ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

Investigation of Differential Item Functioning of Basic Education Certificate Examination (BECE) Mathematics Items in Akwa Ibom State, Nigeria

[1]Evans, G. U., [1]Uko M. P. and [2]Ekim, R. E. D.
[1]Department of Educational Foundations, Akwa Ibom State College of Education, Afaha Nsit
[2]Department of Educational Foundations, University of Uyo

Abstract

Differential Item Functioning (DIF) of multiple-choice test is a potential bias source in measurement, which can be misleading in any evaluation process. Therefore, it is pertinent to identify the possibilities of DIF in multiple-choice items. The descriptive survey design was adopted to study the differential item functioning of 2020 Basic Education Certificate Examination (BECE) Mathematics multiple-choice items. The study population consisted of 19,560 Junior Secondary three (JSS3) students, who registered and sat for the 2019/2020 BECE in Akwa Ibom State. From the population of 1,956 students were selected from 530 secondary schools using stratified, proportionate and sample random sampling techniques. Three research questions were raised and subjected to item differential functioning analysis. The result showed that fourteen items out of 50 items functioned differentially with respect to gender. The result also revealed that sixteen items out of 50 items functioned differently with respect to school location, while 22 items functioned differently with respect to school ownership. It was recommended that, DIF analysis should be incorporated into educational assessment to obtain valid psychometric properties of test and valid educational assessment. Also, examination bodies and test developers should promote analysis of test items bias for every fixed response test.

Keywords: DIF, IRT, Item Effectiveness, BECE, Difficulty Index

Introduction

Test is a process of determining examinees' ability to complete certain task or demonstrate mastery of skill or knowledge of contents. Therefore, test is often conducted to measure examinees' knowledge, skill, aptitude, physical fitness, classification, social development, emotional maturity and mastery of some concepts or achievements in a specific area (Asim, Ekuri & Eni, 2013). Test can be administered verbally, on paper, on a computer, or in a confined area that requires the examinees to physically perform a set of skills. Test produces qualitative or quantitative information which can be used to adjudge students' achievement. It can further be used in providing incentives and goals for the students. In addition, it can be used to certify attainment as well as provide information for decision-making (Rutayuga & Amiri, 2005; Evans, 2016). McAllister and Guidice (2012) and AERA *et al.* (2014) sees test as an item or set of items presented to an individual or set of individuals to which they are expected to respond under specific a condition, with the intent to determine the extent to which traits are present or absent in the respondents.

In Nigeria, multiple-choice tests are frequently deployed by most external examination bodies as well as subject teachers to examine students' mastery of a subject, including mathematics. The reason is that multiple-choice test items can be structured to cover a wide area of a curriculum; it can measure most educational outcomes such as knowledge, understanding, judgment and problem solving (Asim, Evans & Idaka, 2020). Of course, it can always be used to test a large number of students at a time. In addition, multiple-choice item allows for computer-based testing, hence, marking and collation of results processes can be automated. It also aids broad contents coverage, rapid scoring and provision of quick feedback to students (McAllister & Guidice, 2012; Bassey, Joshua & Asim, 2014). Therefore, multiple-choice item can be considered as a panacea to the problem of large class size, which has always been an issue challenging test credibility and

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

civility in the Nigerian educational system.

Measurement of educational constructs in schools is made possible through the use of test items. The cognitive construct of students, like mathematics cognitive ability, can be measured with mathematics achievement test. Items on a test instrument ought to measure the intended ability irrespective of the parameter of subgroup of the students. A test item provides more advantages for one group over another to the extent that their comparative interpretation using measurement theory becomes biased (Zumbo, 2015). Measurement theory commonly refers to as psychometric theory provides the foundation for evaluating tests, their uses and interpretations. Measurement theory espouses validity and reliability properties of test items to define the quality of a test. Validity is the extent to which evidence and theory occur with interpretations driven by the test result as captured in test usage proposal. It is also the degree to which a test actually measures the variables it purports to measure (Nitko, 1996; Cronbach, 1988; Kpolovie, 2002; Joshua, 2005).

The truthfulness of a measured item is its validity (Shaughnessy & Zechmeister 2003; Adebalaziz, 2010). Together, the validity and reliability of a test provide the basis for rating the technical credibility and appropriateness of the use as well as interpretations of test results. Test validity is most important when evaluating the uses of test and the inferences made from test results. However, other considerations such as fairness and comparability of test results for different test takers or from one occasion to another are also important. In most tests, fairness and comparability of test are readily subsumed under more fundamental subjects of validity and reliability. Hence a test that is meant for different sub population but which give each sub-population different approaches in responding to the test result is not merely biased, but unfair to all the population of students. Such a test lacks validity and objectivity because the test will need different interpretations for each sub-population of the students (Umobong & Tommy, 2019).

Theoretical Foundations of Testing

Test objectivity is measured by two conditions namely the calibration of test instrument and measurement of examinees' ability. The calibration of measuring instrument must definitely be independent of people that are used for the calibration. Secondly, the measurement of examinees' ability must be independent of the instrument used for measuring (Nenty, 2004; Umobong, 2004; Joshua, 2008). The objectivity property is not found in psychological or behavioural measurement. It is lack of the objectivity property in behavioural measurement based on classical measurement model that makes the resulting data basically uninterpretable and only meaningful in extremely limited situations. For a long time, the classical test theory (CTT) has dominated the development of measuring instrument in Nigeria and other part of Africa (Turkcebilgi, 2010). Fortunately the solution is gradually changing with the introduction of the item response theory (IRT).

In CTT, students' scores have been based on the summation of the scores obtained by the examinees on the test items of a given test. Using the summed up or total score of all the items as the basis to determine the performance of students in a given subject tends to hide the characteristic of both the testees and the item that constitute the test (Obinne, 2011). Certainly, the situation does not ensure objective and adequate or accurate decision about the testees and the items. Therefore, it is necessary to use the modern test theory to check those hidden psychometric properties of both the testees and the items as a way of improving test development. In view of above, item response theory (IRT) became the paradigm for the design, analysis and scoring of test, questionnaires. Similar instrument can be used for measuring ability, attitude or other parameters (Fox, 2010). Item response theory as set of mathematical models designed to describe the fundamental relationship between examinees' abilities and performance on an item (Cohen & Walls, 2001). By item response theory (IRT) standards, test items should not depend on the characteristics of the sample. With the use of

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

IRT framework, the analysis of test items, Psychometricians have established that some items in a test may function differently from what the test is meant for. It means that such items have interaction with the characteristics of the examinees taking a test. Therefore, the items are described as having differential function. IRT assumes there is Mathematical function that relates the probability of correct respond on an item to an examinee's ability (Sa'ad, Adamu & Sadiq, 2014).Differential item functioning is a difference in proportion of correct responses between equal-ability respondent in two groups. A test that exhibits differential item functioning (DIF) is one that is unfair to a subgroup population in which it is being used. DIF occurs when two groups, the reference and focal groups, possess relevant knowledge and skill but performed differently in a given test (Umoinyang, 2011). DIF is said to exist when people from different groups but with the same ability systematically responding differently to specific test items. It is a threat to test validity and invalidates interpretation of a test result (Pido, 2012; Omorogiuwa & Iro-Aghedo, 2016). It occurs when examinees of the same ability do not have equal probability of getting an item correctly (Roever, 2005). Linacre, 2007 noted that despite the existence of statistically significant difference, the impact of one item could have too small an influence on the meaning of test result. Consequently, it is important to study statistically significant and the logit difference of at least 0.5 in order to address test item bias. Irrespective of the amount of DIF detected, with a satisfactory fit model, the DIF can be negligible. Trivial impact is the difference in the person estimate from the two analyses of less than 0.5 logits (Wright & Panchapakesan, 1969). In differential item functioning analyses, abilities of various samples are evaluated based on the test items linked with demographical features like male and female in similar performance grade or school ownership and school location of the examinees (Greer, 2004; Perrone,2006; Pedraijita, 2009; Gomez-benito, 2017).

Abeddalaziz (2010) investigated gender-related differential item functioning of Mathematics test items. The researcher determined the DIF of mathematics items and concluded that the percentage of agreement among the three approaches in detecting DIF was relatively low. Omorogiuwa and Iro-Aghedo (2016) investigated DIF by gender in National Business and Technical Examination Board (NBTEB) using 2015 Mathematical multiple-choice test items (Dichotomous). The results of the study indicate that male and female students functioned differently in 17 items (representing 34%) while there was no difference in 33 items (representing 66%). Adedoyin (2010), studied gender biased items in Mathematics examination found that out of 16 test items that fitted the three parameters logistis (3PL) item response theory statistical analysis, 5 items were gender biased. Lyons-Thomas et al. (2014), examined gender differential items function (DIF) across four jurisdictions that took part in a large-scale international assessment in Canada, Shangai, Findland and Turkey. They observed that some items performed differently among the examinees from the different countries. Six items representing 12% had DIF effect, while the other 44 items representing 88% had no DIF effect.

Research reports indicate that test items could function differentially for subgroup defined by location. For instance, Uruema and Adams (2013) conducted a study on differential item functioning method as an item bias indicator as an item bias indicator using logistic regression on NECO Economics examination items. They detected items that have DIF against subgroup of students in urban and rural schools. The study further revealed that from the 60 multiple-choice items in NECO Economics examination, 18 items showed DIF. Their findings implies that items used in assessing students' ability have element of bias that place the rural school examinees at disadvantage and the urban schools' examinees were favoured. However, the study by Lee and Mclntire (2001) reveals that there is no significant difference between performance of rural and urban students.

Research studies have been done on differential item functioning of items due to school type. Amuche and Fan (2017) using logistic regression statistics detected item that have DIF against subgroups such as public and private

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

schools' examinees. The result shows that out of 60 items in NECO Biology examination of 2012, 10 items were identified to display DIF of which, six items favoured private school students, while the public school students were at disadvantage because only four items actually favoured public schools. Adeosun (2020) investigated item biasness of early reading literacy for pre-school children in Oyo State. He uses DIF analysis system software version 5.0 to conduct in order to establish items that function differentially between private (reference group) and public (focal group).The result was presented in line graph of items and it indicated DIF in favour of private and public school children. The result shows that out of 226 items, only 70(42%) items were biased with respect to school type. Ogbebor and Onuka (2013) investigated DIF method as item bias indicator. They used logistic regression statistics to identify test items that have DIF against sub-groups such as public and private schools; urban and rural schools. It was discovered that 11 items favoured public; another set of eleven items favoured private schools.

In the world over, concerted efforts have been made towards improving students' achievement in Mathematics for effective scientific and technological development. In Akwa Ibom State, Nigeria, the State Government usually conduct qualifying examination for all Junior Secondary three (JSS3) students, to assess their ability to write Basic Education Certificate Examination (BECE**)**. In addition, priority is given to Mathematics teachers during employment to ensure their availability in public schools to promote teaching-learning of Mathematics. This is due to the importance of Mathematics in science and technological (Eduwem & Umoinyang, 2014). Going by the importance of Mathematics in senior secondary education, evidence by the special attention paid to mathematics by stakeholders in the educational section and the subsequent grouping of students into classes based on BECE results. In spite of these, the achievement of students in the subject has not improved, evidenced by the rate of students' achievement in mathematics from 2011- 2015 stated as: 45.4% in the year 2011, 48.4% in the year 2012, 45.8% in the year 2013, 37.0% in the year 2014 and 47.2% in the year 2015. But attention has not been given to the objectivity of the test items from whose results are used for decision making (Asim, Evans & Idaka, 2020).

In attempting to solve the problems of students' poor achievement in mathematics, Asim *et al.* (2013); Eduwem and Umoinyang (2014) noted some factors capable of influencing students' poor achievement in mathematics to include; motivation, class attendance, class size, subject matter, study time, teaching methods, external activities, test formats and school location. Tata, (2013), attributed students' poor achievement in mathematics to the decline in students' interest in mathematics, fear and poorly distribution of library materials in schools. Researchers such as Sa'ad, Adamu & Sadiq (2014); Eduwem & Umoinyang, (2014); Thawabieh, (2016) have traced the downward trend in students' achievement in mathematics to the inability of the students to understand the test items. Asim, Evans and Idaka (2020) emphasized on the type of multiple-choice test items dominating a testing material, and the problem of biasness of the test formats.

Other works still attempting to offer solutions to the problem of item biasness include; Ojerinde, Popoola, Ojo and Onyeneho (2012); Ogbebor and Onuka (2013); Bulus (2018). They observed the presence of some irrelevant elements in the test items significantly responsible for differential performance of testees with the same ability but of different demographic setting. This is because differential item functioning may occur without the judgment of unfairness producing weak performance. From the above, little or no efforts have been made to investigate items characteristics of Mathematics test in BECE in order to determine the influence of differential item functioning on the fluctuating performance of students in Mathematics in the BECE, which is the focus of this study. This was achieved through the following objectives:

1. to investigate the number of Mathematics multiple-choice items in 2020 BECE that functioned differently among students in Akwa Ibom State based on gender,
2. to determine the extent that the 2020 BECE test items in Mathematics function differently

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

due to school location,

3. to examine the extent to which the 2020 BECE test items in mathematics function differently based on school proprietorship.

Research questions

The following questions were raised to guide the study:

1. What is the number of Mathematics multiple-choice items in 2020 BECE that functioned differently among students in Akwa Ibom State based on gender?
2. To what extent do the 2020 BECE test items in mathematics function differently due to school location?
3. To what extent do the 2020 BECE test items in mathematics function differently due to school proprietorship?

Research Methodology

The research methodology applied to find solutions to the research questions raised to guide the study are covered in the subsections below.

Research Design

The research design adopted for this study is the descriptive survey design. Therefore, quantitative method using some set of predetermined questions, which include written and oral was deployed to investigate a group of JSS3 students of Batch 2020. The entire JSS3 students in Akwa Ibom State were assumed to be of homogenous characteristics ((Shaughessy et al., 2003; Rover, 2005; Fowler, 2009, Anikweze, 2009; Isangedighi, 2012). Besides, survey research design can be applied to large and small population by selecting and studying a sample of the population from which inference are made to discover relative incidence, contributions and interrelations of sociological and psychological variables. Survey design provides a quantitative description of trends, aptitudes, opinions, feelings, perception or achievement of a population by studying a sample of that population at a particular time.

Research Area

The research was carried out in Akwa Ibom State, Nigeria. The State is among those classified as educationally advantaged States in the country, as many of her citizens are exposed to all levels of education with literacy rate of 78.84% (National Bureau of Statistics, 2017).

Population

The population for the study comprises 19,560 JSS3 students from public and private secondary schools in Akwa Ibom State who wrote the 2020 BECE.

Sample/Sampling Technique

A sample of 1,956 students (10%) of the students' population consisting of 1027 males and 929 females was selected for study using multistage sampling procedure.

Instrument

The 2020 BECE Mathematics examination items were adopted and used to obtain secondary data. The examination consists of two sections; objective and essay sections, however, the study used only the objective section whose items were structured in the multiple-choice formats. The objective section consists of 50 items marked dichotomously as correct (1) and incorrect as (0). Every item has four options indicated as A to D, of which one of the options is correct, while the other three options functioned as distractors. The items were deemed valid and reliable because they had been validated by examination bodies. Mathematics question paper and optical mark recognition (OMR) marked scripts were obtained from Akwa Ibom State Ministry of Education by the researchers for data analysis.

Analysis of Data

Data were analysed using Phase 2 module of BILOG MG. Item with a difference greater than ±0.5 indicates the presence of significant DIF.

Results

Research Question 1: What is the number of mathematics multiple-choice items in 2020 BECE that functioned differently among students in Akwa Ibom State based on gender? Phase 2 module of BILOG MG was used to establish the number of items in Mathematics Multiple-choice paper in 2020 BECE function differently among students in Akwa Ibom State based on gender (Table 1).

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 1: IRT Analysis of DIF with respect to gender on selected 50 items from 2020 BECE Mathematics

S/N	Gender		DIF Difference	Decision	Remark
	Male	Female			
1	-0.478	-0.782	0.304	N0 DIF	
2	0.222	-0.091	0.313	No DIF	
3	0.552	0.324	0.228	No DIF	
4	0.116	0.246	-0.130	No DIF	
5	-2.537	-2.476	-0.041	No DIF	
6	10.075	5.358	4.7I7	DIF	Favoured Male
7	-0.629	-0.351	-0.278	No DIF	
8	15.872	10.983	4.889	DIF	Favoured Male
9	-1.913	-2.118	0.205	No DIF	
10	0.294	-0247	0.541	DIF	Favoured Male
11	11.237	11.490	-0.253	No DIF	
12	-1.214	-1.300	0.086	No DIF	
13	-0.851	-0.717	-0.134	No DIF	
14	17.577	21.179	-3.602	DIF	Favoured Female
15	0.234	-0.180	0.414	No DIF	
16	-1.86	-1.947	0.087	No DIF	
17	-3.169	-3.498	0.329	No DIF	
18	-0.134	-0.604	-0.13	No DIF	
19	-1.085	-1.922	0.837	DIF	Favoured Male
20	0.331	0.646	-0.315	No DIF	
21	-12.441	7.798	-20.239	DIF	Favoured Female
22	0.581	0.452	0.129	No DIF	
23	-2.495	-1.078	-1.417	No DIF	
24	12.056	-0.427	12.483	DIF	Favoured Male
25	-1.908	-1.805	-0.103	No DIF	
26	0.042	0.325	-0.283	No DIF	
27	0.20	-0.205	0.405	No DIF	
28	7.79	9.126	-1.216	DIF	Favoured Female
29	-0.399	-0.394	-0.005	No DIF	
30	-0.058	-0.017	-0.041	No DIF	
31	19.11	18.048	1.062	DIF	Favoured Male
32	7.39	2.951	4.392	DIF	Favoured Male
33	-0.419	-0.642	0.228	No DIF	
34	5.329	1.417	3.912	DIF	Favoured Male
35	3.920	4.300	-0.380	No DIF	
36	-1.240	-1.092	-0,148	No DIF	
37	-0.562	-0.368	-0.194	No DIF	
38	-0.601	-0.537	-0.064	No DIF	
39	0.166	2.246	-0.080	No DIF	
40	-1.360	-1.464	0.140	No DIF	
41	0.162	0.299	-0.067	No DIF	
42	-1.044	-0.968	-0.076	No DIF	
43	0.744	-0.582	-0.162	No DIF	
44	0.423	1.122	-0.699	DIF	Favoured Female
45	-0.065	-0.074	0.009	No DIF	

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

46	-0.59	-0.748	0.158	No DIF	
47	0.937	-0.7	-0.237	No DIF	
48	-10.188	-4.594	-5.594	DIF	Favoured Female
49	-1.134	-1.018	-0.116	No DIF	
50	0.812	0.252	0.560	DIF	Favoured Male

Table 1 is the IRT DIF statistics on examined items, performance with respect to respondents' gender. The result shows that 14 (representing 28%) items out of 50 items were gender bias and have group difficulty difference of +0.5. From these 14 items, 9 items (representing 18%) favoured male while 5 (10%) favoured the female. The decision column shows items with DIF and the ones with no DIF with regards to the respondents' gender.

Research Question 2: To what extent do the 2020 BECE test items in mathematics function differently due to school location?

Table 2: IRT Analysis of DIF on 50 multiple choice item of 2019/2020 BECE Mathematics with respect to school location

S/N	School Location		DIF Difference	Decision	Remark
	Urban	Rural			
1	0.436	0.303	0.133	No DIF	
2	-0.43	-0.564	0.134	No DIF	
3	-3.254	-3.806	0.552	DIF	Favoured Urban
4	0.074	0.377	-0.303	No DIF	
5	-0.292	-0.233	-0.059	No DIF	
6	-0.348	-0.246	-0.102	No DIF	
7	0.812	0.252	0.56	DIF	Favoured Urban
8	0.154	-.003	0.157	No DIF	
9	-0.417	-0 .448	0.031	No DIF	
10	0.554	-0.966	0.412	No DIF	
11	0.443	-0.791	1.234	DIF	Favoured Urban
12	-0.566	-0.476	-0.090	No DIF	
13	-19.938	-2.590	-17.346	DIF	Favoured Rural
14	-0.28	-0.337	0.057	No DIF	
15	-1.037	1.351	0.314	No DIF	
16	11.027	15.905	-4.876	DIF	Favoured Rural
17	-0.656	-1.019	0.363	DIF	
18	0.084	0.869	0.953	DIF	Favoured Urban

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

19	-0.554	-0.966	0.412	No DIF	
20	-0.417	-0.448	0.031	No DIF	
21	-0.422	- 0.76	0.338	No DIF	
22	-0.019	-0.333	0.314	No DIF	
23	-0.986	-1.256	0.270	No DIF	
24	-1.66	-2.882	1.215	DIF	Favoured Urban
25	-0.504	-0.820	0.316	No DIF	
26	-5.464	4.700	-10.164	DIF	Favoured Rural
27	-4.336	-0.865	-3.471	DIF	Favoured Rural
28	-0.119	-0.484	0.365	No DIF	
29	-5.665	-1.044	-4.621	DIF	Favoured Rural
30	-0.1.199	-1.514	0.315	No DIF	
31	-0.824	-1.165	0.341	No DIF	
32	-0.504	-0.820	0.316	No DIF	
33	-1.108	-1.436	0.328	No DIF	
34	-0.211	-0.672	0.451	No DIF	
35	-0.623	0.673	0.050	No DIF	
36	-0.422	-0.760	0.338	No DIF	
37	-1.534	-2.802	1.268	DIF	Favoured Urban
38	-0.534	-0.704	0.170	No DIF	
39	-1.199	-1.514	0.315	No DIF	
40	-1.667	-2.882	1.215	DIF	Favoured Urban
41	-0.027	-0.431	0.404	No DIF	
42	-0.702	-0.950	0.248	DIF	
43	-0.119	-0.484	0.365	No DIF	
44	-1.461	-1.787	0.321	No DIF	
45	-0.478	-0.782	0.304	No DIF	
46	7.445	10.587	-3.142	DIF	Favoured Rural
47	0.056	-1.204	1.260	DIF	Favoured Urban
48	6.312	-1.653	7.965	DIF	Favoured Urban
49	0.785	-1.142	0.357	No DIF	
50	-0.657	-0.299	0.648	DIF	Favoured Urban

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 2 is the IRT analysis of DIF statistics on 2020 BECE multiple-choice items in mathematics performance with respect to respondents' location. The result showed that sixteen (16) item out of 50 have group difficulty difference of +0.5. Out of the 50 items, 16 items exhibited DIF, which 10 items (20% of the items) favoured urban schools, while 6(12%) favoured rural schools.

Research Question 3: To what extent do the 2020 BECE test items in mathematics function differently due to school proprietorship?

Table 3: IRT analysis of DIF with respect to school type on 50 multiple choice item of 2019/2020 BECE Mathematics

S/N	School Proprietorship		DIF Difference	Decision	Remark
	Private	Public			
1	0.093	-1.434	1.527	DIF	Favoured Private
2	-19.938	-2.592	-17.346	DIF	Favoured Public
3	-0.422	-0.76	0.338	No DIF	
4	-0.566	-0.476	-0.09	No DIF	
5	0.443	-0.791	1.234	DIF	Favoured Private
6	-0.555	-0.926	0.371	No DIF	
7	-14.096	-3.397	-10.699	DIF	Favoured Public
8	-0.419	-0.647	0.228	No DIF	
9	0.20	-0.205	0.406	No DIF	
10	0.552	0.324	0.228	No DIF	
11	-3.169	-3.498	0.329	No DIF	
12	6.3.12	1.653	7.965	DIF	Favoured Private
13	-7.838	-0.902	-6.936	DIF	Favoured Public
14	-0.196	0-.094	-0.102	No DIF	
15	-0.042	0.325	-0.284	No DIF	
16	-0.986	-1.256	0.27	No DIF	
17	-10.188	-4.594	-5.594	DIF	Favoured Public
18	-0.554	-0.966	0.412	No DIF	
19	0.2	-0.205	0.405	No DIF	
20	-0.566	-0.476	-0.09	No DIF	
21	-2.495	-1.078	-1.417	DIF	Favoured Public
22	11.234	11.49	-0.253	No DIF	

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

23	0.285	-0.877	1.162	DIF	Favoured Private
24	-2.765	-1.47	-1.295	DIF	Favoured Public
25	-0.59	-0.748	0.158	No DIF	
26	-1.24	-1.092	-0.148	No DIF	
27	0.116	0.246	-0.13	No DIF	
28	-12.441	7.798	-20.239	DIF	Favoured Public
29	-1.652	-1.959	0.307	No DIF	
30	17.577	21.179	-3.602	DIF	Favoured Public
31	-4.336	-0.865	-3.471	DIF	Favoured Public
32	-0.28.	-0.0.337	0.057	No DIF	
33	-0.601	-0.537	0.064	No DIF	
34	0.423	1.122	-0.699	DIF	Favoured Public
35	-0.504	-0.082	0.316	No DIF	
36	-0.027	-0.431	0.404	No DIF	
37	-0.785	-1.142	0.357	No DIF	
38	-1.204	0.056	1.26	DIF	Favoured Public
39	-0.019	-0.333	0.314	No DIF	
40	-0.988	-2.188	1.2	DIF	Favoured Private
41	-1.037	-1.351	0.314	No DIF	
42	-0.445	-1.681	1.26	DIF	Favoured Private
43	-7.712	-2.371	-5.341	DIF	Favoured Public
44	--0.555	-0.926	0.371	No DIF	
45	0.162	0.229	-0.067	No DIF	
46	0.294	-0.247	0.541	DIF	Favoured Private
47	7.445	10.587	-3.142	DIF	Favoured Public
48	-0.702	-0.95	0.243	No DIF	
49	0.166	0.246	-0.08	No DIF	
50	7.91	9.126	-1.216	DIF	Favoured Public

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 3 is the IRT analysis of DIF statistics on examined item performance with respect to respondents' school Proprietorship. The result reveals that out of the fifty (50) items, twenty two (22) indicate group difficulty difference of +0.5. From the twenty (22) items, seven (14%) favoured private school while 15(30%) favoured public schools.

Discussions

Differential item function being used to identify the degree of effectiveness of the 2020 BECE items in measuring different examinees' abilities for the numbers of the sub-groups for purpose of eliminating bias items.

The result indicates that, the items of the multiple-choice section function differently for gender, with fourteen (14) items out of the fifty (50) items been gender biased. The study further revealed that, from the 14 biased items, 9 of the items which is 18% favoured male, while five 5 (10%) favoured the female. The items measured construct having something to do with gender of the students other than their ability in mathematics. The findings are in agreement with Bulus (2018) who had earlier made such observation. The study also agrees with Omorogiuwu and Iro-Aghedo (2016) that items of the National Business and Technical Examination Board of 2015 function differently based on gender.

The study also revealed significance differential item functioning in the 2020 BECE multiple-choice Mathematics due to school location. 16 of the 50 items show no significance DIF, while 34 items function differently among students of urban and rural schools having the same ability in Mathematics. 10 of the items favoured rural schools, while 6 items favoured urban schools. This implies that the performance of the examinees on the items do not only depend on the ability in Mathematics, but also on the school location. The finding agrees with that of Uruema and Adams (2013), which show that, there exist differential item functioning between Urban and Rural students. The findings also agreed with Lyons-Thomas et al (2014), they observed that some items performed differently among the testees from the different countries.

Lastly, the result revealed that, out of the 50 items in 2020 multiple-choice Mathematics, 22 were biased with respect to school type. 7 items favoured students of private schools, whereas, 15 items favoured students of public schools. It means that, the performance of the examinees on the items do not only depend on their ability in Mathematics, but also on the school type. This findings is congruence with Ogbebor and Onuka (2013). They used logistic regression statics to identify items that have DIF against sub-group such as public and private schools and urban and rural areas and discovered that 11 items favoured public schools, while 11 items also favoured private schools.

Conclusion

The study reveals DIF in the 2020 BECE multiple-choice examination in Mathematics. The 2020 BECE multiple-choice Mathematics items show differential item functioning with reference to gender, school location and school type. Based on the study, the performance of students in the multiple-choice Mathematics items of 2020 BECE did not only depend on the ability, but also on their gender, school location and school type.

Recommendations

Based on the results obtained from the study, the following recommendations were made.

1. DIF analysis should be incorporated into educational assessment so as to obtain valid psychometric properties of test and valid educational assessment.
2. Examination bodies and test developers should step up to include item bias during item analysis.
3. Measurement or test experts should acquaint themselves the opportunity of obtaining knowledge and skills involved in DIF.

References

Adebalaziz, N. (2010). A gender related differential item functioning of Mathematics test items. *International Journal of Educational Research and Psychological Assessment,* 5;101- 106.

Adedoyin, O. O. (2010). Using item response theory approach to detect gender biased items in public Examinations. *Educational Research and Reviews*

Academic Journal, 5(7): 385-399.

Adeosun, O. O. (2020). An investigation of item biasness of early reading literacy test developed for pre-school children in Oyo State. *The African Journal of Behavioral and Scale Development Research (AJB-SDR), 2*(1), 17-26.

American Educational Research Association (AERA), American Psychological Association & National Council on Measurement in Education (2014). Standards for educational and Psychological testing. Washington, DC: American Educational Research Association.

Amuche, C. I. & Fan, A.F. (2017). An assessment of item bias using differential item functioning technique in NECO Biology conducted examinations in Taraba state Nigeria. *America International Journal of Research in Humanities, Art and Social Science*. Retrieved from http// www.lasir.ne.

Anikweze, C. M. (2009).*Simplified Approach to Educational Research*. Enugu, Snaap Press Limited.

Asim, A.E, Ekuri, E.E. & Eni, E.I (2013).A diagnostic study pre-service teachers' competency in multiple choice item development. Research in Education, 89, 13-22.

Asim, A. E., Evans, G. U. & Idaka, I. E. (2020). Analysis of Multiple choice Item Format and Secondary School Students" Achievement in mathematics in Akwa Ibom State, Nigeria. *African Journal of Theory and Practice of Educational Assessment*, 2(8): 72-87

Bassey, S. W., Joshua, M.T. & Asim, A. E. (2014). Gender differences and mathematics achievement of rural senior secondary students in cross river state, Nigeria. Proceedings of International Conference to Review Research in Science, Technology and Mathematics Education (epi STEME-3).

Bulus, C. H. (2018) Development and calibrating of a diagnostic test in quantitative aspect of Economics in senior Secondary schools in Plateau State, Nigeria.Unpublished Ph.D thesis, University of Jos.

Cohen, B. & Walls, J. (2001). Using item response theory to assess effects of Mathematics in structure in special population. *Journal of Exceptional Children, 68*(1):25-46.

Cronbach, L. J. (1998*). Five Perspective on Validity Argument*. In Wainer, I. T. and Braun, H.I (Eds) *Lawrence Erlbaum*. New Jersey: Hillsdale.

Eduwem, J. D. & Umoinyang, I, E. (2014). Item Types and Upper Basic Education Students' performance in Mathematics in the southern Senatorial District of Cross River State, Nigeria. *Journal of Modern Education Review, 4* (1), 57-73.

Evans, G. U. (2016). Students' perception of multiple choice item format and mathematics achievement test in junior secondary schools in Uyo Educational Zone of Akwa Ibom State, Nigeria.*Academic Journal of Educational Research 4*(7), 111-116.

Fowler, F. J. (2009).*Survey Research Methods*. London, SAGE, 201pp.

Fox, J. P. (2010). Bayesian item response modeling. New York, Springer.

Gomez-Benito, J. (2017). Detecting differential item functioning in behavioural indicators across parallel forms.*Psicothema, 29*(1), 91-95

Greer, T. G. (2004). Detection of differential item functioning (DIF): A comparison of four methods: Mentael-Haenszel, logistic regression, simultaneous item bias and likelihood ratio test. Unpublished Doctoral Dissertation, University of Houston.

Isangedighi, A. J. (2012).*Essentials of Research and Statistics in Education and Social Science.*Calabar, Eti-Nwa Associate, pp66-67.

Joshua, M. T. (2005). *Application of item Response Theory to Practical Testing*

Problems. New Jersey, Lawrence Eribaum Associates.

Joshua, M.T. (2008). Fundamental of test and measurement in education. Calabar,. University of Calabar Press.

Kpolovie, P. J. (2002). Test Measurement and Evaluation in Education. Port Harcourt: Emhai Publishers.

Lee, J. & Mclritire, W. (2001).Interstate variation in the Mathematics achievement of rural and non-Students.*Journal of Research in Rural Education, 16(3),* 15-21.

Linarce, J. M. (2007). A user's guide to winesteps-ministeps Rasch- model computer programs, Chicago, II. Winesteps' com.

Lyons-Thomas, J., Sandilandas, D. & Eriikan, K. (2014). Gender differential item functioning in Mathematics in four International jurisdiction. *Education and Science, 39(172), 20-32.*

McAllister, D. & Guidice, R. M. (2012). This is only test: A machine-graded improvement to the multiple-choice and true-false examination. Teaching in Higher Education, 17(2), 193-207

National Bureau of Statistics (2017).Data on Index of Nigeria Literacy. Abuja, National Bureau of Statistics.

Nenty, H. J. (2004). From classical test theory (CTT) to item response theory (IRT): An introduction to a desirable transition. In O. A. Afemikehe and J. C. Adewale (Eds), Issues in educational measurement and evaluation in Nigeria (in honour of Wole Falayoajo) (pp371-584).Ibadan, Institute of Education, University of Nigeria.

Nitko, A. J (1996).Educational Assessment of student (2nd Ed). New Jersey, USA. Prentice hall.

Obinne, A. D. E. A. (2011). Psychometric analysis of two major exams in Nigeria.*International Journal of Education Service, 3*(2), 137-144.

Ogbebor, U. & Onuka, A. (2013).Differential item functioning method as an item bias indicator. Educational Research, 4(4), 367-373.

Ojerinde, D.,Popoola, K., Ojo, F. & Onyeneho, P.(2012) Introduction to item response theory: Parameter model estimating and application, Abuja: Marvelous Mike press.

Omoro, K. O. & Iro-Aghhedo, E. P. (2016). Determination of differential item functioning by gender in the National Business and Technical Examination Board (NADTEB) 2015 Mathematics Multiple choice Examination. International Journal of Education Learning and Development, 4 (10):25-35.

Pedraijita, J. Q. (2009). Using logistic regression to detect biased test items performance in English Language and Mathematics at West African Senior School. *The International of Education and psychological Assessment Journal, 33*, 2-8.

Perrone, M. (2006). Differential item functioning and item bias: Critical consideration in test fairness. Columbia University working paper in TESOL &Applied linguistics, 6(2), 1-3.

Pido, S. (2012). Comparison of item analysis result obtained using item response theory and classical test theory approaches. *Journal of Educational Assessment in Africa, 7*,192-207.

Rover, C. (2005). That is not fair fairness, bias and differential item functioning in language testing. Retrived from http//www2 hawaii.edu/rover/brownbag.

Rutayagu, A. B. & Amiri, K. (2005). Influences of competence / outcomes based education on assessing technical education and training in Tanzania: NACTE'S perspective, paper presented at 31st Annual Conference of International Association of Education Assessment, Abuja, Nigeria.

Sa'ad, T. U., Adamu, A. & Sadiq, A. M. (2014).The causes of poor performance in mathematics among public senior secondary school students in Azare Metropolis of Bauchi State, Nigeria. Journal of Research &Method in Education, 4(6), 32-40.

Shaughessy, J. J., Zechmeister, E.B. & Zechmeister, J. S. (2003).*Research*

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Methods in Psychology (6th Ed) New York, McGraw Hill Publishers.

Thawabieh, A. M. (2016). A comparison between two test item formats: multiple-choice items and completion items. *British Journal of Education, 4*(8), 32-43.

Tata (2013).Mathematical Modeling dalam Pendidikan Matematika. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika1pp 408-414.

Turkcebilgi, C. (2010). Item Respond Theory. Retrieve respond-theory.

Umobong, M. E. (2004). Item response theory introducing objectivity into educational measurement. In O. A. Afemike and Adewale, J. C.(Eds.), measurement and evaluation in Nigeria (in honour of Wole Faiayalo). Ibadan University of Ibadan Press, 25-32.

Umobong, M. E. & Tommy, U. E. (2019).Comparative analysis of item response theory methods of detecting differential item functioning in educational assessment. *African Journal of Theory and Practice of Educational Assessment,8*, 1-18.

Umoinyang, I. E. (2011).The challenge of removing consistent errors in achievement test using differential item functioning (DIF) detecting methods. Paper presented at the 30th Conference of the Association of Educational Assessment Africa, Nairobi, Kenya.

Uruemu, O. & Adams, O. (2013).Differential item functioning method as an item bias indicator.

Educational Research, 4(4), 367-373.

Wright, B. D. & Panchapakesan, N. N. (1969).A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23-48.

Zumbo, B. D. (2015). A methodology for Zumbo's third generation DIF Analysis and the Ecology of Responding. *Language Assessment Quarterly, 12*(1), 136-151.