Journal of Theoretical and Empirical Studies in Education
Vol. 9 Issue 1 December, 2024

DIFFERENTIAL ITEM FUNCTIONING IN NATIONAL EXAMINATIONS COUNCIL TESTS

Nneka Chinyere Ezeugo (PhD)[1]
Email: nc.ezeugo@unuizik.edu.ng

Chukwuemeka Benedict Ikeanumba (PhD)[2]
Email: chukwuemekaikeanumba@gmail.com

Chinweuche Chioma Okafor[3]
Email: chayojehovah@gmail.com
Department of Educational Foundations, Faculty of Education
Nnamdi Azikiwe University, Awka

Abstract
Educational tests result serve as important yardsticks through which society evaluates the outcomes of its educational system. Consequently, it is essential that educational institutions and examination bodies design assessments (test) that accurately capture the intended characteristics of their examinees. This study examined Differential Item Functioning (DIF) in the tests administered by the National Examinations Council (NECO) and assessed its impact on the tests' reliability. The study was guided by two research questions and tested two hypotheses, employing an ex-post-facto research design. The population included all candidates who took the 2023 NECO Mathematics and Economics examinations in Nigeria. Sampling was conducted in two stages: selecting the subjects and then choosing the participants. Data collection utilized the 2023 NECO past questions for these four subjects, which had already been validated by NECO's Examinations and Standard Unit in Nigeria. Since NECO provides these standardized instruments, the researcher did not independently determine their reliability. Candidate responses to selected test items were collected directly from NECO for analysis. The research questions were analyzed using an item response theory parameter logistic model with Marginal Maximum Likelihood estimation, along with mean and standard deviation calculations. Hypothesis 1 was tested through Thissen's Likelihood Ratio Test (LRT) for DIF, and Hypothesis 2 was examined using a test of equality of alpha coefficients at a 0.05 significance level. Findings among others indicated that, on average, the test items for NECO Mathematics and Economics favored male students more than their female counterparts. The study recommended that NECO should conduct DIF analysis on their assessments to ensure the quality and fairness of test items.

Keywords: Differential Item Functioning, NECO, Tests, and Academic Achievement

INTRODUCTION

Secondary school, as the name suggests, is the level of education that pupils proceed to after they have completed primary school. It is a school that offers secondary education to children. The 6-3-3-4 education system in Nigeria requires students to spend six years in primary school, three years in junior secondary school and three years in senior secondary school, and four years in a tertiary institution. The nine years of the 9-3-4 system are made up of six years in primary school and three years in junior secondary school (Enuku and Enuku, 2019). Altogether, the students must spend a minimum period of six years in secondary school. During this period, Okagbare, Ossai, and Osadebe, (2023) opined that student are expected to spend three years in junior secondary school and three years in senior secondary school. At the end of the first three years in senior secondary school, a student is expected to sit for Senior School Certificate Examination (SSCE) that qualifies such student to proceed to the tertiary institutions. The assessment for Senior School Certificate Examination (SSCE) is therefore a summative examination taken by candidates at the end of three years. It is usually conducted by each the West Africa Examination Council (WAEC) across all the state of the federation.

The national and state examinations conducted by West African Examination Council (WAEC), National Examinations Council (NECO), National Business and Technical Education Board (NABTEB) and Joint Admission and Matriculation Board (JAMB) Examination cater for candidates from various background all over the states of the federation. In some cases, an item in these examinations conducted by various examination bodies could be more difficult for particular group of examinees while very easy for other group of examinees. When there is element in an item that makes examinees that are of equal ability level but from different

population to perform differently, we say such an item shows differential item functioning (Okoye, Fejokwu and Ikeanumba, 2021).Good knowledge and mastery of secondary school subjects do not only play a pivotal role in the secondary school education but serves as a bedrock for good academic performance in higher institution and its applications are also felt in all fields of human endeavors. It is therefore not an overstatement to refer to SSCE as an indispensable element of development. Although there are various examining bodies in Nigeria that have the mandate of certifying the proficiency of candidates in SSCE in Nigeria.

Despite the enormous importance of secondary school education and the fact that it is needed for admission into higher institutions of learning, the performance of students in Mathematics and Economics at the Senior School Certificate Examination (SSCE) is not still encouraging. This poor performance is evidenced in the statistics of the performances of students in the SSCE (NECO) results shown in students' Enrolment and Number/Percentage that had five (5) credit passes including Mathematics and Economics in NECO SSCE from 2017 to 2023 (WAEC, Chief Examiner Report, 2024). Based on the information presented by WAEC Chief Examiner (2024), it is evident that the performance of Nigerian students in SSCE has been poor from 2017 to 2020. However, in 2021-2023 the percentage of candidates that was able to obtain the minimum of five credit pass was consistently above average in contrast to what obtained in the previous years. The simple implication of this is that from 2017-2020, about 70% of Nigerian students might have not gotten the opportunity to secure admission into the various higher institutions of learning.

The issue of educational measurement in research point towards enhancing the equity of test or examination across sub groups of examinees is very essential in decisions based on scores of the examinees. Test consists of a set of uniform items

to which an examinee is to respond independently and the result of which can be treated in such a way as to provide a quantitative comparison of the achievement among different students (Faremi and Jimoh, 2022). Testing is done to determine whether or not an objective or goal has been obtained or achieved. In other words, testing concerns specific achievement of a student in terms of a given objective. By implication, test items have to measure the same thing for individuals from different groups who have the same subject ability. However, as viewed by Okoye, *et al.,* (2021), using test items that measure different things from different subgroups of examinees that are of the same latent traits is contrary to the principle of an egalitarian society that emphasized no deprivation of any kind. Examinees of the same latent trait should respond to test item correctly irrespective of their gender, school type, school ownership and school location.

A test is an instrument for systematic measure of examinees behaviour. A test as seen by Okagbare, Ossai and Osadebe (2023) is referred to as a series of items/tasks or a set of questions that learners respond to orally or in writing that makes it possible to examine differences between learners. Tests are used to gain useful information about examinees' knowledge, skills and progress; it helps each professional to perform work effectively. Tests are used in promotion, placement, selection, certification and decision making (Ihendinihu, 2022). Sometimes, the results of these tests are incorrect due to differences in test performances among various groups of testes. A fair test is one which is comparably valid for all groups of examinees and that affords all testes equal opportunity to demonstrate the skills and knowledge that have been acquired which are relevant to the tests purpose. As test results are often the basis for decision that affects student's educational future, test should provide the same and equal opportunities for all students to demonstrate their skills, ability and knowledge (Okoye, Fejokwu and Ikeanumba, 2021). It

therefore becomes necessary to avoid bias which may negatively influence examinees scores. The existence of differential item functioning is an issue to be addressed because tests are used as a gateway for educational purposes and it is very necessary that test items be fair for all examinees irrespective of their background and locations.

Several studies have been conducted in Nigeria to isolate the factors responsible for the observed poor performance in SSCE. Esomonu and Ikeanumba, (2021) found factors such as assessment factors; teachers' pedagogical content knowledge (Jacob, *et al.,* 2020); measurement errors and measurement practices (Ikeanumba and Esomonu, 2024; Nduka, Ikeanumba and Fejokwu, 2023; Okoye and Nduka, 2022) among others. However, researchers also came up with various recommendations which were intended to improve students' performance in NECO SSCE. These include reformation of better pattern of analysis of findings; better ways of estimating measurement errors, teacher training education in Nigeria; priority being given to pedagogy study; better assessment patterns; conferences, workshops, seminars on the need and importance of teachers' self-efficacy to effectiveness; among others. Several authors recommended that qualified and sufficient teachers should be employed in order to have an improved performance from the students. Okoye and Fejokwu (2021) explained that various factors affecting the teaching and learning of Mathematics and Economics in Nigeria especially at the secondary school level include political instability, state of economy and academic problems. The researcher stressed further that the academic problems include students' unparalleled hatred and indifference among others. The researcher further recommended that students' interest in Mathematics, Economics other science subjects should be stimulated deliberately by teacher during classes.

Despite solutions being proffered by assessment experts, test evaluation experts, developers, educators and researchers, the poor performance of students at the NECO SSCE persisted. An aspect that has not received much research attention in NECO SSCE in Nigeria is how assessment practices affect tests' utility and ultimately examinees' test performance. Some researchers have explored the relative advantage of item response theory (IRT) for scoring achievement test (Nduka, Ikeanumba and Fejokwu, 2023; Okoye & Nduka, 2022; and Okoye & Fejokwu, 2021). They found out from the result of their research that there was no significant difference between candidates' test performance estimated with the classical test theory (traditional scoring approach) and IRT scoring approach. The psychometric quality of NECO SSCE test items is yet to be fully explored by researchers. This involves the assessment practice which includes the psychometric analysis adopted by examiners in ensuring that developed tests have utility and examinees' test scores represent the true ability of examinees in the subject of interest. Simply put, test's utility refers to how useful a test is. Reliability and validity are the fundamental facets of the psychometric quality; as such the nature and strength of the reliability and validity of items are keys to assessing its psychometric quality (Ikeanumba & Esomonu, 2024). Although validity and reliability address respectively the precision of test scores and the degree to which test scores can be interpreted in terms of its specific psychological construct, they are concerned with how well a test performs among identifiable groups in a population (Kelani and Faleye 2022). A number of considerations are involved in making a judgment about the utility of a test.

By collecting data in a representative administration context and attending to dimensionality, reliability and validity, researchers and test developers enhance the possibility that the scale or test will be useful and psychologically informative (Ihendinihu, 2022; and Ogunsanmi, Ibikunle, & Shogbesan, 2023). However, the

present study is limited to reliability. Test reliability shows how consistent a measure is of a particular element over a period of time, and between different participants. For example, a test measuring personality traits should yield the same answers for a subject after several times completing the test, and with a short period of time between (so long as the individual has not inherently changed personality traits). The reliability of a test according to Okoye and Fejokwu (2021) is important, specifically when dealing with psychometric tests; there is no point in having a test that will yield different answers each time measured, particularly when it can influence the decisions of employers and who they may employ to lead their company. In educational and psychological testing, reliability refers to the precision of the measurement process, or the consistency of scores produced by a test (Aituariagbon and Osarumwense, 2022). Several psychometric analyses have been adopted into the assessment framework of public examining bodies in Nigeria including Differential Item Functioning (DIF). This is because the assessment framework, classical test theory (CTT) being used by the examining bodies in the course of test development as explained by (Ikeanumba, Esomonu and Anikpe, 2024) does not measure ability which is the basic ground on which the differential functioning of item can adequately be assessed.

Differential item functioning is somewhat a neutral term used to refer to differences in the statistical properties (discrimination, difficulty and or guessing indices) of an item between groups of examinees of equal ability. Differential item functioning refers to the situation where members from different groups (age, gender, race, education, culture) on the same level of the latent trait (disease severity, quality of life) have a different probability of giving a certain response to a particular item (Okoye and Fejokwu, 2021). No doubt, a test whose item displays different statistical properties among identifiable subgroups in a population will definitely

result to compromised test scores. This may be one of the reasons why students are performing poorly in NECO SSCE level. Thus, research should focus towards assessing the presence of DIF in the SSCE being used by public examining bodies in the assessment of students' performance and ultimately its effects on the utility of the test and examinees' test performance. According to Okagbare, *et al.,* (2023), DIF is a phenomenon that arises when the probability of answering an item correctly is independent of true ability but dependent on membership to a group. It occurs if different sub-member of a group, who are of equal standing on the concept the test is designed to measure, display different probabilities of passing an item or of endorsing an item. This was supported by Okoye, *et al.,* (2022) who explained that DIF occurs when an item is not equally difficult or equally popular for groups that have been matched in terms of the concept being measured. According to the authors, Differential item functioning analyses involve a judgment of the performances of subgroups that have been matched in terms of the relevant construct or concept and hence do not require equal test scores for the groups that are involved. DIF occurs when examinees from different groups have different likelihoods of success on an item, after they have been matched on the ability of interest (Clauser and Mazor, 2018). The presence of DIF is as a result of some characteristics in an item that result in differential performance for individuals of equal ability but from different groups. Items may be judged relatively more or less difficult for a particular group by comparison with the performance of another group drawn from the same population. Differential item functioning of an item can therefore be understood as a lack of conditional independence between an item response and group membership (often gender, location or ethnicity) given the same latent ability or trait (Osawe, 2021). IRT which is a method of DIF require larger sample size and more computational load for parameter estimation. Non-IRT methods are only better than

IRT methods when the sample size is small (e.g., 100), because they do not require item parameter estimation. However, without item parameter estimation, it is more difficult to figure out the source of DIF when an item is flagged as a DIF item. A major identified problem associated with the use of non-IRT method of DIF assessment, is that observed score, which may not be reliable are used as an indicator of ability level (Okoye and Nduka, 2021). Therefore, in the present study, the IRT based method of DIF assessment is used.

In Nigeria, positive academic achievement at any level of education is recognized with certification of those who successfully completed a course of study with good academic records. Academic test results are important yardsticks by which society pronounces the product of its educational system. As such, it is of paramount importance that educational institutions and examination bodies should conduct tests that will enable them establish the desired characteristics of their examinees. Similarly, for any examination to be judged good; the conduct of the examination and the published results should be deemed fair and achieve high level of acceptance by the public. More so, the examination should ensure that no particular candidate or candidates has/have an unfair advantage over others irrespective of their sex, location, social-economic status, school type, cultural/ethnicity among others.

Researchers posit that test with differential item functioning could bring about underachievement of minority groups in a subject area. Questions of differential item endorsement and related concerns about fair test use and interpretation have demanded a considerable amount of the attention of measurement specialists. Researchers observed that it is not socially or politically viable to use different decision rules for the members of different groups on the same matter. Test with differential item functioning distorts the purpose of the test result and the decision that is based on it for some groups. It is against this background that the researchers

deemed it fit to assess differential item functioning in National Examinations Council test and its impact on the tests' reliability using multiple choice questions.

Research Questions

The following research questions guided the study:

1. What are the item parameters of 2023 NECO multiple-choice tests in each of Mathematics and Economics for male and female populations?
2. What are the reliability estimates of 2023 NECO multiple-choice tests in Mathematics and Economics when differentially functioning items are included and the estimates when they are excluded?

Hypotheses

The following null hypotheses were tested at 0.05 level of significance

1. Items of 2023 NECO multiple-choice tests in Mathematics and Economics do not function differently with respect to gender.
2. There is no significant difference between the reliability estimate of the 2017 NECO multiple-choice test in each of Mathematics and Economics, obtained when differentially functioning items are included, and the reliability estimates obtained when differentially functioning items are excluded.

Materials and Methods

The study employed an ex-post-facto research design. The population included all candidates who took the 2023 NECO Mathematics and Economics examinations in Nigeria. Sampling was conducted in two stages: selecting the subjects and then choosing the participants. Data collection utilized the 2023 NECO past questions for these four subjects, which had already been validated by NECO's Examinations and Standard Unit in Nigeria. Since NECO provides these standardized instruments, the researchers did not independently determine their reliability. Candidate responses to selected test items were collected directly from

NECO for analysis. The research questions were analyzed using an item response theory parameter logistic model with Marginal Maximum Likelihood estimation, along with mean and standard deviation calculations. Hypothesis 1 was tested through Thissen's Likelihood Ratio Test (LRT) for DIF, and Hypothesis 2 was examined using a test of equality of alpha coefficients at a 0.05 significance level.

Results

Research Question 1: What are the item parameters of 2023 NECO multiple-choice tests in Mathematics and Economics for male and female populations?

This question will be answered with respect to each subject and the subjects are taken up one after the other in the sections that follows. In each of the tables, the letters "a" "b" and 'c" stand for item discrimination index, item difficulty index and guessing factor respectively.

Table 1: Item Parameters of 2023 for Mathematics Multiple Choice Test in Male and Female Examinees Population

	Female			Male		
Item	**A**	**B**	**C**	**A**	**B**	**C**
IT1	1.05	-1.52	0.00	1.09	-1.48	0.00
IT2	0.43	1.61	0.25	0.44	2.50	0.33
IT3	0.93	0.32	0.62	1.06	0.69	0.70
IT4	1.46	-0.95	0.00	1.59	-0.96	0.00
IT5	0.14	-4.10	0.10	0.05	-6.56	0.15
IT6	1.55	0.31	0.67	1.77	0.41	0.70
IT7	1.53	-0.79	0.00	1.60	-0.81	0.00
IT8	1.81	-0.24	0.50	1.89	-0.27	0.52
IT9	1.59	-0.57	0.00	1.66	-0.56	0.00
IT10	1.67	0.13	0.39	1.90	0.14	0.44
IT11	1.84	-0.53	0.00	1.94	-0.55	0.00
IT12	1.47	-0.24	0.00	1.62	-0.29	0.00
IT13	-0.14	-18.38	0.00	-0.08	-33.26	0.00
IT14	1.51	-0.38	0.02	1.59	-0.44	0.02
IT15	1.28	-0.26	0.00	1.36	-0.23	0.00
IT16	2.14	-0.01	0.44	2.31	0.00	0.48
IT17	1.66	-0.76	0.00	1.76	-0.73	0.00
IT18	0.48	0.02	0.00	0.54	0.21	0.00

IT19	1.54	-0.41	0.68	1.70	-0.23	0.72
IT20	1.23	-1.03	0.03	1.30	-1.00	0.05
IT21	0.50	0.88	0.00	0.56	0.77	0.00
IT22	1.50	-0.92	0.00	1.67	-0.84	0.00
IT23	0.34	-0.23	0.00	0.29	-0.13	0.00
IT24	1.81	-0.25	0.31	1.86	-0.21	0.35
IT25	0.08	0.21	0.01	-0.03	-6.56	0.06
IT26	1.14	-1.10	0.00	1.18	-1.04	0.00
IT27	1.00	-0.28	0.00	1.00	-0.26	0.00
IT28	0.46	1.43	0.00	0.45	1.26	0.00
IT29	0.54	0.94	0.00	0.60	0.96	0.00
IT30	1.49	-0.18	0.44	1.57	-0.06	0.48
IT31	1.67	-0.45	0.00	1.84	-0.43	0.00
IT32	1.94	-0.55	0.00	2.11	-0.53	0.00
IT33	1.71	-0.61	0.00	1.89	-0.57	0.00
IT34	1.91	-0.67	0.00	2.09	-0.64	0.00
IT35	1.86	-0.74	0.03	1.96	-0.71	0.02
IT36	1.49	-0.27	0.00	1.53	-0.26	0.00
IT37	1.70	0.07	0.08	1.74	0.03	0.06
IT38	2.83	0.18	0.42	3.00	0.16	0.43
IT39	2.48	0.07	0.44	2.56	0.05	0.43
IT40	1.42	-0.12	0.01	1.47	-0.15	0.00
IT41	2.36	0.23	0.41	2.64	0.20	0.42
IT42	1.69	0.23	0.12	1.82	0.23	0.12
IT43	1.63	-0.61	0.05	1.82	-0.58	0.05
IT44	0.30	-0.76	0.00	0.27	-0.67	0.00
IT45	1.68	-0.46	0.16	1.83	-0.36	0.22
IT46	1.63	-1.01	0.00	1.74	-0.96	0.00
IT47	0.78	-0.37	0.00	0.80	-0.32	0.00
IT48	2.19	-0.02	0.33	2.36	-0.01	0.37
IT49	1.92	-0.16	0.47	2.11	-0.11	0.49
IT50	1.15	-0.28	0.16	1.25	-0.26	0.20
IT51	1.03	-1.78	0.00	1.15	-1.58	0.00
IT52	1.45	-1.47	0.00	1.53	-1.40	0.00
IT53	2.19	-0.04	0.50	2.61	0.06	0.55
IT54	1.38	-1.04	0.00	1.41	-1.06	0.00
IT55	0.36	1.46	0.00	0.34	1.82	0.00
IT56	1.80	0.13	0.42	2.05	0.21	0.47
IT57	1.30	-1.00	0.00	1.30	-1.01	0.00
IT58	1.30	-0.99	0.00	1.29	-1.07	0.00
IT59	1.23	-0.64	0.00	1.31	-0.64	0.01
IT60	0.58	-0.15	0.00	0.57	-0.12	0.00
Mean	**1.35**	**-0.65**	**0.13**	**1.44**	**-1.00**	**0.15**
STD	**0.63**	**2.47**	**0.21**	**0.70**	**4.45**	**0.22**

Table 1 shows the item parameters of 2023 NECO Mathematics test in the male and female populations. The table shows that on average the test items showed better discrimination power in male population (a = 1.44) than in the female population (a = 1.35). However, in terms of difficulty level, the test items were more difficult for the female students (b = -0.65) than for the male students (b = -1.00). With respect to the vulnerability of guessing of the test items, the result shows that the test items were more vulnerable to guessing in the male population (c = 0.15) than in the female counterpart (c = 0.13). The result revealed that the item parameters of 2023 NECO Mathematics test differs among male and female populations. The result implies that item parameters of 2023 NECO Mathematics differ with respect to the gender of examinees that sat for the test.

Table 2: Item Parameters of 2023 NECO Economics Multiple Choice Test in Male and Female Examinees Population

	Female			Male		
Items	**a_1**	**b_1**	**c_1**	**a_2**	**b_2**	**c_2**
IT1	1.22	-1.25	0.00	1.14	-1.45	0.00
IT2	1.93	-0.60	0.40	1.92	-0.64	0.45
IT3	1.94	-0.44	0.38	1.90	-0.51	0.43
IT4	-0.64	-2.28	0.00	-0.80	-2.03	0.00
IT5	10.59	0.60	0.20	12.92	0.60	0.20
IT6	1.23	-1.36	0.00	1.24	-1.48	0.00
IT7	1.56	-1.38	0.00	1.59	-1.49	0.00
IT8	1.83	-0.79	0.11	1.73	-1.06	0.02
IT9	1.89	-0.88	0.02	1.91	-1.03	0.00
IT10	0.98	-1.92	0.00	1.13	-1.82	0.00
IT11	1.78	8.44	0.51	-0.44	0.20	0.00
IT12	1.66	-1.13	0.01	1.65	-1.26	0.00
IT13	1.52	-1.86	0.00	1.52	-1.96	0.00
IT14	35.45	0.55	0.34	39.19	0.56	0.34
IT15	28.63	0.57	0.22	34.05	0.57	0.22
IT16	-8.51	-2.72	0.15	-8.12	-2.68	0.14
IT17	1.73	-1.15	0.00	1.79	-1.20	0.00
IT18	1.48	-0.74	0.01	1.51	-0.82	0.01
IT19	1.19	-1.49	0.00	1.25	-1.52	0.00
IT20	1.80	-0.78	0.34	1.69	-0.86	0.40
IT21	1.77	-0.78	0.33	1.80	-0.82	0.38

IT22	2.04	-0.93	0.04	1.98	-1.11	0.00
IT23	-0.73	-2.16	0.00	-0.81	-2.06	0.00
IT24	-0.66	-16.59	0.26	0.68	21.99	0.23
IT25	1.55	-0.12	0.43	1.51	-0.20	0.43
IT26	0.01	95.56	0.03	-0.11	-4.63	0.00
IT27	1.63	-0.93	0.00	1.60	-1.03	0.00
IT28	1.95	-1.05	0.05	1.81	-1.24	0.00
IT29	1.48	-0.92	0.07	1.53	-1.06	0.00
IT30	-0.19	0.18	0.00	-0.33	0.23	0.00
IT31	2.09	-0.65	0.09	2.03	-0.80	0.05
IT32	1.63	-1.66	0.00	1.68	-1.70	0.00
IT33	1.78	-0.79	0.01	1.83	-0.93	0.00
IT34	1.30	-1.47	0.00	1.39	-1.43	0.00
IT35	1.31	-1.80	0.00	1.36	-1.84	0.00
IT36	31.07	0.57	0.26	35.54	0.57	0.26
IT37	1.78	-1.09	0.02	1.85	-1.16	0.00
IT38	1.61	-1.04	0.00	1.69	-1.08	0.00
IT39	-0.25	-6.82	0.00	-0.33	-5.18	0.00
IT40	36.27	0.55	0.36	39.74	0.55	0.35
IT41	33.94	0.57	0.30	35.53	0.57	0.29
IT42	-0.17	0.45	0.00	-0.32	0.36	0.00
IT43	31.61	0.58	0.20	33.97	0.58	0.19
IT44	1.71	-0.85	0.07	1.82	-0.89	0.05
IT45	2.05	-0.72	0.04	2.12	-0.84	0.03
IT46	1.45	-1.84	0.00	1.52	-1.81	0.00
IT47	1.96	-0.49	0.33	1.98	-0.58	0.35
IT48	1.60	-0.84	0.04	1.63	-0.98	0.00
IT49	1.64	-0.89	0.00	1.80	-0.93	0.00
IT50	1.09	-1.59	0.00	1.13	-1.59	0.00
IT51	2.09	-0.57	0.30	2.08	-0.63	0.36
IT52	1.47	-1.30	0.00	1.33	-1.46	0.00
IT53	2.00	-1.23	0.00	1.97	-1.33	0.00
IT54	1.77	-1.20	0.00	1.79	-1.30	0.00
IT55	1.68	-1.27	0.00	1.69	-1.33	0.00
IT56	1.00	-1.12	0.00	0.91	-1.20	0.00
IT57	0.79	-1.64	0.00	0.73	-1.69	0.00
IT58	0.03	273.86	0.49	-0.02	-1.86	0.03
IT59	0.02	10.79	0.06	-0.16	-30.00	0.50
IT60	-0.03	-31.32	0.01	-0.09	-8.36	0.00
Mean	**4.41**	**4.78**	**0.11**	**4.78**	**-1.30**	**0.10**
STD	**9.780**	**37.805**	**0.153**	**10.839**	**5.016**	**0.156**

Table 2 shows the item parameters of 2023 NECO Economics test in the male and female populations. The table shows that on average the test items showed better discrimination power in male population (a = 4.78) than in the female population (a

= 4.41). However, in terms of difficulty level, the test items were more difficult for the female students (b = 4.78) than for the male students (b = -1.30). With respect to the vulnerability of guessing of the test items, the result shows that the test items were more vulnerable to guessing in the female population (c = 0.11) than in the male counterpart (c = 0.10). The result revealed that the item parameters of 2023 NECO Economics test differ among male and female populations. The result implies that item parameters of 2023 NECO Economics differ with respect to gender of examinees that sat for the test.

Research Question 2: What are the reliability estimates of 2023 NECO multiple-choice tests in Mathematics and Economics when differentially functioning items are included and the estimates when they are excluded?

Table 3: Reliability Estimates of Tests when DIF Items are Excluded and when they are Included

S/NO	SUBJECTS	DIF FREE ITEMS	WITH DIF ITEMS
1.	MATHEMATICS	0.92	0.90
2.	ECONOMICS	0.95	0.95

Reliability Estimates of Mathematics Multiple-Choice Test with and Without DIF

Table 3 shows the reliability estimates of 2023 NECO Mathematics test with DIF and DIF free. The table shows that the reliability estimate of the Mathematics test when the items showing DIF were removed (r = 0.92) than when the reliability was estimated with DIF items (r = 0.90). The implication of the result is the reliability estimate of 2023 NECO Mathematics multiple-choice test exhibiting DIF differ from the reliability estimates of the 2023 NECO multiple choice test without DIF.

Reliability Estimates of Economics Multiple-Choice Test with and Without DIF

Table 3 also shows the reliability estimates of 2023 NECO Economics test with DIF and DIF free. The table shows that the reliability estimate of the Economics test when the items showing DIF were removed (r = 0.95) was the same as the

reliability estimate of the Economics test when it was estimated with DIF items (r = 0.95). The result showed that the reliability of 2023 NECO Economics test exhibiting DIF is the same as the reliability estimate of the test free of DIF. The result implies that there is no difference between the reliability estimates of 2023 NECO Economics multiple-choice test exhibiting DIF and the DIF free version of the test.

Hypotheses

The following null hypotheses were tested at 0.05 level of significance

Hypothesis 1: Items of 2023 NECO multiple-choice tests in Mathematics and Economics, do not function differently with respect to gender.

Table 4: Differential Item Functioning for Mathematics Test

	Female			Male				
Item	**A**	**B**	**C**	**A**	**B**	**C**	**UA**	**Remark**
IT1	1.05	-1.52	0.00	1.09	-1.48	0.00	0.045	NO DIF
IT2	0.43	1.61	0.25	0.44	2.50	0.33	0.668	DIF
IT3	0.93	0.32	0.62	1.06	0.69	0.70	0.141	NO DIF
IT4	1.46	-0.95	0.00	1.59	-0.96	0.00	0.046	NO DIF
IT5	0.14	-4.10	0.10	0.05	-6.56	0.15	9.615	DIF
IT6	1.55	0.31	0.67	1.77	0.41	0.70	0.037	NO DIF
IT7	1.53	-0.79	0.00	1.60	-0.81	0.00	0.029	NO DIF
IT8	1.81	-0.24	0.50	1.89	-0.27	0.52	0.016	NO DIF
IT9	1.59	-0.57	0.00	1.66	-0.56	0.00	0.023	NO DIF
IT10	1.67	0.13	0.39	1.90	0.14	0.44	0.036	NO DIF
IT11	1.84	-0.53	0.00	1.94	-0.55	0.00	0.029	NO DIF
IT12	1.47	-0.24	0.00	1.62	-0.29	0.00	0.067	NO DIF
IT13	-0.14	-18.38	0.00	-0.08	-33.26	0.00	14.936	DIF
IT14	1.51	-0.38	0.02	1.59	-0.44	0.02	0.061	NO DIF
IT15	1.28	-0.26	0.00	1.36	-0.23	0.00	0.045	NO DIF
IT16	2.14	-0.01	0.44	2.31	0.00	0.48	0.016	NO DIF
IT17	1.66	-0.76	0.00	1.76	-0.73	0.00	0.038	NO DIF
IT18	0.48	0.02	0.00	0.54	0.21	0.00	0.250	NO DIF
IT19	1.54	-0.41	0.68	1.70	-0.23	0.72	0.058	NO DIF
IT20	1.23	-1.03	0.03	1.30	-1.00	0.05	0.043	NO DIF
IT21	0.50	0.88	0.00	0.56	0.77	0.00	0.198	NO DIF
IT22	1.50	-0.92	0.00	1.67	-0.84	0.00	0.090	NO DIF
IT23	0.34	-0.23	0.00	0.29	-0.13	0.00	0.422	NO DIF

IT24	1.81	-0.25	0.31	1.86	-0.21	0.35	0.028	NO DIF
IT25	0.08	0.21	0.01	-0.03	-6.56	0.06	37.421	DIF
IT26	1.14	-1.10	0.00	1.18	-1.04	0.00	0.061	NO DIF
IT27	1.00	-0.28	0.00	1.00	-0.26	0.00	0.000	NO DIF
IT28	0.46	1.43	0.00	0.45	1.26	0.00	0.170	NO DIF
IT29	0.54	0.94	0.00	0.60	0.96	0.00	0.152	NO DIF
IT30	1.49	-0.18	0.44	1.57	-0.06	0.48	0.067	NO DIF
IT31	1.67	-0.45	0.00	1.84	-0.43	0.00	0.048	NO DIF
IT32	1.94	-0.55	0.00	2.11	-0.53	0.00	0.038	NO DIF
IT33	1.71	-0.61	0.00	1.89	-0.57	0.00	0.057	NO DIF
IT34	1.91	-0.67	0.00	2.09	-0.64	0.00	0.045	NO DIF
IT35	1.86	-0.74	0.03	1.96	-0.71	0.02	0.034	NO DIF
IT36	1.49	-0.27	0.00	1.53	-0.26	0.00	0.017	NO DIF
IT37	1.70	0.07	0.08	1.74	0.03	0.06	0.037	NO DIF
IT38	2.83	0.18	0.42	3.00	0.16	0.43	0.014	NO DIF
IT39	2.48	0.07	0.44	2.56	0.05	0.43	0.012	NO DIF
IT40	1.42	-0.12	0.01	1.47	-0.15	0.00	0.033	NO DIF
IT41	2.36	0.23	0.41	2.64	0.20	0.42	0.026	NO DIF
IT42	1.69	0.23	0.12	1.82	0.23	0.12	0.030	NO DIF
IT43	1.63	-0.61	0.05	1.82	-0.58	0.05	0.055	NO DIF
IT44	0.30	-0.76	0.00	0.27	-0.67	0.00	0.311	NO DIF
IT45	1.68	-0.46	0.16	1.83	-0.36	0.22	0.085	NO DIF
IT46	1.63	-1.01	0.00	1.74	-0.96	0.00	0.055	NO DIF
IT47	0.78	-0.37	0.00	0.80	-0.32	0.00	0.053	NO DIF
IT48	2.19	-0.02	0.33	2.36	-0.01	0.37	0.019	NO DIF
IT49	1.92	-0.16	0.47	2.11	-0.11	0.49	0.031	NO DIF
IT50	1.15	-0.28	0.16	1.25	-0.26	0.20	0.050	NO DIF
IT51	1.03	-1.78	0.00	1.15	-1.58	0.00	0.204	NO DIF
IT52	1.45	-1.47	0.00	1.53	-1.40	0.00	0.072	NO DIF
IT53	2.19	-0.04	0.50	2.61	0.06	0.55	0.054	NO DIF
IT54	1.38	-1.04	0.00	1.41	-1.06	0.00	0.022	NO DIF
IT55	0.36	1.46	0.00	0.34	1.82	0.00	0.364	NO DIF
IT56	1.80	0.13	0.42	2.05	0.21	0.47	0.052	NO DIF
IT57	1.30	-1.00	0.00	1.30	-1.01	0.00	0.000	NO DIF
IT58	1.30	-0.99	0.00	1.29	-1.07	0.00	0.080	NO DIF
IT59	1.23	-0.64	0.00	1.31	-0.64	0.01	0.040	NO DIF
IT60	0.58	-0.15	0.00	0.57	-0.12	0.00	0.036	NO DIF

Table 4 shows the differential item functioning assessment of the 60 Items of 2023 NECO Mathematics multiple-choice test. The table shows that Item 2

functioned differentially with respect to gender (Raju statistic = 0.668, > 0.6). Thus, the hypothesis that "2023 NECO Mathematics multiple-choice test items do not exhibit differential functioning with respect to gender" was rejected for item 2. The same trend of the result was obtained for other four items. They are 5, 13, and 25. The implication is that four of the 60 items of the 2023 NECO Mathematics test functioned differentially with respect to gender.

Table 5: Differential Item Functioning for Economics Test

	Female			Male				
	A	**b**	**g**	**A**	**B**	**G**	**UA**	**Remark**
IT1	2.16	0.23	0.01	2.56	0.19	0.00	0.067	NO DIF
IT2	2.21	-0.23	0.00	2.65	-0.19	0.00	0.071	NO DIF
IT3	2.94	0.26	0.05	3.21	0.21	0.04	0.050	NO DIF
IT4	2.38	-0.10	0.01	2.63	-0.10	0.00	0.032	NO DIF
IT5	0.17	9.51	0.00	0.15	12.28	0.00	2.762	DIF
IT6	2.13	-0.46	0.00	2.46	-0.41	0.00	0.070	NO DIF
IT7	2.19	-0.35	0.00	2.54	-0.28	0.00	0.079	NO DIF
IT8	2.51	-0.17	0.00	2.95	-0.12	0.00	0.062	NO DIF
IT9	2.79	0.03	0.82	3.10	0.15	0.85	0.020	NO DIF
IT10	2.40	-0.58	0.00	2.86	-0.47	0.00	0.108	NO DIF
IT11	1.66	-1.07	0.00	2.05	-0.82	0.00	0.258	NO DIF
IT12	1.92	-0.35	0.00	2.30	-0.29	0.00	0.088	NO DIF
IT13	-2.60	-1.05	0.06	-2.56	-0.85	0.04	0.194	NO DIF
IT14	2.33	-0.38	0.00	2.74	-0.29	0.00	0.095	NO DIF
IT15	2.41	0.13	0.00	2.92	0.14	0.00	0.061	NO DIF
IT16	2.89	0.24	0.06	3.36	0.21	0.05	0.043	NO DIF
IT17	5.41	0.53	0.58	6.36	0.52	0.64	0.010	NO DIF
IT18	2.24	0.78	0.80	2.84	0.76	0.80	0.016	NO DIF
IT19	1.66	-0.93	0.00	2.07	-0.66	0.00	0.282	NO DIF
IT20	2.42	-0.30	0.00	2.69	-0.26	0.00	0.044	NO DIF
IT21	2.60	-0.43	0.00	3.04	-0.33	0.00	0.103	NO DIF
IT22	1.99	-0.13	0.00	2.37	-0.09	0.00	0.073	NO DIF
IT23	2.77	-0.30	0.01	3.13	-0.25	0.01	0.055	NO DIF
IT24	2.22	0.09	0.04	2.58	0.06	0.03	0.054	NO DIF
IT25	2.73	-0.34	0.00	3.13	-0.31	0.00	0.044	NO DIF
IT26	2.22	-0.95	0.00	2.64	-0.76	0.00	0.195	NO DIF
IT27	11.56	0.69	0.46	13.65	0.62	0.51	0.039	NO DIF
IT28	5.84	0.71	0.53	6.13	0.70	0.58	0.006	NO DIF
IT29	5.81	0.71	0.55	6.39	0.68	0.61	0.014	NO DIF
IT30	2.00	-0.42	0.00	2.37	-0.33	0.00	0.102	NO DIF

IT31	2.40	-0.62	0.00	2.87	-0.49	0.00	0.133	NO DIF
IT32	3.13	-0.13	0.04	3.45	-0.14	0.03	0.023	NO DIF
IT33	2.82	-0.47	0.00	3.30	-0.40	0.00	0.073	NO DIF
IT34	2.97	0.12	0.01	3.35	0.13	0.00	0.032	NO DIF
IT35	0.10	31.17	0.00	0.09	34.85	0.00	3.674	DIF
IT36	4.76	0.47	0.70	5.84	0.47	0.72	0.010	NO DIF
IT37	2.93	0.18	0.04	3.28	0.14	0.03	0.044	NO DIF
IT38	2.92	0.20	0.00	3.44	0.17	0.00	0.048	NO DIF
IT39	2.63	0.25	0.00	3.12	0.21	0.00	0.063	NO DIF
IT40	2.76	0.17	0.06	3.17	0.16	0.05	0.036	NO DIF
IT41	5.84	0.68	0.59	6.99	0.65	0.63	0.016	NO DIF
IT42	3.36	0.02	0.04	3.72	0.00	0.03	0.026	NO DIF
IT43	2.70	-0.01	0.01	3.06	0.00	0.00	0.035	NO DIF
IT44	2.85	0.10	0.01	3.19	0.09	0.01	0.030	NO DIF
IT45	2.03	-0.32	0.00	2.59	-0.20	0.00	0.140	NO DIF
IT46	-0.19	-15.86	0.00	-0.14	-21.87	0.00	6.011	DIF
IT47	2.22	-0.28	0.02	2.53	-0.25	0.01	0.052	NO DIF
IT48	2.69	-0.47	0.00	3.10	-0.41	0.00	0.071	NO DIF
IT49	-2.26	-0.98	0.00	-2.60	-0.79	0.00	0.196	NO DIF
IT50	2.29	-0.59	0.00	2.79	-0.43	0.00	0.159	NO DIF
IT51	2.22	-0.77	0.00	2.83	-0.52	0.00	0.253	NO DIF
IT52	2.31	-0.15	0.00	2.76	-0.09	0.00	0.074	NO DIF
IT53	2.82	-0.10	0.00	3.30	-0.07	0.00	0.048	NO DIF
IT54	2.90	0.00	0.04	3.24	-0.02	0.02	0.032	NO DIF
IT55	2.41	-0.21	0.00	2.88	-0.15	0.00	0.073	NO DIF
IT56	-0.85	-0.64	0.00	-1.21	-0.44	0.00	0.327	NO DIF
IT57	2.02	-0.55	0.00	2.52	-0.40	0.00	0.161	NO DIF
IT58	-1.63	-0.94	0.02	-1.99	-0.71	0.00	0.233	NO DIF
IT59	2.05	-0.42	0.00	2.48	-0.28	0.00	0.143	NO DIF
IT60	1.85	0.01	0.00	2.34	0.04	0.00	0.097	NO DIF

Table 5 shows the differential item functioning assessment of the 60 Items of 2023 NECO Economics multiple-choice test. The table shows that item 5 functioned differentially with respect to gender (Raju statistic = 2.762, > 0.6). Thus, the hypothesis that "2023 NECO Economics multiple-choice test items do not exhibit differential functioning with respect to gender" was rejected for item 5. The same trend of the result was obtained for another two items. They are 35 and 46. The

implication is that three out of 60 items on the 2023 NECO Economics test functioned differentially with respect to gender.

Hypothesis 2: There is no significant difference between the reliability estimate of the 2017 NECO multiple-choice test in each of Mathematics and Economics obtained when differentially functioning items are included, and the reliability estimates obtained when differentially functioning items are excluded.

Table 6: Comparison of Reliability Estimates of 2023 NECO Mathematics Test when DIF Items are included and when they are excluded

	Reliability Estimate	Chisq	Df	p-value
DIF free Items	0.92	4563631.3366	1	0.000
with DIF Items	0.90			

Table 6 shows the comparison of reliability estimate of NECO in Mathematics when DIF items were included and the estimate when the DIF items were excluded. The table shows that the reliability estimate, when the DIF items were excluded (0.92) is greater than the estimate when they were included (r = 0.90). Furthermore, p-value shows that the difference observed in the reliability estimates of the test under the two conditions was significantly different (χ = 4563631.3366 (df = 1, p-value = 0.000). Thus, the hypothesis of no significant difference was rejected with respect to gender.

Table 7: Comparison of Reliability Estimates of 2023 NECO Economics Test when DIF Items are included and when they are excluded

	Reliability Estimate	Chisq	Df	p-value
DIF free Items	0.95	0.0000	1	1.000
with DIF Items	0.95			

Table 7 shows the comparison of reliability estimate of 2023 NECO in Economics when DIF items were included and the estimate when the DIF items were excluded. The table shows that the reliability estimate, when the DIF items were excluded (0.95) is equal to the estimate when they were included (r = 0.95).

Furthermore, p-value shows that the difference observed in the reliability estimates of the test under the two conditions was significantly different ($\chi = 0.0000$, df = 1, p-value = 1.000). Thus, the hypothesis of no significant difference was not rejected with respect to gender.

Discussion and Conclusion

The findings of the study are discussed as follows:

Item Parameters of 2023 NECO Multiple-Choice Tests in Mathematics and Economics for Male and Female Populations

The finding of this study revealed that on the average, the 2023 NECO Mathematics multiple test items showed better discrimination power in male population than in the female population but were more difficult for female than male populations in terms of difficulty level and that males were more vulnerable to guessing than female counterparts. This could be seen in Table 1. Finding of this study again showed that on the average, the 2023 NECO Economics test items showed better discrimination power in male population than in the female population and also more difficult for male students than for the female students in terms of difficulty level. Interestingly, the vulnerability of guessing of the test items, were the same among the male female populations. This was observed in Table 2. In other words, the findings on this study revealed that fact that item parameters of 2023 NECO Mathematics and Economics multiple choice test items differs substantially with respect to gender of examinees that sat for the test. These findings corroborate the finding of Okoye, *et al.,* (2021) which investigated uni-dimensionality and occurrence of DIF in English language and Mathematics items of Delta State Qualifying Examination and found the occurrence of DIF items in both English language and Mathematics multiple-choice items of the OSQE for 2017. The finding on this study also conforms to that of Okagbare, *et al.,* (2023). Thus, this could make

the items measure what they were not supposed to measure. The conclusions of related studies in relation to this study have lay credence to the fact that 2023 NECO Mathematics and Economics multiple-choice tests items contained an appreciable number of items that exhibited DIF could have a major implication on examinees that were not of the same levels. Again, the finding of this study revealed that only 4 of the 60 items on the 2023 NECO Mathematics, and 3 of the 60 items on the 2023 NECO Economics multiple-choice test functioned differentially with respect to gender. Thus, the null hypotheses that 2023 NECO Mathematics and Economics multiple-choice test items do not exhibit differential item functioning with respect to gender were rejected because there were significant differences in examinees' performance in the tests exhibiting DIF.

Reliability Estimates of 2023 NECO Multiple-Choice Tests in Mathematics and Economics when Differentially Functioning Items are Included and the Estimates when they are Excluded

The finding of this study revealed that the reliability estimates of 2023 NECO Mathematics and Economics multiple-choice tests items exhibiting DIF differ from the reliability estimates of the 2023 NECO Mathematics and Economics multiple-choice tests without DIF based on gender. The finding of this study supports that of Okoye, *et al.,* (2021) and Faremi and Jimoh (2022) which found that the greatest DIF effect was less than 2 points on the 0 to 60 total score scale and about 0.15 on the IRT ability scale but that DIF had a limited effect on the ratio of true score variance to observed-score variance (i.e., test reliability). This study has revealed the impact of reliability estimates on examinees in 2023 NECO examinations which could result in the unreliable ability estimates of the examinees. This finding advanced that the implication of neglecting reliability estimates in the NECO examinations or any other examinations for placement could result in lower

performance in examinees future academic endeavours because the scores used were those obtained directly from the examination body. The finding of this study is consistent with the finding of Abba (2015) that out of 100 items 64 were biased. 36 favoured male while 4 favoured female students. The result showed that the items set and administered by NECO/SSCE 2010 significantly and consistently favoured males. The finding of this study agrees with the findings of Sa'ad, *et al.,* (2020), which found that multiple–choice English language items administered by NECO SSSCE2014 functioned differently between male and female students in which sixty-nine (69) out of hundred were flagged gender bias. Sixty-one (61) were in favour of male and four (08) in favour of female students respectively. These findings have shown that 2023 NECO multiple-choice tests items of examinations bodies are not void of differential item functioning which could actually mislead and impact negatively on the performance of the examinees. This means that examines scores in such NECO examinations are determined largely by the group to which an examinee belongs and not by ability.

Conclusion

The findings of this study revealed that item parameters of 2023 NECO Mathematics and Economics multiple-choice test items differ substantially with respect to the gender of examinees that sat for the tests as there were no consistent discriminating and difficulty level as well as vulnerability in the population. This study also revealed that the reliability estimates of 2023 NECO Mathematics and Economics multiple-choice tests items exhibiting DIF differ from the reliability estimates of the 2023 NECO Mathematics and Economics multiple-choice tests without DIF based on gender.

Recommendations

The following recommendations are made in light of the findings of this study:

1. Public examination bodies should ensure that their tests are subjected to DIF analysis to have quality and fair tests items.
2. Public examination bodies should employ both Classical Test and Item Response theory frameworks while conducting DIF analysis of items in all subjects.
3. Public examining bodies staff should be exposed to capacity building training on how the reliability of tests used by them for the assessment and certification of students' performance in school subjects be improved upon by identifying and eliminating items that function differentially with respect to identifiable subpopulations.
4. All tests should be developed by observing the stipulated procedure, appropriate item analysis as well as ensure that test items are fair to all groups of examinees.
5. Test experts and developers should explore the use of DIF approach to detect biased items.

References

Aituariagbon, K. E & Osarumwense, H. J (2022). Non-parametric method of detecting differential item functioning in senior school certificate examination (SSCE) 2019 Economics multiple-choice items. *Kashere Journal of Education, 3* (1):146 - 158.

Clauser, B. E., & Mazor, K. M. (2018). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, *17*(1), 31 - 44.

Enuku, U. E. & Enuku, C. A. (2019). Incorporating AIDS Education in Prison Adult Education Programs in Nigeria. *Journal of Correctional Education*. *50* (3), 96 – 100

Esomonu, N.P.M & Ikeanumba, C.B. (2021). Effect of formative assessment on academic achievement of postgraduate students in advanced educational statistics in public universities in southeast, Nigeria. *International Journal of Advanced Education and Research, 6*(4), 05 - 11.

Faremi, Y.A. & Jimoh, K. (2022). Differential Item Functioning and implications for testing in Nigeria Education System. *Indonesian Journal of Learning Education and Counseling, 5*(1). 1 - 10.

Ihendinihu, U.E. (2020). Assessment of secondary school mathematics teacher's knowledge and utilization of multiple-choice test construction guidelines and procedures. *ASSEREN Journal of Education*, *5* (1):175 - 190.

Ihendinihu, U.E. (2022). Cognitive levels, task numbers and facility indices of Mathematics multiple-choice questions in basic education certificate examination conducted by National Examination Council in Nigeria. *Rivers State University Journal of Education,*25:1 - 11.

Ikeanumba, C.B., Esomonu, N.P.M. & Anike, K.L. (2024). Estimating measurement error of senior secondary school Economics examinations using generalizability theory in Imo State. *Unizik Journal of Educational Research and Policy Studies, 17* (3), 236 - 245

Jacob, F., John, S., & Gwany, D. M. (2020). Teachers' pedagogical content knowledge and students' academic achievement: A theoretical overview. *Journal of Global Research in Education and Social Science*, *14*(2), 14 – 44.

Jimoh, K & Adediwura, A.A (2022). Gender and culture-related differential item functioning in 2016 National Examinations Council Mathematics multiple-choice questions in Nigeria. *Ife Journal of Behavioral Research, 10* (1), 55 - 72.

Kelani, K.O. & Faleye, B.A (2022). Differential Item Functioning of 2016 multiple-choice Economics items of Osun State Unified Promotion Examination. *Ife Journal of Behavioral Research 10* (1), 1 - 9.

Nduka, S.V., Ikeanumba, C.B. & Fejokwu, I.B. (2024). Calibrating the item parameters of 2020 NECO Mathematics multiple-choice questions using

Classical Test Theory Framework. Prof. Ngozi, N, Agu, Christiana, A. Ugodulunwa & Nkechi, P.M. Esomonu (Eds). *Readings in Teaching Pedagogy, Educational Evaluation and Research: Festschrift in Honour of an Academic Legend Prof. Romy Okoye*, 148 - 154.

Ogunsanmi, E. C., Ibikunle, A. H. & Shogbesan, P. I. (2023). Differential item functioning of 2015-2017 Agricultural Science multiple-choice questions conducted by the National Business Technical Examination Board. *Journal of Evaluation, 7* (1), 11 - 24

Ogunsanmi, O.A., Ibikunle, Y.A. & Shogbesan, Y.O. (2023). Evaluation of Differential Item Functioning (DIF) in 2017 West African Examinations Council (WAEC) Physics multiple choice test in Osun State, Nigeria. *Al-Hikmah Journal of Arts & Social Sciences Education*, *5*(1), 114 - 121.

Okagbare, F., Ossai, P.A.U. & Osadebe, P.U. (2023). Assessment of Differential Item Functioning in Physics Multiple Choice used by WAEC among Senior Secondary School Students in Delta State. *West African Journal of Interdisciplinary Research*, *1*(2), 1 - 20.

Okoye R. O. & Fejokwu B. I. (2021). Determination of Differential Item Functioning by gender in the National Business and Technical Examinations Board (NABTEB) 2017 Physics multiple choice examination in Delta State. *International Journal of Advanced Education and Research, 6* (3), 18 - 22.

Okoye, R. O., Fejokwu, I. B. & Ikeanumba, C.B. (2021). Analysis of content validity of teacher-made tests for physics in Sapele Local Government Area, Delta State, Nigeria. *International Journal of Advanced Education and Research, 6(6), 15 - 19.*

Okoye, R.O. & Nduka, S.V. (2022). Application of Item Response Theory in calibrating 2020 NECO Mathematics multiple-choice questions. *International Journal of Advanced Education and Research*, *7*(1). 1 - 5

Osawe, O. U. (2021). *Test item fairness in Economics Achievement Test*. Unpublish M.Ed. Delta State University, Abaka, Nigeria.

Sa'ad, N., Ali, A. A. & Abdullahi, A. (2020). Differential item functioning of 2014 NECO English Language Examination in North Senatorial District of Kano State, Nigeria. *Kano Journal of Educational Psychology* (KaJEP), *2* (2), 2736 – 1373 (Online).