International Journal of Innovative Social & Science Education Research 12(1):92-103, Jan.-Mar., 2024

© SEAHI PUBLICATIONS, 2024 www.seahipublications.org ISSN: 2360-8978

Detection Of Gender-Based Uniform And Non-Uniform DIF Of The NECO Mathematics Multiple-Choice Questions In Imo State, Nigeria

Solomon Chukwu Ohiri

Directorate of Academic Planning,
Alvan Ikoku University of Education, Owerri, Nigeria
E-mail: ohiri.chukwu2018@gmail.com

ABSTRACT
The research aimed at examining whether the 2020, 2021 and 2022 National Examination Council (NECO) June/July mathematics multiple-choice questions exhibited uniform and non-uniform gender-related differential item functioning (DIF) in Imo State. A survey research design was employed. The population was made up of all senior secondary school three (SS3) students of 2022/2023 academic session. The number of sampled candidates used in the study was 2,484 students. This comprised 1,178 male and 1,306 female students. Three research questions were formulated to guide the study. The instruments used for the study were the 2020, 2021 and 2022 June/July multiple-choice mathematics questions set by the National Examination Council (NECO). Each of the instruments consists of 60-items. To detect uniform and non-uniform differentially functioned items by gender, a software called STATA 15 of the logistics regression which is one of the classical test theory methods of DIF detection was applied. The results of the analyses revealed that some items functioned differentially based on gender. Sixteen items (32%) in 2020, sixteen items (32%) in 2021 and twelve items (24%) in 2022, functioned differentially according to the gender of the students. Among the items that displayed DIF, 18% displayed uniform and 18% also displayed non-uniform DIF in 2020. In 2021, 14% and 18% displayed uniform and non -uniform DIF respectively; while in 2022, 16% and 8% were flagged uniform and non-uniform DIF respectively. Sequel to the findings of the study, it was concluded that the NECO June/July mathematics multiple-choice questions of 2020, 2021 and 2022 sets were not free from uniform and non-uniform differential item functioning (DIF). It was also recommended among others that National Examination Council should carry out differential item functioning analysis for all test items as part of test development process.
Keywords: Uniform DIF, Non-uniform DIF, Gender, Logistic Regression, Assessment.

INTRODUCTION

Fairness in assessment of students' achievement in mathematics in our secondary schools is sacrosanct. Fairness is an essential quality of a test; its equitable treatment of all examinees during the testing process, absence of measurement bias, equitable access to the constructs being measured, and justifiable validity of test score interpretation for the intended purpose (Effiom, 2021).

In the pursuance of test fairness, a fundamental requirement of measurement is that test scores should be valid. A test should measure examinee's ability accurately without regard to the examinee's membership in any demographic group. Effort should be made to ensure that sources of variation do not put any subgroup at a disadvantage. If extraneous sources of variance are distributed differently for identifiable subgroups on a test item, the item is considered biased. The presence of bias is a cause for concern

because tests are used to evaluate latent traits such as skills, abilities, and other psychometric characteristics. Nkwocha (2019) said that test is an instrument used to find out whether an object or person possesses a particular attribute or characteristic. Public examination bodies, like the National Examination Council (NECO), use achievement test to access the ability level of candidates who write these examinations. Tests from these examination bodies are expected not to favour any group of testees, either male or female, or other groups for any reason.

Test results can be affected by the test takers' demographic characteristics. Every test taker belongs to a subgroup. Test takers' answer to items tend to be influenced by their demographic membership (Anna-Brew and Cobbinah, 2020). Tests should provide equal opportunities to all examinees without bias to demonstrate their abilities and knowledge irrespective of their socio-demographic factors like gender, location, religious and cultural groups (Amaechi, Eluwa & Madu, 2020). The primary concern in test development and test use as Karami (2012) suggested, is demonstrating that the interpretations and uses we make of test scores are valid. As a result, a test needs to be fair to different test takers. A test is considered fair if different sub-groups of testees with equal ability have the same probability of correct responses to the test items. According to Ogbebor and Onuka (2013), a fair test is a test that enables all examinees to have an equal chance to demonstrate the skills and knowledge which they have acquired and which are vital to the purpose of the test. Ordinarily, it is expected that two individuals at the same level of a latent trait or ability, regardless of what group they belong to, will have the same probability of correctly or affirmatively responding to an item. If this is not true for an item, the item is said to be functioning differentially. Differential item functioning (DIF) occurs when individuals of the same ability level from separate groups have different probabilities of answering an item correctly (Annan-Brew, 2020). DIF is an indicator of bias observed when test takers from different groups have different probabilities or likelihood of responding correctly to an item, after controlling for ability. When a test has the element of differential item functioning (DIF), it shows that the test is not fair to some of the examinees or subgroups. According to Zwick in Oladele, Adegoke and LongJohn (2020), differential item functioning (DIF) analysis is a key component in the evaluation of the fairness and validity of educational tests.

Generally, Differential item functioning is of two types: uniform and non-uniform. Uniform or unidirectional DIF exists when the probability of endorsing an item (answering an item correctly) is greater for one group than for the other group over all the levels of proficiency. The existence of non-uniform or crossing DIF demonstrates that the difference in probabilities of a correct response is not the same at all levels of proficiency between the two comparison groups. That is, the probability of correctly answering an item is higher for one group at some points on the scale, and higher for the other group at other points (Ibrahim, 2017).

In uniform DIF, one group is advantaged throughout the range of the ability, that is, for the advantaged group, the probability of correctly responding to an item is consistently greater than for the other group. Items that show uniform DIF have different difficulty parameters for the groups, meaning that the item is more difficult for one of the groups throughout the ability continuum. In the case when an item shows non-uniform DIF, members of one group find the item more difficult in one part of the ability continuum and then this is reversed for another part of the ability continuum. Items that function differently in a non-uniform way, have different discrimination parameters and potentially different difficulty parameters.

In the analysis of uniform and non-uniform DIF, there are two major psychometric theories, which are classical test theory and item response theory and their corresponding models have been used for addressing DIF studies. The current study is anchored on the classical test theory (CTT). In the classical test theory, the main concern of the item analysis is to describe the statistical characteristics of each item. The total score of a test is considered the sum of scores on the individual items, and the individual item is of interest through its effect on the total test score. Thus, item analysis in classical test theory is focused on the degree to which each item influences the whole measurement.

Different studies have been conducted to investigate the presence or absence of uniform and non-uniform DIF in test items with regards to gender with diverse findings. Ibrahim, Aminu, Yalwaji and Sani (2022) carried out a study on differential item functioning in Senior Secondary School Certificate Examination

multiple-choice items: Implication for technology adoption in large scale assessment in Kano state Nigeria. They discovered that 26 items (43% of the 60 items) in NECO 2017 Mathematics multiple-choice test were found to reveal uniform and non-uniform DIF. Seven (7) out of 26 items displayed uniform DIF; while 19 items exhibited non-uniform DIF. Abedalaziz (2010) found in a study using logistic regression that 10 of the 30 items of the tenth grade students' Mathematics in Jordan at the end of the First semester, school year 2009 – 2010, displayed uniform DIF that favoured the male group, while eight (8) items revealed non-uniform.

Annan- Brew and Cobbinah (2020) carried out a gender-related differential item functioning of 2015 WASSCE core mathematics results in Southern Ghana using Logistic Regression procedure. The results showed that forty-three (43) items or 86% of the items revealed DIF. Out of the 43 items, 9 items revealed statistically significant uniform DIF, whereas 34 items revealed statistically significant non-uniform DIF. The nine items that revealed statistically uniform DIF had 5 items in favour of male candidates and 4 items in favour of female candidates, while the 34 items that showed statistically significant non-uniform DIF had 18 items in favour of male candidates and 16 items in favour of female candidates. Ong, Williams and Lamprianou (2015) explored crossing differential item functioning (DIF) in a test drawn from a national examination of mathematics from 11-years-old pupils in England. An empirical dataset was analyzed to explore DIF by gender in a mathematics assessment. A two-step process involving the logistic regression (LR) procedure for detecting uniform and non-uniform DIF was applied to identify crossing DIF. The results showed 36 uniform (out of the 60 items) and 19 non-uniform statistically significant gender DIF items. Out of the 19 non-uniform DIF items, 10 items were crossing DIF. The 36 items that revealed significant uniform DIF had 25 in favour of males and 11 in favour of females, whereas the 19 items that revealed significant non-uniform DIF had 7 in favour of males and 5 in favour of females.

The empirical studies gathered information related to the present study. Some of these studies were done outside Nigeria. In particular, none of these studies reviewed has explored DIF on basis of uniform and non-uniform gender group membership in Imo State.

Purpose of the Study

The purpose of this study was to investigate the level to which Uniform and Non-uniform DIF exists in mathematics achievement multiple-choice questions administered by the National Examination Council (NECO) for senior secondary school certificate, ranging from 2020 - 2022 in terms of gender.

Research Questions

The following research questions guided the study

1. What percentage of the items in each year showed DIF in favour of each gender?
2. What percentage of the items in each year had uniform DIF?
3. What percentage of the items in each year had non-uniform DIF?

METHOD

The research design adopted for this study was survey research design. The appropriateness of the design stemmed from the postulation of Nworgu (2015), that survey research design is one in which a group of people or items is studied by collecting and analyzing data from only a few people or items considered to be a representative of the entire group. This design was considered appropriate because only a part of the population was studied and findings were used to generalize for the entire population. The Population of the study comprised all the SS3 students of the public secondary schools in Imo State in the 2022/2023 academic session. A combination of purposive sampling and simple random sampling was used for the sample selection. First, co-educational secondary schools were purposively sampled to ensure that students from both gender groups came from the same schools. From the co-educational secondary schools, two schools were obtained through simple random sampling from each Educational Zone. This gave rise to ten co-educational secondary schools. All SS3 students in the ten (10) sampled schools were used for the study, giving rise to 2484 students (1, 178 males and 1, 306 females). The instruments used for data collection were the 2020, 2021 and 2022 NECO June/July multiple-choice mathematics questions. Each of the instruments consisted of 60-items, and each item consists of a stem and a list of

possible answers lettered A - E of which only one option is the correct answer. Each item of the instruments was scored 1 for correct option and 0 for wrong option with maximum score of 60 and minimum of 0 for the entire instruments. The instruments had been validated by experts in the Test Development Division of the National Examination Council (NECO) and therefore required no further validation since they were adopted. Thus, the items were considered appropriate in terms of subject contents and instructional objectives. On the other hand, being instruments of standardized international examination, which were conducted by the National Examination Council (NECO), the instruments were deemed reliable. Hence, the reliability of the instruments was not established by the researchers. To collect pertinent data needed for the study, the instruments were administered to the SS3 students in each of the sampled schools with the help of the mathematics teachers, who served as the research assistants. The researchers, through the teachers, informed the students ahead of time about the exercise and the need to be prepared because it would form part of their continuous assessment. This measure was to ensure that the students put in their best. A software called STATA 15 was used to estimate the item parameters for the reference group (male students) and focal group (female students) for the determination of the differential item functioning (DIF) of the items. Data were analyzed using model equation for logistic regression so as to detect uniform and non-uniform differentially functioned items in terms of gender. The logistic regression model consists of two stages. First, the control variable, usually the "classical" total score, was included in the regression equation. Then, two other variables, related to the group (male and female) and the interaction group score, were included in the equation. The analysis consists in testing if the insertion of these two variables leads to a significant statistical result. For an item to be classified as displaying DIF, the two degrees of freedom Chi-squared test in logistic regression needed to have a p-vale less than or equal to 0.05 (Oratokhai, 2021)

RESULTS AND DISCUSSIONS

Research Question 1. ***What percentage of the items in each year showed DIF in favour of each gender?***

To answer Research Question 1, the odds ratio of logistic regression for DIF was conducted based on gender. Female students in the study were referred to as focal group (coded 1), while male students were referred to as reference group (coded 0). Differential item functioning occurs in favour of female students when the odds ratio of logistic regression test (LRT) of an item is significant ($p < .05$) and the odds ratio is greater than 1. The results obtained from the analysis are presented in Table 1, while the summary is presented in Table 2.

Table 1. Items That Showed DIF in Favour of Each Gender in 2020, 2021 and 2022 June/July NECO Mathematics Multiple-Choice Questions

	NECO 2020			NECO 2021			NECO 2021		
Item	Odds Ratio	p-value	Remarks	Odds Ratio	p-value	Remarks	Odd Ratio	p-value	Remarks
1	0.75	0.44	No DIF	1.11	0.75	No DIF	0.65	0.48	No DIF
2	0.78	0.45	No DIF	1.68	0.02	DIF Female	1.14	0.85	No DIF
3	1.40	0.28	No DIF	1.95	0.03	DIF Female	1.07	0.93	No DIF
4	0.69	0.02	DIF Male	0.75	0.35	No DIF	0.82	0.04	DIF Male
5	0.80	0.37	No DIF	1.39	0.27	No DIF	0.55	0.05	No DIF
6	1.18	0.48	No DIF	0.52	0.00	DIF Male	1.10	0.78	No DIF
7	1.06	0.87	No DIF	0.73	0.02	DIF Male	1.93	0.91	No DIF
8	1.20	0.48	No DIF	1.14	0.01	DIF Female	1.40	0.26	No DIF
9	1.17	0.52	No DIF	1.48	0.00	DIF Female	0.78	0.42	No DIF
10	1.25	0.37	No DIF	0.69	0.00	DIF Male	1.07	0.90	No DIF

11	0.96	0.96	No DIF	1.20	0.65	No DIF	0.84	0.59	No DIF
12	0.79	0.37	No DIF	1.67	0.04	DIF Female	0.85	0.62	No DIF
13	0.76	0.01	DIF Male	0.58	0.04	DIF Male	0.91	0.82	No DIF
14	0.96	0.97	No DIF	0.72	0.31	No DIF	1.34	0.01	DIF Female
15	0.84	0.50	No DIF	1.07	0.95	No DIF	1.21	0.56	No DIF
16	0.71	0.15	No DIF	1.57	0.42	No DIF	1.01	0.92	No DIF
17	1.03	0.96	No DIF	1.65	0.06	No DIF	1.39	0.26	No DIF
18	0.86	0.56	No DIF	0.80	0.46	No DIF	0.68	0.21	No DIF
19	1.02	0.98	No DIF	1.97	0.00	DIF Female	0.57	0.01	DIF Male
20	0.92	0.79	No DIF	0.69	0.00	DIF Male	0.85	0.64	No DIF
21	0.91	0.78	No DIF	1.18	0.54	No DIF	0.94	0.92	No DIF
22	0.74	0.21	No DIF	0.92	0.85	No DIF	1.13	0.74	No DIF
23	1.02	0.98	No DIF	0.91	0.83	No DIF	3.05	0.00	DIF Female
24	1.51	0.08	No DIF	1.24	0.49	No DIF	1.83	0.02	DIF Female
25	1.45	0.11	No DIF	0.83	0.56	No DIF	0.52	0.01	DIF Male
26	0.72	0.00	DIF Male	1.62	0.19	No DIF	0.87	0.71	No DIF
27	1.26	0.36	No DIF	0.89	0.78	No DIF	0.88	0.76	No DIF
28	1.01	0.03	DIF Female	0.91	0.87	No DIF	0.91	0.84	No DIF
29	0.81	0.01	DIF Male	0.74	0.33	No DIF	1.05	0.95	No DIF
30	1.81	0.02	DIF Female	1.11	0.08	No DIF	1.47	0.20	No DIF
31	1.47	0.14	No DIF	0.64	0.02	DIF Male	1.36	0.27	No DIF
32	1.81	0.01	DIF Female	1.06	0.90	No DIF	0.77	0.45	No DIF
33	1.12	0.65	No DIF	0.54	0.06	No DIF	1.02	0.95	No DIF
34	1.25	0.33	No DIF	0.87	0.71	No DIF	0.72	0.28	No DIF
35	1.11	0.71	No DIF	0.73	0.30	No DIF	0.51	0.03	DIF Male
36	1.26	0.36	No DIF	1.34	0.27	No DIF	1.17	0.61	No DIF
37	0.73	0.00	DIF Male	1.51	0.18	No DIF	1.24	0.46	No DIF
38	1.33	0.24	No DIF	1.41	0.03	DIF Female	0.81	0.52	No DIF
39	0.86	0.02	DIF Male	0.64	0.12	No DIF	1.21	0.57	No DIF
40	1.04	0.90	No DIF	1.45	0.17	No DIF	1.54	0.14	No DIF
41	1.01	0.93	No DIF	0.52	0.09	No DIF	1.56	0.14	No DIF
42	0.67	0.01	DIF Male	0.87	0.66	No DIF	1.77	0.00	DIF Female
43	0.90	0.73	No DIF	0.57	0.07	No DIF	0.80	0.02	DIF Male
44	1.66	0.02	DIF Female	0.63	0.02	DIF Male	0.59	0.07	No DIF
45	0.85	0.56	No DIF	0.63	0.10	No DIF	1.75	0.04	DIF Female
46	1.74	0.01	DIF Female	1.60	0.07	No DIF	0.64	0.13	No DIF
47	1.14	0.60	No DIF	1.37	0.27	No DIF	0.76	0.37	No DIF
48	0.97	0.98	No DIF	1.08	0.82	No DIF	0.99	0.01	DIF Male
49	1.18	0.48	No DIF	0.69	0.20	No DIF	1.05	0.94	No DIF
50	1.34	0.23	No DIF	1.28	0.36	No DIF	1.07	0.90	No DIF
51	1.81	0.02	DIF Female	1.11	0.08	No DIF	1.47	0.20	No DIF

52	1.47	0.14	No DIF	0.64	0.02	DIF Male	1.36	0.27	No DIF
53	1.81	0.01	DIF Female	1.06	0.90	No DIF	0.77	0.45	N DIF
54	1.12	0.65	No DIF	0.54	0.06	No DIF	1.02	0.95	No DIF
55	1.25	0.33	No DIF	0.87	0.71	No DIF	0.72	0.28	No DIF
56	1.11	0.71	No DIF	0.73	0.30	No DIF	0.51	0.03	DIF Male
57	1.26	0.36	No DIF	1.34	0.27	No DIF	1.17	0.61	No DIF
58	0.73	0.00	DIF Male	1.51	0.18	No DIF	1.24	0.46	No DIF
59	1.33	0.24	No DIF	1.41	0.03	DIF Female	0.81	0.52	No DIF
60	0.86	0.02	DIF Male	0.64	0.12	No DIF	1.21	0.57	No DIF

Table 2. Percentage of Items That Displayed DIF in Favour of Male and Female Students in 2020, 2021 and 2022 NECO June/July Mathematics Multiple-Choice Questions

Year	NO. of Items	DIF	DIF in favour of male	DIF in favour of Female
2020	50	16	9 (18 %)	7 (14 %)
2021	50	16	8 (16 %)	8 (16 %)
2022	50	12	7 (14 %)	5 (10 %)

The outcome presented in Table 2 shows that out of 50 items in NECO 2020 June/July Mathematics multiple-choice test questions, 16 possess DIF. Out of the 16 items that were flagged DIF, 9 (Items 4, 13, 26, 29, 37, 39, 42, 58 and 60 representing 18 %) were in favour of male students and 7 (Items 28, 30, 32, 44, 46, 51 and 53 representing 14 %) were in favour of female students. In 2021, 16 items out of 50 items were flagged DIF. Out of the 16 that were flagged DIF, 8 (Items 6, 7, 10, 13, 20, 31, 44 and 52, representing 16 %) were in favour of male students, while 8 (Items 2, 3, 8, 9, 12, 19, 38 and 59, representing 16 %) were in favour of female students. In 2022, 12 items were flagged DIF out of the 50 items. Out of the 12 items flagged DIF, 7 (Items 4, 19, 25, 35, 43, 48 and 56, representing 14 %) were in favour of male students while 5 (Items 14, 23, 24, 42 and 45, representing 10 %) were in favour of female students.

When test possesses DIF, it could bring about low achievement for a minority group in a subject matter and this can hamper the meaning of test outcomes and decision that is based on it for some groups, especially core subject like mathematics which is a compulsory criterion for further educational advancement. Therefore, it is important that tests be fair to all and not biased against any group.

Research Question 2. ***What percentage of the items in each year had uniform DIF?***

To answer Research Question 2, logistic regression for uniform DIF was conducted based on gender. Uniform differential item functioning occurs when Uniform Logistic Regression Test (ULRT) of an item is significant ($p < .05$). The results obtained from the analysis are presented in Table 3, while the summary is presented in Table 4.

Table 3. Uniform DIF Analysis of 2020, 2021 and 2022 NECO June/July Mathematics Multiple-Choice Questions Based on Gender

	NECO 2020			NECO 2021			NECO 2022		
Item	ULRT	P-value	Remarks	ULRT	P-value	Remarks	ULRT	P-value	Remarks
1	0.23	0.63	No DIF	0.45	0.50	No DIF	0.63	0.42	No DIF
2	0.02	0.89	No DIF	8.58	0.00	Uniform DIF	0.09	0.76	No DIF
3	3.36	0.06	No DIF	8.60	0.00	Uniform DIF	0.25	0.62	No DIF
4	2.39	0.12	No DIF	0.38	0.53	No DIF	0.49	0.48	No DIF
5	0.91	0.34	No DIF	2.92	0.08	No DIF	1.86	0.17	No DIF
6	1.77	0.18	No DIF	0.39	0.01	Uniform	0.19	0.66	No DIF

						DIF			
7	0.10	0.74	No DIF	0.47	0.01	Uniform DIF	3.77	0.05	No DIF
8	0.53	0.46	No DIF	0.76	0.38	No DIF	1.60	0.20	No DIF
9	0.89	0.34	No DIF	2.75	0.09	No DIF	0.51	0.47	No DIF
10	1.01	0.31	No DIF	0.13	0.72	No DIF	0.43	0.51	No DIF
11	0.01	0.91	No DIF	0.03	0.87	No DIF	0.01	0.92	No DIF
12	1.38	0.23	No DIF	2.31	0.12	No DIF	0.03	0.86	No DIF
13	4.73	0.02	Uniform DIF	0.29	0.02	Uniform DIF	0.03	0.86	No DIF
14	0.02	0.87	No DIF	3.84	0.05	No DIF	2.30	0.12	No DIF
15	0.53	0.46	No DIF	0.01	0.92	No DIF	0.24	0.62	No DIF
16	0.77	0.38	No DIF	0.78	0.37	No DIF	0.05	0.82	No DIF
17	0.83	0.36	No DIF	5.36	0.12	No DIF	3.35	0.06	No DIF
18	1.65	0.19	No DIF	0.63	0.42	No DIF	2.32	0.12	No DIF
19	0.03	0.86	No DIF	6.79	0.00	Uniform DIF	0.80	0.00	Uniform DIF
20	0.50	0.48	No DIF	0.97	0.32	No DIF	0.35	0.55	No DIF
21	0.00	0.96	No DIF	0.18	0.66	No DIF	0.14	0.71	No DIF
22	3.24	0.07	No DIF	0.07	0.78	No DIF	0.35	0.55	No DIF
23	0.62	0.43	No DIF	0.15	0.69	No DIF	21.31	0.00	Uniform DIF
24	2.91	0.08	No DIF	0.61	0.43	No DIF	6.64	0.00	Uniform DIF
25	3.35	0.06	No DIF	0.02	0.89	No DIF	0.79	0.00	Uniform DIF
26	4.00	0.04	Uniform DIF	0.49	0.48	No DIF	0.03	0.86	No DIF
27	3.55	0.05	No DIF	0.18	0.67	No DIF	0.05	0.83	No DIF
28	0.01	0.91	No DIF	0.14	0.70	No DIF	0.21	0.64	No DIF
29	1.06	0.30	No DIF	1.45	0.22	No DIF	0.02	0.89	No DIF
30	0.75	0.38	No DIF	0.58	0.44	No DIF	1.44	0.23	No DIF
31	2.38	0.12	No DIF	0.49	0.48	No DIF	1.60	0.20	No DIF
32	5.11	0.02	Uniform DIF	0.04	0.84	No DIF	0.50	0.47	No DIF
33	0.00	0.98	No DIF	1.07	0.30	No DIF	0.00	0.95	No DIF
34	0.69	0.40	No DIF	1.27	0.25	No DIF	0.75	0.38	No DIF
35	0.00	0.99	No DIF	0.36	0.55	No DIF	0.92	0.01	Uniform DIF
36	0.65	0.42	No DIF	0.41	0.52	No DIF	0.77	0.37	No DIF
37	5.69	0.01	Uniform DIF	1.14	0.28	No DIF	1.92	0.16	No DIF
38	0.04	0.83	No DIF	0.14	0.70	No DIF	0.35	0.55	No DIF
39	0.72	0.02	No DIF	1.48	0.22	No DIF	0.09	0.76	No DIF
40	0.39	0.53	No DIF	1.83	0.17	No DIF	2.85	0.09	No DIF
41	0.18	0.67	No DIF	1.02	0.31	No DIF	2.59	0.10	No DIF
42	5.80	0.01	Uniform DIF	0.14	0.70	No DIF	0.85	0.35	No DIF
43	0.00	0.97	No DIF	0.90	0.34	No DIF	2.16	0.14	No DIF
44	5.43	0.01	Uniform DIF	0.42	0.01	Uniform DIF	3.19	0.07	No DIF
45	0.00	0.97	No DIF	0.59	0.44	No DIF	5.39	0.02	Uniform DIF
46	5.87	0.01	Uniform DIF	0.81	0.36	No DIF	2.34	0.12	No DIF
47	1.67	0.19	No DIF	1.29	0.25	No DIF	1.14	0.28	No DIF

48	0.01	0.90	No DIF	1.11	0.29	No DIF	0.03	0.04	Uniform DIF
49	2.35	0.12	No DIF	1.16	0.14	No DIF	0.07	0.78	No DIF
50	1.87	0.17	No DIF	1.37	0.24	No DIF	0.16	0.69	No DIF
51	0.79	0.50	No DIF	1.28	0.27	No DIF	0.78	0.34	No DIF
52	0.00	0.99	No DIF	0.39	0.75	No DIF	5.91	0.11	No DIF
53	0.69	0.41	No DIF	0.42	0.57	No DIF	0.71	0.33	No DIF
54	5.59	0.21	No DIF	1.17	0.24	No DIF	1.93	0.18	No DIF
55	0.04	0.84	No DIF	0.18	0.71	No DIF	0.65	0.51	No DIF
56	0.76	0.02	No DIF	1.48	0.23	No DIF	0.09	0.04	Uniform DIF
57	0.49	0.51	No DIF	1.88	0.27	No DIF	2.75	0.08	No DIF
58	0.19	0.57	No DIF	1.04	0.35	No DIF	2.49	0.12	No DIF
59	5.70	0.02	No DIF	0.17	0.77	No DIF	0.95	0.33	No DIF
60	0.00	0.03	Uniform DIF	0.91	0.44	No DIF	2.26	0.18	No DIF

Table 4. Percentage of Items that Displayed Uniform DIF in NECO 2020, 2021 and 2022 June/July Mathematics Multiple-Choice Questions

Year	No. of Items	Uniform DIF	Percentage of items with Uniform DIF	Items with uniform DIF
2020	50	8	16 %	13, 26, 32, 37, 42, 44, 46, 60
2021	50	7	14 %	2, 3, 6, 7, 13, 19, 44
2022	50	8	16 %	19, 23, 24, 25, 35, 45, 48, 56

The outcome presented in Table 4 shows that 8 items, that is, Items 13, 26. 32, 37, 42, 44, 46 and 60 representing 16% of the total items in NECO 2020 June/July mathematics multiple-choice test displayed uniform differential item functioning, with Items 13, 26, 37, 42 and 60 favouring male students, while Items 32, 44 and 46 favoured female students. In 2021, 7 items, that is, Items 2, 3, 6, 7, 13, 19 and 44 representing 14% of the total items displayed uniform differential item functioning, with Items 6, 7, 13 and 44 favouring male students, while Items 2, 3 and 19 favoured female students. In 2022, 8 items, that is, Items 19, 23, 24, 25, 35, 45, 48 and 56 representing 16% of the total items displayed uniform differential item functioning, with Items 19, 25, 35, 48 and 56 favouring male students, while Items 23, 24 and 45 favoured female students.

These findings proved the effectiveness of Binary logistic regression in detecting uniform and non-uniform DIF. In uniform DIF, the item favours the advantaged group, while the other group is less favoured with respect to difficulty of the item(s) at different ability levels of the examinees. When an item possesses uniform DIF, all the individuals in one group perform better on an item than all the individuals in the other group. This can cause discouragement to the disadvantaged group. The results of the analyses for the three consecutive years are in agreement with the previous research work of Ibrahim, Aminu, Yalwaji and Sani (2022) in which they discovered that 26 items (43% of the 60 items) in NECO 2017 Mathematics multiple-choice test were found to reveal uniform DIF. Seven (7) out of 26 items displayed uniform DIF. Again, the outcomes of the current study agree with the works of Annan-Brew and Cobbinah (2020) in which the results showed that forty-three (43) items or 86% of the items of 2015 WASSCE core mathematics results in Southern Ghana, revealed DIF. Out of the 43 items, only 9 items revealed statistically significant uniform DIF, of which, 5 items were in favour of male candidates and 4

items were in favour of female candidates. The current study also conforms with the works of Ong, Williams and Lamprianou (2015) in which the results showed 36 items (out of the 60 items) had uniform DIF. The 36 items that revealed significant uniform DIF had 25 in favour of males

Research Question 3. ***What percentage of the items in each year had non-uniform DIF?***
To answer Research Question 3, logistic regression for non-uniform DIF was conducted based on gender. Non-uniform differential item functioning occurs when Non-Uniform Logistic Regression Test (NLRT) value of an item is significant ($p < .05$). The results obtained from the analysis are presented in Table 5, while the summary is presented in Table 6.

Table 5
Non-Uniform DIF Analysis of 2020, 2021 and 2022 NECO June/July Mathematics Multiple-Choice Questions Based on Gender.

	NECO 2020			NECO 2021			NECO 2022		
Item	NLRT	P-value	Remarks	NLRT	P-value	Remarks	NLRT	P-value	Remarks
1	2.48	0.11	No DIF	0.28	0.59	No DIF	0.13	0.72	No DIF
2	0.82	0.36	No DIF	0.42	0.51	No DIF	0.03	0.86	No DIF
3	0.12	0.72	No DIF	1.90	0.16	No DIF	0.28	0.59	No DIF
4	0.88	0.02	Non-uniform DIF	0.09	0.76	No DIF	0.92	0.00	Non-uniform DIF
5	1.66	0.19	No DIF	0.13	0.72	No DIF	0.78	0.37	No DIF
6	1.15	0.28	No DIF	1.44	0.22	No DIF	2.67	0.10	No DIF
7	0.02	0.87	No DIF	0.49	0.48	No DIF	0.15	0.69	No DIF
8	0.34	0.56	No DIF	6.59	0.01	Non-uniform DIF	0.01	0.92	No DIF
9	0.11	0.73	No DIF	7.56	0.01	Non-uniform DIF	0.33	0.56	No DIF
10	0.29	0.58	No DIF	0.36	0.00	Non-uniform DIF	0.75	0.38	No DIF
11	0.13	0.71	No DIF	2.53	0.11	No DIF	0.89	0.34	No DIF
12	1.90	0.16	No DIF	4.78	0.02	Non-uniform DIF	0.09	0.76	No DIF
13	0.43	0.51	No DIF	3.40	0.06	No DIF	0.94	0.33	No DIF
14	0.01	0.91	No DIF	0.66	0.41	No DIF	6.18	0.01	Non-uniform DIF
15	1.41	0.23	No DIF	2.40	0.12	No DIF	3.76	0.05	No DIF
16	1.10	0.29	No DIF	2.22	0.13	No DIF	0.97	0.32	No DIF
17	0.00	0.95	No DIF	4.47	0.43	No DIF	1.93	0.16	No DIF
18	0.46	0.49	No DIF	3.19	0.07	No DIF	1.80	0.18	No DIF
19	2.83	0.09	No DIF	0.49	0.48	No DIF	0.29	0.58	No DIF
20	0.14	0.71	No DIF	0.91	0.00	Non-uniform DIF	0.02	0.90	No DIF
21	1.23	0.26	No DIF	0.31	0.58	No DIF	0.02	0.89	No DIF
22	0.39	0.53	No DIF	0.92	0.33	No DIF	0.27	0.60	No DIF
23	1.70	0.19	No DIF	1.20	0.27	No DIF	1.47	0.22	No DIF
24	0.98	0.32	No DIF	0.27	0.60	No DIF	0.04	0.85	No DIF
25	0.10	0.74	No DIF	0.11	0.73	No DIF	2.45	0.11	No DIF

26	4.63	0.13	No DIF	0.07	0.78	No DIF	0.02	0.88	No DIF
27	0.60	0.43	No DIF	0.83	0.36	No DIF	1.09	0.29	No DIF
28	4.51	0.03	Non-uniform DIF	0.86	0.35	No DIF	0.15	0.69	No DIF
29	0.94	0.01	Non-uniform DIF	0.05	0.82	No DIF	0.01	0.90	No DIF
30	4.41	0.03	Non-uniform DIF	0.09	0.76	No DIF	2.03	0.15	No DIF
31	0.02	0.88	No DIF	0.42	0.01	Non-uniform DIF	2.59	0.10	No DIF
32	0.37	0.54	No DIF	0.76	0.38	No DIF	1.08	0.29	No DIF
33	0.67	0.41	No DIF	0.10	0.75	No DIF	0.06	0.80	No DIF
34	0.10	0.75	No DIF	0.19	0.65	No DIF	0.12	0.73	No DIF
35	0.53	0.46	No DIF	0.64	0.42	No DIF	0.24	0.62	No DIF
36	0.53	0.46	No DIF	2.56	0.10	No DIF	1.42	0.23	No DIF
37	3.58	0.05	No DIF	0.01	0.91	No DIF	3.79	0.05	No DIF
38	1.04	0.30	No DIF	9.20	0.00	Non-uniform DIF	1.80	0.17	No DIF
39	0.93	0.01	Non-uniform DIF	0.02	0.89	No DIF	0.02	0.88	No DIF
40	0.20	0.65	No DIF	0.34	0.55	No DIF	0.16	0.69	No DIF
41	0.07	0.79	No DIF	3.44	0.06	No DIF	2.62	0.10	No DIF
42	0.07	0.78	No DIF	0.97	0.32	No DIF	4.30	0.03	Non-uniform DIF
43	1.78	0.18	No DIF	0.49	0.48	No DIF	0.43	0.03	Non-uniform DIF
44	0.07	0.79	No DIF	1.13	0.28	No DIF	1.06	0.30	No DIF
45	0.47	0.49	No DIF	0.02	0.87	No DIF	0.13	0.71	No DIF
46	0.37	0.54	No DIF	2.42	0.11	No DIF	0.98	0.32	No DIF
47	1.40	0.23	No DIF	0.00	0.97	No DIF	0.55	0.45	No DIF
48	0.00	0.99	No DIF	3.66	0.05	No DIF	5.01	0.52	No DIF
49	2.10	0.14	No DIF	0.91	0.33	No DIF	0.40	0.52	No DIF
50	0.44	0.50	No DIF	2.97	0.08	No DIF	0.31	0.57	No DIF
51	4.51	0.03	Non-uniform DIF	0.86	0.35	No DIF	0.15	0.69	No DIF
52	5.94	0.21	No DIF	0.05	0.02	Non-uniform DIF	0.01	0.90	No DIF
53	4.41	0.03	Non-uniform DIF	0.09	0.76	No DIF	2.03	0.15	No DIF
54	0.02	0.88	No DIF	5.42	0.61	No DIF	2.59	0.10	No DIF
55	0.37	0.54	No DIF	0.76	0.38	No DIF	1.08	0.29	No DIF
56	0.67	0.41	No DIF	0.10	0.75	No DIF	0.06	0.80	No DIF
57	0.10	0.75	No DIF	0.19	0.65	No DIF	0.12	0.73	No DIF
58	0.53	0.04	Non-uniform	0.64	0.42	No DIF	0.24	0.62	No DIF

			DIF						
59	0.53	0.46	No DIF	2.56	0.01	Non-uniform DIF	1.42	0.23	No DIF
60	3.58	0.05	No DIF	0.01	0.91	No DIF	3.79	0.05	No DIF

Table 6. Percentage of Items That Displayed Non-uniform DIF in 2020, 2021 and 2022 NECO June/July Mathematics Multiple-Choice Questions Based on Gender.

Year	No. of Items	Non-Uniform DIF	Percentage of items with non-uniform DIF	Items with non-uniform DIF
2020	50	8	16 %	4, 28, 29, 30, 39, 51, 53, 58.
2021	50	9	18 %	8, 9, 10, 12, 20, 31, 38, 52, 59.
2022	50	4	8 %	4, 14, 42, 43

The outcome presented in Table 6 shows that 8 items, that is Items 4, 28, 29, 30, 39, 51, 53 and 58, representing 16% of the total items in NECO 2020 June/July mathematics multiple-choice test questions displayed non-uniform differential item functioning, with Items 4, 29, 39 and 58 favouring male students, while Items 28, 30, 51 and 53 favoured female students.

In 2021, 9 items, that is Items 8, 9, 10, 12, 20, 31, 38, 52 and 59, representing 18% of the items displayed non-uniform differential item functioning, with Items 10, 20, 31 and 52 favouring male students, while Items 8, 9, 12, 38 and 59 favoured female students. In 2022, 4 items, that is, Items 4, 14, 42 and 43, representing 18% of the items displayed non-uniform differential item functioning, with Items 4 and 43 favouring male students, while Items 14 and 42 favoured female students.

Non-uniform DIF occurs when there is an interaction between test takers' ability level and their performance on an item contributing to change in the direction of DIF along the ability scale. In non-uniform DIF, interaction is found between trait level, group assignment and item responses. The results of the analyses for the three consecutive years are in agreement with the previous research work of Ibrahim, Aminu, Yalwaji and Sani (2022) in which they discovered that 26 items (43% of the 60 items) in NECO 2017 Mathematics multiple-choice test were found to reveal non-uniform DIF, in which 19 items exhibited non-uniform DIF. The results of the analyses for the three consecutive years were in agreement with the previous research work of Annan-Brew and Cobbinah (2020) in which the results showed that forty-three (43) items or 86% of the items of 2015 WASSCE core mathematics results in Southern Ghana, revealed DIF. Out of the 43 items, 34 items that showed statistically significant non-uniform DIF had 18 items in favour of male candidates and 16 items in favour of female candidates. The current study also conforms with the works of Ong, Williams and Lamprianou (2015) in which the results showed 19 items (out of the 60 items) had non-uniform DIF. The 19 items that revealed significant non-uniform DIF had 7 items in favour of males and 5 items in favour of females.

CONCLUSION

The study detected uniform and non-uniform gender-related differential item functioning in NECO mathematics multiple-choice standardized test in Imo State. Based on the findings of the study, it was concluded that for the three consecutive years under review, NECO June/July mathematics multiple-choice questions were not free from uniform and non-uniform DIF.

RECOMMENDATIONS

The following recommendations were made based on the findings and conclusion of the study.

a) The findings of this study have provided supporting evidence that the use of DIF statistical procedures are plausible alternatives to the use of expert judgement in determining the psychometric properties of both achievement and psychological tests
b) National Examinations Council (NECO) should carry out differential item functioning analysis for all test items as part of test development process. This is necessary because of large population of students take the examinations. These students are from different demographic backgrounds, like gender, ethnic groups, etc. Analyzing items for differential functioning would help the examination body to identify items that are biased.

REFERENCES

Abedalaziz, N. (2010). A gender-related differential item functioning of mathematics test items. *The International Journal of Educational and Psychological Association 5*, 101-116

Annan-Brew, R. (2020). *Differential item functioning of West African senior school certificate examination core subjects in Southern Ghana*. Unpublished Ph.D. Dissertation, Department of Education and Psychology, University of Cape Coast, Ghana.

Annan-Brew, R. & Cobbinah, A. (2020). Gender-related differential item functioning of 2015 WASSCE core Mathematics results in Southern Ghana using Logistic Regression procedure. *European Scientific Journal, 16*(16), 188-198.

Amaechi, C. E., Eluwa, I. O. & Madu, A. O. (2020). Detection of school type and ownership differential item functioning in economic multiple-choice standardized test in Nigeria. *Nigerian Journal of Educational Research and Evaluation. 19*, 188-198.

Effiom, A. P. (2021). Test fairness and assessment of differential item functioning of mathematics achievement test for senior secondary students in Cross River State, Nigeria using item response theory. *Global Journal of Educational Research, 20*, 55-62.

Ibrahim, A. (2017). Relative effectiveness of generalized Mantel-Haenszel, simultaneous item bias test and logistic discriminate function analysis for detecting differential item functioning in ordinal test items. *Ife Psychologia (An International Journal), 25*(1), 104-132.

Ibrahim, S.; Aminu, I. H.; Yalwaji, S. & Sani, A. K. (2022). Differential item functioning in Senior Secondary School Certificate Examination multiple-choice items: Implication for technology adoption in large scale assessment. *African journal of theory and practice of educational research, 11*(1), 1 – 9.

Karami, H. (2012). An introduction to differential item functioning. *The International Journal of Educational and Psychological Assessment, 11*(2), 59-76.

Nkwocha, P. C. (2019). *Basics of educational measurement and evaluation (Revised ed).* Owerri: Mercy Divine Publishers.

Nworgu, B. G. (2015). *Educational research: Basic issues and methodology*. (3rd Ed.). Nsukka: University Trust Publishers.

Oladele, B. K., Adegoke, B. A. & LongJohn, D. A. (2020). Assessment of differential item functioning in public examinations mathematics constructed- response tests. *Journal of Positive Psychology and Counselling.*

Ogbebor, U. C. & Onuka, A. O. U. (2013). Differential item function of economics question papers of National Examination Council in Delta State, Nigeria. *Nigerian Journal of Educational Research and Evaluation, 12* (1), 45-60.

Ong, Y. M.; William, J. & Lamprianou, I. (2015). Exploring crossing differential item functioning by gender in mathematics assessment. *International Journal of Testing, 15*(4), 337-355.

Oratokhai, D. I. (2021). *Investigating differential item functioning in National Business and Technical Examination English language multiple choice test items.* Unpublished Ph.D. research seminar, Department of Educational Evaluation and Counseling Psychology, University of Benin, Benin City.