

DIFFERENTIAL ITEM FUNCTIONING OF 2020-2021 NIGERIAN BASIC EDUCATION CERTIFICATE EXAMINATION MULTIPLE CHOICE QUESTIONS IN MATHEMATICS IN ABIA STATE

Ihendinihu, Uchechi Ezinwanyi

Department of Science Education, College of Education,
Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria.
E-mail:ihendinihu.uchechi@gmail.com; Tel: +2348063938750

Abstract

The study identified the Mathematics multiple choice test items of the Basic Education Certificate Examination (BECE) administered by the Nigerian National Examination Council (NECO) in 2020-2021 that exhibited Differential Item Functioning (DIF) with respect to gender and socio-economic status, using ex-post facto research design. Participants of the study consists of three hundred and fifty (350) Junior Secondary School year three (JSS 3) students selected from two coeducational schools in Abia state of Nigeria. Data collected from the responses of students to the 120 Mathematics multiple-choice questions were analyzed using Scheuneman modified chi-square test statistic, chi square statistic and independent sample t-test to answer the four research questions and test the four hypotheses that guided the study. Result showed that Mathematics multiple-choice test items administered by NECO for BECE in 2020 and 2021 contain items that functioned differentially for examinees based on gender and socio economic status (SES). Whereas no significant difference exist in the percentage of items that functioned differentially in favour of both gender, the examination contain test items that significantly functioned differentially for examinees based on SES in favour of the high SES. The research recommends continual conduct of pilot study to analyze DIF for possible modification, replacement or equitable distribution of bias items among notable subgroup, especially socio economic status. In addition, the researcher advocates for the use of very simple English while setting mathematics BECE questions to avoid misinterpretation².

Key words: Differential item functioning, gender, socio-economic status, multiple choice questions, mathematics, and Basic Education Certificate Examination.

Introduction

The relevance of assessment in the teaching and learning process makes it imperative that assessment tools should be carefully developed and scrutinized. Assessment plays a very important role in the educational process and development as it is the basis for taking most decisions in the educational sector and society at large (Ihendinihu, 2020). In education and indeed in the school system, test is crucial in ascertaining students' academic success. Test is the most prominent tool for assessing cognitive achievement of students.

A test is an instrument of measurement, the administration and scoring of which leads to the quantification of psychological characteristics. (Ihendinihu, 2022). Test scores are basis for promotion, certification, appraisal, recruitment and placement. To ensure further validity of tests, it is pertinent that test developers present items that are unbiased and fair, that could be used to examine the ability of test-takers from both homogenous and heterogeneous settings, as the value of such measure would be domiciled in its quality (Kelani & Faleye, 2022). Test fairness and equality are crucial issues in testing because the scores obtained from it have important consequences for all examinees, the authors noted. Test items should be carefully analyzed in a bid to ensuring that no individual or group of examinees responding to the instrument is disadvantaged/deprived in anyway. A fair test should not discriminate against sub-groups of candidates or give an advantage to other groups.

In other to ensure that psychometric qualities of test items are maintained, item analysis, which includes differential item functioning (DIF) of test items should be investigated, established and incorporated in test construction. DIF is a statistical characteristic of an item that shows the extent to which the item might be measuring different abilities for members of separate sub-groups (Agbaegbu, Akanwa & Ihechu, 2022). According to shepherd, Camilli and Averill (2011), DIF specify whether individuals of equal ability have the same probability of getting a given item correct. DIF occurs when different groups of examinees with similar over all ability, or similar status on an appropriate criterion, have, on average, systematically different responses to a particular item. Put simply, the presence of DIF in test items means that two people who are at similar levels of the latent construct being measured but who belong to two different cultural, race or gender groups, respond differently to a particular question purporting to measure that construct, resulting in differences in the level of 'Performance' measured.

Testing in education and psychology is an attempt to measure examinee's knowledge, intelligence or other characteristics in a systematic approach (Aituariagbon & Osarumwense, 2022). Testing various performances of examinees could be based on local language or it could be based on unfamiliar format or emphasis on a learning domain (affective, cognitive and psychomotor) that is not stressed in the education system of the examinees school or location. (Yaghoobi in Ihendinihu, 2022). It is unreliable for a test to be fair. It can only be equitable, hence, the reason why examinees respond to an item differently. Test fairness in the context of test items does not prevent items from functioning differentially for different subgroups. More so, DIF can be regarded as a statistical difference between the probability of a specific population group getting the item right and a comparison population group getting the item wrong given that both groups have equal level of expertise with respect to the content being. Hisehfeld, Moore & Brown (2005) noted that difference in students' test scores in some subjects such as Mathematics could be attributed to social and cultural influences.

Analyzing test items for DIF is a core step to determine if subgroups (gender, school location, socioeconomic status, and school type) to which candidates belong substantially bias any item score. The study of DIF has become an integral part of determining the

validity and reliability of standardized tests, which cut across different groups of examinees. One of such examination is Basic Education Certificate Examination organized by NECO. The students that write this examination come from different socio-economic backgrounds, (high and low) different locations (urban and rural), different ethnic groups and they belong to different gender (male and female). This creates the need to investigate whether the items in these tests function differentially among these groups and whether biased items are distributed equitably among the groups with Mathematics as a focal subject.

Okeke in Ihendinihu (2022) described gender as socially or culturally constructed characteristics, qualities, behaviours and roles, which different societies ascribe to female and males. Therefore, with the introduction of differential item functioning in the BECE mathematics multiple choice item validation process, it is expected that male and female students will respond to the test items equally irrespective of their groups. Thus, the call for the detection of DIF in order to maintain valid assessment items for proper decision-making. Socioeconomic status is a way of describing people based on their education, income, type of job, family size and relationships. Socioeconomic status can be low, medium, and high. People with a lower socioeconomic status have less access to financial, educational, and social and health resources than those with a higher socioeconomic status. Students' socioeconomic status is directly linked to, that of their parents or guardians.

Several scholars have investigated the DIF of various subjects in examinations conducted by different bodies. Kasali and Adeyemi in Ihendinihu (2022) estimated item parameter indices of 2016 NECO mathematics and further investigated the presence of items that functioned differentially across cultural environments in Nigeria and gender. The study adopted ex-post facto design with a sample size of 276,338. The results showed, among others, that the items functioned differentially in cultural environments and gender.

Madu (2012) studied Differential Item Functioning (DIF) by gender in mathematics examination conducted by West African Examination Council (WAEC) in 2011 in Nigeria using a descriptive survey research design. Data obtained from a sample of 1,671 students and analyzed using Scheuneman modified chi-square statistics indicated that items significantly function differently by gender in 39 items and 11 items did not exhibit DIF.

Aituariagbon and Osarumwense (2022) analyzed and compared non-parametric methods of detecting Differential Item Functioning in NECO Senior Secondary Certificate Examination 2019 Economics Multiple Choice items using survey research design. Three (3) DIF methods considered are Mantel Haenszel, standardized P-diff and transform item difficulty (delta plot). The study showed that the three (3) methods displayed DIF. However standardized p-difference and transform item difficulty are most suitable methods than Mantel Haenszel statistics in detecting DIF of 2019 Economics multiple-choice items.

Agbaegbu, Akanwa and Ihechu (2022) examined the differential item functioning of National Business Technical Examinations Board Agricultural Science multiple-choice test items in (2015-2017) in South East zone using descriptive and ex-post-facto research design. Data were collected from a sample of 728 SSS3 students and analyzed using Scheuneman modified chi-square statistics and chi-square test statistics. Results showed

that Agricultural Science multiple-choice test items used in NABTEB 2015-2017; contain test items that significantly functioned differentially for examinees based on gender and school location.

Furthermore, Kelani and Faleye (2022) identified the Economics items of the Osun state Unified Promotion Examination that exhibited Differential Item Functioning (DIF) with respect to gender and school location. The study, which adopted ex-post facto research design, collected data from a sample of 2500 senior secondary school two students. Result of analysis done using Mantel-Haenszel statistics indicated that six items exhibited DIF in items of school location.

Odili (2014) studied the effect of language manipulation on differential item functioning in Biology multiple choice test items used by West African Examination Council (WAEC), in the Senior School Certificate Examination (SSCE), in Delta state of Nigeria through a descriptive survey design and post-test only control group experimental design. Data were collected from a sample of 1,022 students and analysed using Scheuneman modified chi-square statistics and t-test statistics. The study revealed that Biology multiple-choice questions used by WAEC in SSCE contain test items that significantly function differently for students from high and low socio-economic status, urban and rural geographical locations and male and female examinees.

Orhiwene and Queensoap (2019) used DIF approach to detect item biased in Chemistry Achievement Test (CAT) using descriptive comparative research design to describe and compare examinees of the four ethnic groups. A statistical and content analysis were done with Scheuneman modified chi-square. Based on the analysis, the study established that the CAT has some items that showed biases and not all items that exhibited differential functioning were flagged biased. The study concludes that there was an ethnic bias in the CAT.

A look at literature review shows that investigation of DIF in various subjects centered on examinations administered to senior secondary school students with less emphasis on examinations administered to JSS. There is need to extend the investigation to examinations administered to junior secondary schools, a gap which this study intends to fill. In addition, literature is depleting with investigation of SES as a group in DIF. This accounts for its inclusion as a variable of study.

Purpose of the study

The purpose of this study is to investigate differential Item Functioning of 2020-2021 Nigerian Basic Education Certificate Examination Multiple Choice Questions in Mathematics Specifically, the study sought to determined:

1. Determine if the Mathematics multiple choice test (MCT) items in 2020 - 2021 NECO BECE function differentially among test takers based on gender.
2. Determine if the Mathematics MCT items in 2020-2021 BECE function differently among the examinees based on socio-economic status.
3. Compare the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics test items based on gender.

4. Compare the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics test item about Socio Economic Status of students.

Research Question

The following research questions guided the study:

1. What percentage of items in the 2020-2021 NECO Basic Education Certificate Examination (BECE) Mathematics multiple choice test items functioned differentially by gender?
2. What percentage of items in the 2020-2021 NECO BECE Mathematics multiple choice test items functioned differentially by socio-economic status?
3. What is the difference between the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics test items based on gender?
4. What is the difference in the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics test item based on SES?

Hypotheses

The following null hypotheses guided the study and were tested at 0.05 level of significance

1. The 2020-2021, NECO BECE Mathematics multiple choice test items do not significantly function differentially with respect to gender.
2. The 2020-2021, NECO BECE Mathematics multiple choice test items do not significantly function differentially with respect to SES of students.
3. There is no significant difference in the mean value of items that functioned differentially in 2020 and those of 2021 NECO BECE Mathematics multiple choice test items with respect to gender.
4. There is no significant difference in the mean value of items that functioned differentially in 2020 and those of 2021 NECO BECE Mathematics multiple choice test items based on SES.

Methods

The researcher adopted the ex-post facto research design, as the variables (gender and socio-economic status) of the study cannot be manipulated. Besides, the study also based on instrumentation design because it involved analyzing the items of an instrument, that is, determining the extent to which each of the items function differentially by gender and socio-economic status. The participants of the study consist of 350 junior secondary school three (JSS 3) students selected from two coeducational schools in Abia state of Nigeria. Purposive sampling technique was used to select the only coeducational unity secondary school and one private secondary school (coeducational) that participate in NECO BECE. All the JSS3 students were expected to participate, but only those who agreed to participate were used. The distribution is 151 males and 199 females with 217 from high SES and 133 from low SES. The study used two instruments for data collection. The first is a multiple choice test items used for junior secondary schools certificate examination for 2020 and 2021 in mathematics. The instrument is a standardized test administered by NECO for all

unity secondary schools and some private and state schools that opted for it. It contains 120 items of options A-D with only one correct answer. There are 60 items for 2020 and 60 items for 2021 giving at 120 items. Since the BECE by NECO is a standardized instrument, and is considered valid and reliable. However, the instrument was trial tested on a sample of 60 students outside the participants for the study for purposes of reestablishing its reliability. Kudar Richardson -20indicated a coefficient of .76. The second instrument used for the study is socio-economic status questionnaire developed by the researcher. It comprised 15 items on a 4-point scale of Strongly Agree =4 points, Agree=3 points, Disagree = 2 points and Strongly Disagree =1 point. The researcher used this instrument to isolate or distribute students according to socio-economic background. Three experts from measurement and evaluation validated the SESQ. The internal consistency of the items of the instrument was determined using Cronbach alpha and a coefficient of .72 obtained. Data obtained for the study were analyzed using the Scheuneman modified chi-square statistic to answer all the research questions and chi-square test statistic was used to test null hypothesis 1 and 2 while hypotheses 3 and 4 were tested using t-test all at .05 level of significance.

Results

Research question 1

What percentage of items in the 2020-2021 NECO BECE Mathematics multiple choice test items functioned differentially by gender?

Table 1: Scheuneman chi-square Gender Differential Item Functioning Indices for 2020-2021 BECE Mathematics multiple choice test items

Items	2020 χ^2	2021 χ^2	Items	2020 χ^2	2021 χ^2
1	5.98	15.87 ^{**} _m	31	7.27	16.07 ^{**} _m
2	15.54 ^{8*} _m	13.86 ^{**} _f	32	10.66 ^{**} _m	11.03 [*] _f
3	0.87	1.97	33	5.74	7.03
4	14.76 ^{**} _m	0.76	34	7.32	10.32 [*] _f
5	2.64	5.23	35	6.53	14.63 ^{**} _m
6	12.32 [*] _f	8.12	36	1.76	10.61 [*] _m
7	9.96 [*] _f	15.65 ^{**} _m	37	16.98 ^{**} _m	12.05 [*] _f
8	0.95	12.98 [*] _f	38	8.61	8.61
9	10.16 [*] _m	13.82 ^{**} _m	39	9.87 [*] _f	12.71 [*] _m
10	11.85 [*] _m	17.86 ^{**} _f	40	17.04 ^{**} _m	4.21
11	4.87	1.54	41	7.23	14.51 ^{**} _m
12	12.20 [*] _f	6.98	42	2.53	3.42
13	7.96	10.85 [*] _f	43	.98	1.49
14	12.35 [*] _f	17.86 ^{**} _f	44	13.98 ^{**} _m	4.21
15	9.98 [*] _f	0.97	45	4.34	3.63
16	2.84	5.32	46	2.34	0.54
17	11.71 [*] _m	13.87 ^{**} _f	47	2.67	1.37

18	6.24	6.77	48	2.56	2.43
19	10.21 ^{*_m}	8.65	49	6.45	1.35
20	10.85 ^{*_m}	14.53 ^{**_m}	50	6.45	5.36
21	13.71 ^{**_f}	9.98 ^{*_m}	51	5.12	4.14
22	3.22	9.05	52	5.12	4.46
23	14.85 ^{**_m}	7.82	53	2.34	15.34 ^{**_m}
24	17.63 ^{**_f}	3.86	54	10.43 ^{*_f}	3.91
25	2.88	1.87	55	7.58	2.74
26	10.54 ^{*_m}	10.54 ^{*_m}	56	4.66	4.61
27	3.14	1.89	57	11.45 ^{*_m}	3.77
28	14.73 ^{**_m}	5.31	58	6.09	2.37
29	15.91 ^{**_f}	11.76 ^{*_f}	59	10.17 ^{*_f}	5.74
30	6.76	3.88	60	3.56	3.54

Table 1 indicates that 48 out of 120 items representing 40% in 2020-2021 NECO BECE Mathematics multiple choice test items significantly function differentially for testes based on gender. In 2020 examination, 26 items representing 43 % significantly functioned differentially for testes based on gender, while 22 items representing 37% significantly functioned differentially for testes based on gender in 2021. Generally, the Scheuneman modified chi-square comparing NECO BECE Mathematics multiple choice test items for female and male, flagged 48 items with significantly differential item functioning ($P < 0.05$). The result showed that, 26 items out of 48 items representing 54 % in NECO BECE 2020-2021, that displayed DIF favored male while 22 items out of 48 items representing 46 % were in favor of female.

Hypothesis 1:

The 2020-2021 NECO BECE Mathematics test items do not significantly function differentially with respect to gender.

Table 2: Chi-square Summary of 2020-2021 NECO BECE Mathematics Differential Item functioning by gender

GENDER	OBSERVED DIFFAVOURED ITEMS	EXPECTED DIFFAVOURE D ITEMS	DF	CHI SQUARE VALUE CALCULATED	SIG/CHI SQUARE VALUE FROM TABLE
MALE	26	24			
FEMALE	22	24	1	.3334	3.84

From table 2, the chi-square calculated value of 0.3334 is less than the tabulated chi-square value of 3.84 at .05 level of significance. Hence, the null hypothesis is not rejected. The implication is that there is no significant difference between male and female students on

the percentage of items which functioned differentially in the 2020-2021 NECO BECE Mathematics multiple choice test items.

Research Question 2

What percentage of items in the 2020-2021 NECO BECE Mathematics multiple choice test items functioned differentially by socio-economic status.

Table 3: Scheuneman chi-square SES Differential Item Functioning Indices for NECO BECE Mathematics multiple choice test items used in 2020 - 2021.

Items	2020 χ^2	2021 χ^2	Items	2020 χ^2	2021 χ^2
1	17.54** _h	8.97	31	7.88	12.29* _h
2	11.98* _h	4.88	32	9.89* _h	17.37** _h
3	15.13* _h	5.65	33	6.02	5.66
4	6.88	12.19* _t	34	5.18	8.32
5	5.65	17.11** _h	35	8.04	14.67** _l
6	5.20	14.67** _l	36	12.45* _h	17.11** _h
7	11.32* _t	8.32	37	3.91	12.19* _l
8	16.43* _h	5.66	38	15.46** _h	5.65
9	6.32	17.37** _h	39	15.39** _t	4.88
10	4.62	12.29* _h	40	6.77	8.97
11	7.38	6.45	41	9.13	6.43
12	14.81** _h	11.65* _l	42	6.41	4.65
13	6.63	4.86	43	14.76** _l	13.66** _h
14	5.97	15.54** _h	44	13.65** _h	11.89* _h
15	15.43** _h	10.19* _l	45	8.12	13.87** _h
16	8.26	7.94	46	4.76	6.22
17	12.22** _t	7.05	47	14.76** _h	9.89* _h
18	4.32	10.87* _h	48	13.94** _h	7.32
19	17.56** _h	16.18** _h	49	8.80	15.54** _h
20	10.87* _h	5.76	50	10.44** _l	11.03* _l
21	10.44* _t	8.37	51	4.62	5.76
22	8.80	6.06	52	6.32	16.18* _h
23	13.94** _h	14.87** _h	53	16.43** _h	10.87* _h
24	14.76** _h	9.97* _h	54	11.32* _l	7.05
25	4.76	12.02* _h	55	5.20	7.94
26	8.12	5.74	56	5.65	10.19* _l
27	13.65** _h	7.12	57	6.88	15.54** _h
28	14.76** _l	16.03** _h	58	15.13** _h	4.86
29	6.41	14.21** _t	59	11.98* _h	11.65* _l
30	9.13	8.37	60	17.54** _h	6.45

Result of analysis presented in table 3 shows that a total of 29 items representing 43 % of Mathematics multiple choice test items used in NECO BECE in 2020 differentially functioned for students with respect to socio-economic status. In 2021 examination, 30

items representing 50% showed evidence of differential item functioning for examinees based on SES. Generally, the Scheuneman modified chi-square comparing NECO BECE for 2020-2021 mathematics multiple-choice test items for SES flagged 59 items with significantly differential item functioning ($P < 0.05$), representing 49% of all the items. The result showed that, 40 items out of 59 representing 67.8 % in NECO 2020-2021 BECE, displayed DIF in favor of students in high SES while 19 items out of 59 items representing 32.2% were in favor of students from low SES status.

Hypothesis 2

The 2020-2021 NECO BECE Mathematics test items do not significantly function differentially with respect to SES of students.

Table 4: Chi-square Summary of 2020-2021 NECO BECE Mathematics multiple choice test items.

SES	OBSERVED DIF FAVORED ITEMS	EXPECTED DIF FAVORED ITEMS	DF	CHI SQUARE VALUE CALCULATED	CHI SQUARE VALUE FROM TABLE
HIGH	40	29,5			
LOW	19	29.5	1	7.48	3.84

Data in table 4 showed that the chi-square value calculated (7.48) is greater than the chi-square value from table (3.84) at .05 significance level. Hence the null hypothesis is rejected. The implication is that there is a significant difference between students from high and low SES on the percentage of items that functioned differentially in 2020-2021 NECO BECE Mathematics multiple choice test examination in favor of high SES group.

Research Question 3

What is the difference between the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics test items based on gender?

Table 6: Mean and standard deviation of DIF of 2020 and 2021 NECO BECE test items based on gender.

Variable	Observation	Mean	SD	Standard Error mean
2020 DIF items	26	12.6862	2.51890	.49400
2021 DIF items	22	13.4864	2.39778	.51121
Mean diff		0.8002		

The mean and standard deviation of gender based DIF for 2020 are 12.6862 and 2.5189 respectively, whereas the mean and standard deviation of gender based DIF for 2021 are

13.486 and 2.39778 respectively. This shows a mean difference of 0.8002 in favour of 2021. A test of hypothesis will ascertain the statistical significance of the difference.

Hypothesis 3: There is no significant difference in the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics multiple choice test items with respect to gender.

Table 6: Independent sample t-test analysis of DIF of 2020 and 2021 NECO BECE Mathematics items based on gender.

variable	Observation (N)	Df	t- value	Sig	remarks
DIF for 2020	26				
DIF for 2021	22	46	1.121	.268	Accept Ho

Result from table 6 show that t-value of 1.121 is not significant at .05 level ($P = .268 > .05$). Hence, H_0 is not rejected. The conclusion is that the mean of DIF for 2020 and 2021 based on gender is not statistically different.

Research Question 4

What is the difference in the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics test item with regards to SES of students?

Table 7: Mean and standard deviation of DIF of 2020 and 2021 BECE Mathematics test items based on SES.

Variable	Observation	Mean	SD	Standard Error mean
2020 DIF items	29	13.9303	2.24738	.41733
2021 DIF items	30	13.5603	2.47960	.45271
Mean diff		0.3700		

Result shown in table 7 indicates that the mean and standard deviation of SES based DIF for 2020 are 13.9303 and 2.24738 respectively while the mean and standard deviation of SES based DIF for 2021 are 13,5603 and 2.47960 respectively. This shows a mean difference of 0.3700 in favour of 2020. A test of hypothesis will confirm the significance of this difference.

Hypothesis 4

There is no significant difference in the mean value of items that functioned differentially in 2020 and 2021 NECO BECE Mathematics multiple choice test items based on SES of testees.

Table 8: Independent Sample t-test of DIF of 2020 and 2021 NECO JSCE Maths items based on SES.

Variable	Observation (N)	Df	t- value	Sig	remarks
DIF for 2020	29				
DIF for 2021	30	57	.600	.551	Accept Ho

From table 8, t-value of .600 is not significant at .05 level ($P = .551 > .05$). Hence, H_0 is not rejected. The implication is that the mean of DIF for 2020 and 2021 with respect to SES is not statistically different.

Discussion of findings

The findings of this study based on analysis done using Scheuneman modified chi-square indicates that there were cases of gender and socio-economic status differential item functioning in the 2020 and 2021 NECO BECE Mathematics test items. The result also showed that the differences in the means of DIF for gender and SES for the two years are not significant. Hence, the DIF issues for the two years are the same.

The fact that more items functioned differently in favour of the students from high socio-economic status than those from low socio-economic level is in line with a priori expectation. Students from this group may have access to additional study materials that will expose them to diverse modes of constructing questions in the content area. Some of them may have home tutors that will provide additional enlightenment, which will be of advantage to them.

The findings of this study align with the findings of previous studies. Kasali and Adeyemi (2022) reported differential item functioning of 2016 NECO Mathematics multiple test items based on gender. Similarly, Kelani and Faleye (2022) found that economics items of the Osun state unified promotion examination exhibited differential item functioning with gender. The result also agrees with Madu (2012) who worked on the analysis of gender-related differential item functioning in Mathematics multiple choice items administered by WAEC and reported that male and female examinees function differentially in 39 items out of 50 items. The finding also corroborate that of Agbaegbu, Akanwa and Ihechu (2022) who found that Agricultural Science multiple-choice test items used in NABTEB 2015-2017 contain test items that significantly functioned differentially for testees on the basis of gender. In addition, Adediwura (2013) identified differential item functioning in relation to gender and student's course of study using IRT and GLM methods. Orluwene and Queensoap (2019) also used Differential Item Functioning via Scheuneman chi-square statistics to detect item bias based on ethnic groups in Chemistry Achievement Test. The results of this study agree with Adedoyin (2010) who identified gender-biased items in Botswana Junior Certificate Examination. The methodology of detecting DIF adopted in this study is in line with that of Agbaegbu, Akanwa and Ihechu, Orluwene and Queen soap (2019), Odili (2014), Madu (2012).

Conclusion

Based on the findings of the study, it is concluded that the BECE Mathematics multiple choice test items administered by NECO in 2020 and 2021 functioned differentially among the examinees on the basis of the subgroups of gender and socio-economic status. In addition, the DIF based on gender and SES in 2020 and 2021 is not significantly different.

Sequel to the findings and conclusion, the researcher recommends that test practitioners should conduct trail testing and perform DIF analysis prior to test administration. This will help test developers to modify, replace or equitably distribute items that function differentially for different test taking groups. In addition, the researcher advocates for the use of very simple English while setting mathematics BECE questions to avoid misinterpretation. A similar study should be conducted that will consider other subgroups like school location, school type, ethnic groups and so on, in Mathematics or other subjects in both junior and senior secondary school levels.

REFERENCES

- Adediwura, A.A (2013). A comparative study of item response theory and generalized linear model method of detecting differential item functioning in dichotomous test. *Research journal in organizational psychology and educational studies* 2 (6):308-316
- Adedoyin, O.O (2010). Using IRT approach to detect gender biased items in public examinations. *Educational research and reviews academic journals* 5 (7): 385-399
- Agbaegbu, C.N, Akanwa, U.N & Ihechu, K.J.P (2022). 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
- 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
- Hisehfeld, M, Moore, R.L & Brown, E (2005). Exploring the gender gap on the GRE subject test in economics. *Journal of economic education*.26:316-322
- Ihechu, K.J & Madu, A.O (2016). Differential Item functioning of National Examination Council Agricultural science questions for 2015. *Journal of education and leadership development* 8 (2):21-16
- 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
- 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
- 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

- Madu, B.C (2012). Analysis of gender-related differential item functioning in Mathematics multiple-choice items administered by West African Examination Council (WAEC). *Journal of education and practice* 3(8):71-78
- Odili, J.N (2014). *Effect of Language manipulation on Differential Item Functioning of Biology multiple-choice test* unpublished PhD Thesis submitted to the department of Science Education, University of Nigeria Nsukka
- Orluwene, G.W & Queensoap, M (2019). Detecting item bias with Scheuneman chi-square in a chemistry achievement test in Nigeria. *International journal of innovative social & science education research* 7 (1):88-101
- Shephard, C, Camilli, G & Averill, M (2011). Comparison of procedures for detecting test item bias with both internal and external ability criteria. *Journal of educational statistics*. 6(4):317-375
- Swaminathan, H & Rogers, H.J (2010). Detecting differential item functioning using logistic regression procedures. *Journal of educational measurement* 27 (4):362-370.