement and other dimensions of learning can be measured by a variety of yardsticks or measuring instruments, the result of which are types of scores, ranks or grades. However, this academic achievement could be affected by assessment methods or testing instruments. As such, some test scores fail to produce the true learning outcome. This could be seen in the poor examination result of Senior Secondary School in Nigeria. By implication, most students cannot gain admission into tertiary institutions to achieve very well.

Test is an instrument for systematic measure of a person's behaviour or academic traits with the aid of a numerical scale. It is the easiest instrument to measure the cognitive domain of learners. As seen by Obinne and Amali (2014) and Amajuoyi (2015), test is an instrument used for measuring ability, achievement, interest and other traits. Test results help educators to know how much students have learnt and to provide feedback for restructuring and is used for modifying the teaching-learning process. Test consists of a set of uniform questions or tasks to which a student is to respond independently and the result of which can be treated in such a way as to provide a quantitative comparison of the performance in different students (Nworgu, 2015).

Test can as well be classified based on item format, as essay and objective tests. Under the objective test, there is the most commonly used type of test called multiple choice test. Here, the testees are required to choose one response among the set of many alternatives (Opara, 2016). It comprises of a stem, a key and distracters. It's agreed that through testing, latent abilities of examinees

are revealed. The assessment of learning outcomes in English Language by using tests is one of the crucial and basic discussions in education.

Despite these functions of tests and the fact that modern societies have adopted testing as the most objective means of decision making in education, various criticisms have been raised on issues concerning testing in recent times (Amajuoyi, 2015). From the description of test and testing above, a test is supposed to measure students'/examinees' ability/performance or other latent traits irrespective of the sub-group they belong. Tests are expected to provide equal opportunities to all examinees to demonstrate their latent construct, abilities and knowledge irrespective of their socio-demographic factors like gender; location; ethnic, cultural and religious groups. Latent construct here means unseen abilities possessed by a learner.

Buttressing the assertion, Robitzsch (2023) commented that the response to a particular test item is determined by the latent construct of interest, referred to as theta (θ), being measured. The achievement tests, have been faulted (Nworgu, 2015; Moyo & Nenty, 2017) for limitations and bias in what they intend to measure. The existence of bias introduces measurement error and hence decreases the validity of the entire test. Bias is the presence of some characteristic of a test and/or an item in a test that results in differential performance for two individuals of the same ability but from different ethnic, sex, cultural or religious groups. When the whole test is the unit of concern, it is referred to as test bias. Amajuoyi (2015) defined test bias as objective statistical indices that examine the patterning of test scores for relevant sub-populations.

Differential Item Functioning (DIF) is a concept used to detect whether test items function differently for examinees of the same ability but from different demographic groups. When an item shows DIF, it means that students of equal proficiency have different probabilities of answering the item correctly based on external factors rather than their true ability. DIF analysis is crucial in ensuring fairness and validity in standardized tests (Umar et al, 2024, Jimoh et al, 2023) The absence of DIF is determined by the fact that the conditional probability distribution of the response is not dependent on group membership. This is applicable to dichotomous scored tests. Dichotomous scores are scores from a test which has two categories of outcome which may be pass and fail, yes and no or true and false, usually coded as 1 and 0 respectively.

Uniform DIF is a form of DIF which occur when a sub-group of examinees with ability levels, uniformly answer a particular item or subset of items than the other group, that particular sub-group is said to be advantaged over the other group and can be considered as having a superior ability over the less favoured group and is termed as "reference" group, while the less (inferior group) advantage group is the "focal" group in bias analysis comparatively While Non-uniform DIF occurs when there is an interaction between the higher ability and lower ability group for an item seems to be difficult for those at higher level in one group, and more difficult for those at lower level in the other group (Bizumic, et al, 2023; Walker, 2011). For a dichotomously scored test, the probability of correctly responding to an item is the same for members of either group once matched on the same attribute. This indicates that there is no DIF because members of the reference and focal groups with the same underlying ability or attribute have the same

probability of responding correctly. Therefore, there is no disadvantage for one group over the other.

There are two major test theories; Classical Test Theory (CTT) and Item Response Theory (IRT). Classical Test Theory (CTT) Focuses on observed scores as a combination of true ability and measurement error while Item Response Theory (IRT) which is a modern approach that examines the relationship between student ability and test item characteristics. IRT is particularly useful for DIF analysis, as it allows researchers to examine whether students of equal ability but different demographic backgrounds have the same probability of answering an item correctly. In addition, IRT also called modern theory is a set of latent variable techniques (Latent Trait Theory) especially designed to model the interaction between a subject's "ability" and the item level stimuli (difficulty, guessing, etc.). It is a modeling technique that tries to describe the relationship between an examinee's test performance and the latent trait underlying the performance (Xue & Chen, 2023).

There are several methods to detect if items have DIF effects. Some of the methods include but not limited to: chi-squares techniques: Mantel-Haenszel (M-H)/Generalized Mantel-Haenszel (GMH) methods, Scheuneman Modified Chi-square ($SS\chi^2$), Cochran's Chi-square Test Method ($CT\chi^2$) and Lord's Chi-Square (χ^2) Test (Magis et al., 2010). Mantel-Haenszel and logistic Regression methods are going to be used for the present research.

According to Ogbebor and Onuka (2013) claimed that the overall impact of item DIF accumulates across the whole test. By implication, test developers should try as much as possible to construct test items that may have minimal DIF effect. The socio-demographic variables involved in this study are; gender, location, school type, and subject combination.

Gender is a set of characteristics distinguishing between males and females, particularly, in the case of man and woman which, depending on the context, may vary from sex to social role and to gender identity (Bland, 2013). Nevertheless, Chang (2013) reported that although there is a decrease in the gap in gender difference in students' performance in subjects, female representation in subjects involving calculation like sciences is still low in comparison with their male counterparts. Considering socio-cultural background, Oludipe (2012) observed that in Nigeria, certain vocations and professions have traditionally been regarded as men's and others; women. Some of these vocations are Medicine, engineering, architecture and nursing, catering, typing respectively. The society's socio-cultural construct of females as weaker sex, together with females' self-perception of themselves as weaker sex, inferior and subordinate to the males, have imposed some socio-cultural limitations on female aspirations and achievement in sciences (Akinlolu et al., 2023; Wang & Degol, 2017). Similarly, Onyeocha et al., (2023) inferred that the socio-cultural upbringing of females within most Nigerian homes tends to shape the girl-child away from science and science related disciplines. For instance, in most homes, what are regarded as complex and difficult tasks are allocated to boys whereas girls are expected to handle the relatively easy and less demanding tasks. Consequently, fewer females opt for science subjects thereby creating some differences in the number of males and females in science discipline in favour of the males.

School type generally refers to boarding and day schools/mixed or single school is another contributory variable in DIF. The type of school a student attends, to a large extent, influence one's academic performance. Researches, such as Moonpreneur (2023), Archbishop Murphy High School (2022) and Thomas High School (n. d.), have reported that private schools generally outperform their public counterparts.

The type of school a student attends (single–sex school or mixed- sex school) could to a large extent, influence one's academic performance. Single-sex school refers generally to education at the elementary, secondary, or post- secondary level in which males or females attend school exclusively with members of their own sex. Mixed-sex school on the other hand refers generally to education in which both male and female attend (Single-sex schools VS mixed-gender schools: Comparison impacts, 2023; Eisenkopf et al., 2015). The present study researcher added to researches on boarding and day schools or mixed and single sex schools' dichotomy in English language achievement.

Having looked at some of the socio-demographic variables that could indicate differential item functioning, the question then is: can differential item functioning in standardized multiple choice English language test in Gombe state Mock Examination be detected and examined?

Statement of the Problem

In an ideal educational system, standardized tests such as the Senior Secondary Certificate Examinations (SSCE) conducted by WAEC, NECO, and NABTEB should fairly and accurately assess students' attainments proficiency without any form of bias. These tests should be valid and reliable, ensuring that all students regardless of their demographic backgrounds have equal opportunities to demonstrate their true abilities. Item Response Theory (IRT) principles, particularly unidimensionality and local independence, should be upheld, ensuring that test items measure only the intended construct (English language proficiency) without being influenced by external factors such as gender, school type, location, or socio-economic status.

The National Examinations Council (NECO) releases yearly school certificate results every year, and underachievement in English language has continued to be a recurring pattern, leading to a high failure rate (Premium Times, 2023; Tribune Online, 2023). That was supported by the outcomes of the mock exams taken by students in Gombe State over the previous five years. Only 25% of the 23,571 students who took the mock exam in 2019 and 20% of the 27,305 candidates who took it in 2020 passed with English. According to data compiled by the Gombe State Ministry of Education, out of 25,946 candidates, only 32% had English proficiency in 2021. In 2022 and 2023, that percentage was 38% and 34%, respectively, and for those with five credit passes and above, English language was included in mock exams for public schools. This could be associated with item bias, lack of one-dimensionality of test items, gender bias, ethnicity, high-low ability, school types, school locations, socioeconomic status, race bias, religion bias, construct bias, method bias, among others.

Many studies had been conducted on Differential Item Functioning of tests used in public, and private secondary schools in Nigeria. For instance, Ogbogo and Opara (2019) carried out research on differential item functioning in English language test

using item response theory for ethnic groups in south-south Nigeria. Shanmugan (2018) studied on determining gender differential item functioning for mathematics in coeducational school culture. Onuka (2013) carried out a study on differential item functioning method as an item bias indicator in Delta State. Patrick and Bright (2018) studied on Assessment of differential item functioning in social studies multiple choice questions in basic education certificate examination in Delta State Nigeria. Bakama, Malik and Chiroma (2018) Studied on the impact of Pre-National Diploma programme on academic performance in Northern polytechnics in Nigeria. Abba (2014) carried out a study on analysis of differential item functioning of senior secondary certificate (NECO) English examination in Dawakin Kudu educational zone, Kano State.

However, to the best of researcher's knowledge, none of these studies within the reach of the researcher looked in to Analysis of Differential Item Functioning of 2023 English Language Multiple-choice Mock Examination in Senior Secondary Schools in Gombe State as such the researcher seeks to explore whether certain items within the English language mock examination demonstrate DIF based on students' characteristics which may impact the validity and fairness of the assessment outcomes.

Objectives of the Study

The Objectives of this study was to detect differential Item Functioning of 2023 English Language Multiple-choice Mock Senior School Certificate Examination in Gombe state through the following objectives:

1. To find out the number of items functioned differentially between Boys' and Girls' students of Mega schools in the 2023 English Language Multiple-choice Mock Examination.
2. To find out the number of items functioned differentially between Boys' and Girls' students of urban schools in the 2023 English Language Multiple-choice Mock Examination.
3. To find out the number of items functioned differentially between Boys' and Girls' students of science-based schools in the 2023 English Language Multiple-choice Mock Examination.

Research Questions

The following research questions were raised and answered:

1. What number of items in the 2023 English Language Multiple-choice Mock Examination functioned differentially between boys' and girls' students of Mega schools?
2. What number of items in the 2023 English Language Multiple-choice Mock Examination functioned differentially between boys' and girls' students of urban schools?
3. What number of items in the 2023 English Language Multiple-choice Mock Examination functioned differentially between boys' and girls' students of science-based schools?

Methodology

This study focused on non-manipulated variables (School type, School location, etc), the study employed an Ex-Post Facto research design. An ex-post-facto design sought to find out the factors that are associated with certain occurrences of already existing condition or state of affairs and searching back in time for plausible causal factors retrospective. It's stated that,

an Ex-Post Facto design is carried out as an inquiry but did not have direct control of the independent variables because their manifestation had already occurred. In this study the design is used to examine whether items in English language multiple choice mock senior school certificate examination (SSCE) set and administered by Gombe state ministry of education in 2023 functions differently for students by gender, location, school type, and subject combination. Thus, this design was found suitable since the researcher was using an existing data which cannot be manipulated anymore. The population of the study comprised 25,173 candidates that sat for the 2023 English language multiple choice Mock examination from various public senior secondary schools which entails 14,861 and 10,312 for male and female respectively administered by the Ministry of Education, Gombe State. It has been recommended that a relatively large sample size be used in IRT 3PLM analysis. According to Karami (2015) "the Three-parameter model adds guessing and other deviant behaviours to the analysis, but requires samples of 1,000 subjects". Thus, the researcher is going to use ten percent of the target population as sample size. Therefore, the sample size for this study is 756. Multi stage sampling technique was used to determine sample population for the study. Furthermore, random sampling and purposive sampling techniques was used to sample sixteen (16) public senior secondary schools from seven (7) Local Government Areas (LGAs) in Gombe state. At the second stage, proportionate sampling technique was employed to select the percentage of students that can make the sample size according to their stratification. At the third stage, simple random sampling technique was used to avoid bias during the real selection of the participants. The instruments for data collection are Socio-Demographic Inventory (SDI) and English language Paper I Multiple Choice Test. The SDI was used to measure the socio-demographic data of the students with consideration to their gender, location, school type, and subject-combinations. The English language Paper I Multiple Choice Test used by Ministry of Education for 2023 MOCK SSCE was adopted for the study. It has sixty (60) items of A-D options but for the purpose of this write up only twenty (20) were taken. There, only one option is the correct answer. In this study, Logistic Regression (LR) method was employed using SPSS software version 26 for data analysis. Logistic Regression has been chosen due to the fact that, it is used to analyze dichotomous data such as 1 for right and 0 for wrong response and chi-square test to evaluate the level of significance of items exhibiting DIF, and descriptive statistics using frequency cross tabulation was also used in identifying the percentage of correct response for each item that functioned differently by gender, school type, school location and subject combination.

Result

Research Question One: What number of items in the 2023 English Language Multiple-choice Mock SSCE functioned differentially between boys' and girls' students of Mega schools?

Table 1: Summary of Binary Logistic Regression Analysis of 2023, English Language Mock SSCE in Mega Schools by Gender

Item No.	ΔR2	Sig.	Remarks	Item No.	ΔR2	Sig.	Remarks
1	0.001	0.684	No DIF	11	0.813**	0.000	DIF
2	0.007	0.125	No DIF	12	0.003	0.287	No DIF
3	0.061*	0.000	DIF	13	0.001	0.622	No DIF
4	0.000	0.587	No DIF	14	0.000	0.905	No DIF
5	0.006	0.157	No DIF	15	0.082**	0.000	DIF
6	0.000	0.678	No DIF	16	0.097**	0.000	DIF
7	0.003	0.304	No DIF	17	0.074**	0.000	DIF
8	0.008	0.139	No DIF	18	0.006	0.184	No DIF
9	0.094**	0.000	DIF	19	0.001	0.572	No DIF
10	0.002	0.423	No DIF	20	0.084**	0.000	DIF

Negligible DIF:$\Delta R^2 < 0.035;^{*} Moderate DIF: 0.035 \leq \Delta R^2 < 0.07,^{**} Large$ DIF:$\Delta R^2 \geq 0.07$

From the above table, the items that functioned differently with regards to gender in 2023 English language Mock SSCE at mega schools are items 3, 9, 11,15, 16,17 and 20 with significant values of 0.000,0.000,0.000,0.000,0.000,0.000 and 0.000 respectively.

Research Question Two: What number of items in the 2023 English Language Multiple-choice Mock Examination functioned differentially between boys' and girls' students of urban schools?

Table 2: Summary of Binary Logistic Regression Analysis of 2023, English Language Mock SSCE in Urban Schools by Gender

Item No.	ΔR2	Sig.	Remarks	Item No.	ΔR2	Sig.	Remarks
1	0.001	0.684	No DIF	11	0.813**	0.000	DIF
2	0.007	0.125	No DIF	12	0.003	0.287	No
3	0.061*	0.000	DIF	13	0.001	0.622	No DIF
4	0.076**	0.000	No DIF	14	0.088**	0.000	DIF
5	0.006	0.157	No DIF	15	0.082**	0.000	DIF
6	0.000	0.678	No DIF	16	0.097**	0.000	DIF
7	0.003	0.304	No DIF	17	0.074**	0.000	DIF
8	0.008	0.139	No DIF	18	0.077**	0.000	DIF
9	0.094**	0.000	DIF	19	0.001	0.572	No DIF
10	0.002	0.423	No DIF	20	0.084**	0.000	DIF

Negligible DIF:$\Delta R^2 < 0.035;^{*} Moderate DIF: 0.035 \leq \Delta R^2 < 0.07,^{**} Large$ DIF:$\Delta R^2 \geq 0.07$

From the above table, the items that functioned differently with regards to gender in 2023 English language Mock SSCE at Urban schools are items 3, 4, 9, 11, 14, 15, 16,17, 18, and 20 with significant values of 0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000 and 0.000 respectively.

Research Question Three: What number of items in the 2023 English Language Multiple-choice Mock Examination functioned differentially between boys' and girls' students of science-based schools

Table 3: Summary of Binary Logistic Regression Analysis of 2023, English Language Mock SSCE in Science-based Schools by Gender

Item No.	ΔR2	Sig.	Remarks	Item No.	ΔR2	Sig.	Remarks
1	0.001	0.684	No DIF	11	0.813**	0.000	DIF
2	0.007	0.125	No DIF	12	0.003	0.287	No DIF
3	0.045	0.132	No DIF	13	0.001	0.622	No DIF
4	0.076**	0.000	No DIF	14	0.000	0.907	No DIF
5	0.006	0.157	No DIF	15	0.082**	0.000	DIF
6	0.000	0.678	No DIF	16	0.097**	0.000	DIF
7	0.003	0.304	No DIF	17	0.074**	0.000	DIF
8	0.008	0.139	No DIF	18	0.000	0.927	No DIF
9	0.094**	0.000	DIF	19	0.001	0.572	No DIF
10	0.002	0.423	No DIF	20	0.001	0.0975	No DIF

Negligible DIF:$\Delta R^2 < 0.035$;* $ModerateDIF: 0.035 \leq \Delta R^2 < 0.07$,** $Large$ DIF:$\Delta R^2 \geq 0.07$

From the above table, the items that functioned differently with regards to gender in 2023 English language Mock SSCE at Urban schools are items 4, 9, 11, 15, 16, and 17 with significant values of 0.000,0.000,0.000,0.000,0.000, and 0.000 respectively.

Discussion

The findings of this study revealed that some items in the 2023 English Language Multiple-choice Mock Examination functioned differently across gender, school type, and location. Specifically, items 3, 9, 11, 15, 16, 17, and 20 exhibited DIF in mega schools, while items 3, 4, 9, 11, 14, 15, 16, 17, 18, and 20 exhibited DIF in urban schools. In science-based schools, items 4, 9, 11, 15, 16, and 17 exhibited DIF.

These findings are consistent with previous studies that have reported DIF in educational tests (Ogbogo & Opara, 2019; Abba, 2014). The presence of DIF in these items suggests that they may be biased against certain groups of students, which could impact the validity and fairness of the assessment outcomes.

Conclusion

This study investigated the presence of differential item functioning in the

2023 English Language Multiple-choice Mock Examination in Gombe State, Nigeria. The findings revealed that some items exhibited DIF across gender, school type, and location. These results highlight the importance of examining DIF in educational tests to ensure fairness and validity.

Recommendations

Based on the findings of this study, the following recommendations are made:

1. Test developers should examine DIF in educational tests: Test developers should regularly examine DIF in educational tests to ensure fairness and validity.

2. Items exhibiting DIF should be reviewed or replaced: Items exhibiting DIF should be reviewed or replaced to prevent bias against certain groups of students.

3. Further research is needed: Further research is needed to investigate the causes of DIF in educational tests and to develop strategies for preventing or minimizing DIF.

Reference

Abba, A. (2014). Analysis of differential item functioning of senior secondary certificate (NECO) English examination in Dawakin Kudu educational zone, Kano State.

Akinlolu, A. A., Oludipe, D. O., & Onyeocha, A. O. (2023). Factors influencing students' performance in English language in Nigeria.

Amajuoyi, R. E. (2015). The role of testing and evaluation in educational development.

Bakama, M. A., Malik, S. A., & Chiroma, A. (2018). The impact of pre-national diploma programme on academic performance in northern polytechnics in Nigeria.

Bizumic, B., Reynolds, K. J., Turner, J. C., & Subasic, E. (2023). Avowed versus unavowed prejudice: A study of differential item functioning.

Chang, H. (2013). The impact of socio-cultural factors on girls' education in Nigeria.

Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. W. (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland.

Federal Republic of Nigeria (2013). *National policy of education.* Abuja: Federal Government press.

Jimoh, M. I., Daramola, D. S., Oladele, J. I., Sheu, A. L., & Ogunjimi, M. O. (2020). Differential Item Functioning of Mathematics Joint Mock Multiple-Choice Test Items in Kwara State, Nigeria. *Journal of Multidisciplinary Studies, 9(1), 67-92.* http://dx.doi.org/10.7828/jmds.v9i1.1451

Moonpreneur. (2023). Private vs public schools: Which is better?

Obinne, J. I., & Amali, E. O. (2014). The impact of testing and evaluation on students' academic performance.

Ogbebor, P. O., & Onuka, A. O. (2013). *An investigation of differential item functioning in mathematics items using Rasch model.*

Ogbogo, B. E. (2019). An introduction to item response theory.

Ogbogo, B. E., & Opara, J. A. (2019). Differential item functioning in English language test using item response theory for ethnic groups in south-south Nigeria.

Onuka, A. O. (2013). *Differential item functioning method as an item bias indicator in Delta State.*

Onyeocha, A. O., Oludipe, D. O., & Akinlolu, A. A. (2023). *Factors influencing students' performance in English language in Nigeria.*

PremiumTimes. (2023). NECO releases 2023 SSCE results.

Robitzsch, A. (2023). Item response theory: A review of the literature.

Shanmuga, A. (2018). *Determining gender differentia item functioning for mathematics in coeducational school culture.*

Umar, U.Y., Zakka, K., & Abubakar, M. (2024). Differential item functioning of general studies mathematics examination in Gombe State Polytechnic, Bajoga, Nigeria. *Nigerian Journal of Educational Research and Evaluation, 23.*