DETECTION OF GENDER- RELATED DIFFERENTIAL ITEM FUNCTIONING OF WEST AFRICAN SENIOR SCHOOL CERTIFICATE EXAMINATION MULTIPLE CHOICE MATHEMATICS ITEMS IN EKITI STATE

Prof. M.S. Omirin and Audu Godwin
Faculty of Education, Ekiti State University, Ado-Ekiti
godwinaudu610@gmail.com 08033963827, 08067409235

Abstract

The study investigated detection of gender-related differential item functioning of West African senior school certificate examination multiple choice mathematics items in Ekiti State, Nigeria and its implications for educational assessment. It begins by identifying Test as a major tool for educational assessment and then looked into concept of differential item functioning. The research adopted the descriptive research of the survey type. The population for the study consisted of all senior secondary school students in the 203 public and 235 private secondary school students in Ekiti State. The sample for the study consisted of 1200 senior secondary school III students drawn from 12 public and 12 private secondary schools in Ekiti state. The sample was selected using multistage sampling procedure. The instrument used for data collection was multiple-choice Mathematics items administered by WAEC 2019, which consisted of 50 items. The data collected were analyzed using a two-parameter logistic model (2PL) implemented in BILOG-Mg version 3.3.0 software to answer the general questions. Hypotheses were tested using Bilog M.G. software statistical analysis which was used to generate IRT Item Characteristic Curve (ICC) for each WAEC examination items to show whether items function differentially. An independent sample t-test was implemented using R programming language software version 4.1.1, and Analysis of Variance (ANOVA) were also used to test the postulated hypotheses. All the hypotheses were tested at 0.05 level of significance. One of the findings indicated that the multiple-choice Mathematics items administered by WAEC 2019 gives males examinees greater opportunities in respond to mathematics items than their female counterparts. It was recommended among others that the mathematics teachers to pay more attention and encourage the female students in the class.

Keywords: Differential item functioning, test, multiple-choice item, item performance.

Introduction

Mathematics is a subject that is seen by the society as the foundation of scientific and technological knowledge that is vital in socio- economic development of the nation. Due to this fact, Mathematics is made a compulsory subject in schools at all levels in Ekiti State and Nigeria as a whole. This is why Mathematics is so important that every child must study it for six years in primary school, three years in junior secondary school and three years in senior secondary school. Students are expected to pass this subject before being

promoted to the next class or gain admission into a higher institution of learning in Nigeria. Mathematics is one of the senior secondary school subjects that require assessment to ascertain students' basic knowledge and skills and understanding of the concepts and the nature of mathematical problems in any society.

These objectives can be achieved by the use of different assessment instruments such as; essay tests and objective tests which are utilized by the teacher depending on the aims of the measurement. The focus of this study is on objective tests. Objective test is one of the assessment instrument used in testing or assessing students' academic achievement in any given instruction. The objective test is the most commonly used test format in all school levels, also in entrance examinations to secondary and tertiary institutions. The tests are to make free choice of one correct or best answer from the alternatives given to a question (Omirin & Ajayi, 2013). It is an instrument designed and used to elicit sample of behavior, a set of well constructed stimuli presented to a testee to find out whether learning has already taken place (Adebule & Oluwatayo, 2011)

In Nigeria, there exist a number of national examination bodies and they include National Examination Council (NECO), West African Examination Council (WAEC), National Business and Technical Examination Board (NABTEB), and Joint Admission Matriculation Board (JAMB). These bodies cater for candidates of various backgrounds all over the country. Candidates who participate in the examinations conducted by these examination bodies are in different settings and therefore differently toned for personal and environmental reasons. As a result of this, the problem of test item bias cannot be ruled out in these examinations.

The environment (urban/rural) which a child finds himself/herself goes a long way to determine one's academic achievement in life. Children who come from rich environment have better academic achievement than those from poor environment. Urban areas are well equipped with learning facilities, qualified teachers, good roads and good communication networks which puts them at an advantageous position compare to their rural counterparts where such opportunities are inadequate or somehow lacking. According to Akubuiro cited by Anagbogu (2009), urban learning environment has a greater access to socio-cultural and economics facilities and services and as produce a high performing leaner. While the rural students who have not been exposed to these favourable experiences and rich, environments find it difficult to bridge the gap and so results to poor performance in their various subjects.

To find out if differential item function exists in Ekiti state unified mathematics Examination (ESUME) and confirm if the test items function in different ways for groups of test- takers. Nworgu (2011) in his contribution opined that current research data have implicated test used in national examination as functioning differently with

respect to different subgroups. That is students' scores in such examination are determined largely by the testees' abilities. Ogbebor and Onuka (2013) supported this in their findings that 10 items were bias out of 60 items of Economics multiple-choice items in NECO Examination with respect to school type (public or private) and 8 items were biased in respect to school location (urban and rural). Scores generated from a test that contains items that are biased against one group or the other or test result from unfair testing procedures cannot be used to make a valid quality decisions in education.

According to Berret (2001) multiple choice tests are generally biased towards male while the female students experience more difficulties with questions involving numerical, spatial or high reasoning skills. Also, Adebule (2013) observed that questions always arise concerning whether high average test scores by certain groups are due to actual achievement differences, bias in test or combination of both. Conversely, the favoured groups are the advantaged group during promotion and admission or selection into science-based courses in the high intuitions while the disadvantaged groups on the other hand are disallowed due to some factors tagged extraneous and irrelevant variables that interfere with measurement of the underlying psychological construct being measured. These factors relate to the group like gender, location, school type have significant influence on the testees' response to item.

DIF is necessary, but not sufficient condition for item bias. Thus, if DIF is not apparent for an item, then no item bias is present. However, if DIF is present the its presence is not a sufficient condition to declare the item bias, rather one would have to apply a follow-up item bias analysis (e.g content analysis, empirical evaluation) to determine the presence of item bias. For instance, if in a mathematics test, girls display higher probability of answering any item correctly more than boys of equal ability level due to the fact that the content in the test are biased against boys, then we say the item exhibit DIF and should be considered for modification or removal from the test item.

An item is biased if it discriminates between members of different groups who have the same ability on what is being measured. Put different groups who have the same trait level differ in their score on the item. In his contribution, Adebule (2009) defines bias in test as a situation when item is in an achievement; tests are found to favour one group over another for reasons not explainable by differences in achievement level between groups.

The West Africa Examination Council was established in 1952 with the sole responsibility of conducting examinations required in the public interest in West Africa, which are conducting examinations and awarding certificates that are equivalent to those of examining authorities in Unite Kingdom. The West African countries include five Anglophonic countries which are Ghana, Nigeria, Sirerra Lone, Liberia and the Gambia as well as a

part of Cameroon. The council conducts four different categories of examinations; they are international Examinations, National Examinations, Examination conducted in collaboration with other examining bodies, and Examinations conducted on behalf of other examining bodies. The international exams are exams taken in the five countries with WAEC ordinance. It consists of West African Senior School Certificate Examination (WASSCE). From 1952 to 1968, WAEC performed its duties well without much criticism.

Criticisms started becoming louder in 1969 as a result of massive failure and other variables. (Anigbo 2018). However, from 1970s, some issues appeared to be getting too much for WAEC to handle such as timely release of result, massive failure, uncontrollable population explosion of candidates, overloading of works, cases of leakage of examination papers and increased rate of examination malpractice (Kolawole, 2007)

Examination bodies often carry out empirical verification for detecting biased items in their respective examinations in order to redeem and exclude items found to be biased so that all the examinees can be assured of equity in the examination and also to ensure that the ability of examinees are reliably assessed Examination bodies are expected to construct test items in such a manner that test items are free from writing errors such as wordiness, irrelevancy, offensiveness, and excessive stimulations, so that when an inadequacy exists between groups' examination item scores, the disparity will be attributed to true differences in whatever the test purports to measure in the examinees (Aborisade,2016). As educators take cognizance of the possibility of test item bias in national testing situation, candidates from educationally disadvantaged areas and low socio-economic status would be certain to be fairly treated.

National examination bodies often over-predict or under-predict some candidates from certain states during the selection exercise, to the extent that some examining bodies have different policies of awarding the final grade to examinees. For instance, JAMB has accepted different cut-off points for selecting candidates into Nigerian tertiary institutions based on merit, catchment area, educationally disadvantaged states and institutional discretion. Nigerian as a nation is a heterogeneity setting and there is an assumption that human development is a process dependent upon interaction between inherited qualities and environmental forces.

Researcher on this note have been interested in knowing whether WASSCE multiple choice mathematics items administered by WAEC function differentially between groups based on gender, school location and school type, many researchers has therefore on different occasion, time and location compared the difficulty and the discriminating power of WASSCE using the same or different subject. Their findings however has brought confusion

as to what exactly the position of the situation should be as researchers continue to come out with different findings and drawing different conclusions.

Differential item functioning occur when a characteristic of the item that is not relevant to the test purpose differently influences responses of examinee (Ercikan & Lyong-Thomas, 2013). There is an expectation that if an item on a test is not biased, then examinees from two groups who have equal overall ability ought to have the same probability of correctly responding to it. When examinees from different groups that has comparable ability levels have different probabilities of getting an item correct, differential item functioning (DIF) is said to occur (Battuaz, 2015).

A test that exhibits item bias is one that is unfair to a subgroup of the general population in which it is being used. Item bias occurs when two groups (reference group and focal group) that are matched in terms of their relevant knowledge and skills perform differently in an item (Umoinyang, 2011). DIF is a threat to test validity and invalidates interpretation of the test results for some groups of the same population (Pido, 2012). Item bias occurs when examinees of the same ability do not have the equal probability of getting an item correct (Ojerinde, 2016). This arises mainly due to the sex, cultural, ethnic, religious, or class background of the examinees. Item bias manifest itself in context, language and item structure and format bias (Congbogo & Opara, 2019).

Content bias refers to a situation where knowledge and or skills tested are not part of the educational background of the examinees. Lack of familiarity with content in test items disadvantages individuals in their performance. The individual's responses to items are not based on other irrelevant abilities. Language bias occurs where words in items have different or unfamiliar meanings for different examinee subgroups. The item has difficult vocabulary, group specific language, and vocabulary and reference pronouns. Item structure and format bias occurs where there is ambiguity in the instructions, items stem or options. The content or clues and explanations given to successfully complete the task provided disadvantage individuals in some subgroups (Karami & Nodoushan, 2016).

Since gender equality in education is important in terms of social justice and human rights, it has always been one of the most popular areas in international reports and studies. When examining gender differences in Mathematics, boys shows higher achievement than girl in 28 out of 31 participating countries in Programme for International Assessment Projects (PISA) 2000, and in 38 out of 40 countries in PISA 2004, in 38 out of 65 countries, boys shows higher success tha girls, while in only 5 out countries this situation is reverse. In 22 countries, the achievement of girls and boys are similar. (OECD, 2004).

When the differences according to gender are considered in terms of item type, it is seen that boys have higher achievement

than girls in multiple-choice item (Ebisine, 2013). Researches have shown that as a cause of this condition, boys tend to take more risk and do not refrain from responding to items even if they are not sure whereas girls prefer to leave the items blank, in addition, Madu (2012) illustrate that boys are superior to girls in complex multiple-choice mathematics items in PISA 2003. However, a different situation is observed in terms of constructed response items Ebisine, (2013) found that girls show higher achievement tha boys in constructed response items. As a result of this situation Abedalaziz, Leng and Alahenadi (2010) stated that girls express their thoughts more effectively because of their language skills are higher lane,

A paper presented at the 33rd Annual Conference of International Association for Education Assessment (IAEA) held in Azerbaijan 16-21, 2007 by Executive Secretary and Deputy Director of Nigeria Education Research Development Council (NERDC) title "The Predictive validity of Public Examinations: A Case Study of Nigeria" revealed that poor prediction might be due to the quality of assessment instruments used by public examination bodies like WAEC, NECO, NABTEB, UME etc hence, recommended that there may be need to address the psychometric properties of the test instruments used in national assessment (Obioma & Salau, 2007).

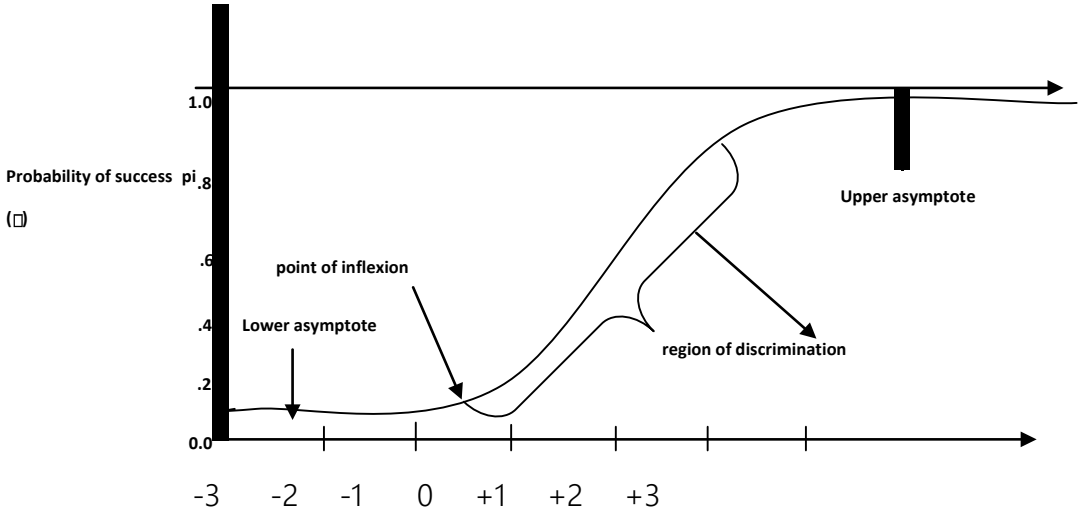
In this study, the West African Senior School Certificate Examination mathematics multiple choice items constituted the area of focus. This paper therefore, looked into test as a major tool in educational assessment especially when it involves making decision(s) about students and its validity in terms of its fairness to the different sub-groups of the examinees.

Differential item functioning is a collection of statistical methods utilized to determine if examination items are appropriate and fair for testing the knowledge of difference group of examinee (e.g male and female). DIF methods, therefore, assess the test-takers' response patterns to specific test item. DIF occurs when a statistically significant difference is evident in the probability that test-takers from the two district groups who have the same underlying ability on the measured construct demonstrate differing probabilities of correctly answering the item (Adebule, 2013). Thus, if DIF is not evident for an item, then there is no item bias. Therefore, DIF is required but not sufficient for item bias. That is if DIF is apparent, then its presence is not sufficient to declare item bias (Roever, 2005). Consequently, a difference in the performance of groups of examinees with different abilities on specific items is not indicative of item bias, but rather item impact (Aborisade, 2016)

Methods of Detecting Test Item Bias in the Measurement of Ability

Many methods of detecting test item bias in the measurement of ability exist these include: item characteristic curve, regression method, chi-square method and transformed item difficulty method among others.

Item characteristic curve approach of detecting test item bias, states that a test is unbiased if all the individuals having the same underlying ability have equal probability of getting the item correct regardless of subgroup membership (Aborisade, 2016). In other words, an item is said to be unbiased if the characteristic curves for the item measured on two groups are identical. If the situation does not hold, then the item is biased and the area between the group ICCs serve as a measure of the item aberrance (Lord, 2002). All item characteristic curves are plotted from test data and form curves of the same general form: from left to right, beginning low, inclining sharply, and leveling off dramatically as illustrated thus:



Item characteristic curve for a 3-parametic logistic model

Two things are to be noted in this diagram above and they are: (i) the slope of the curve is monotonic; that is, it always rises and never exactly horizontal, (ii) the two asymptotes, the upper and the lower, which may approach but never, actually reach 1.00 and 0.00 respectively.

Research Hypotheses

The following research hypotheses were formulated for the study:

1 Items in WASSCE multiple-choice Mathematics administered by WAEC do not function significantly between male and female examinees.

2 Items in WASSCE multiple-choice Mathematics administered by WAEC in 2019 do not significantly function differentially across the school location.

Methodology

The study adopted the descriptive research of the survey type. The population for this study consisted of all senior secondary school students (SSS3) in the 203 public and 235 private secondary school students in Ekiti State. The sample for the study consisted of 1200 senior secondary school III students drawn from 12 public and 12 private secondary schools in Ekiti state.

The sample was selected using multistage sampling procedure. The instrument was reviewed and vetted for face and content validity by the researcher's supervisor and other two experts in Tests and Measurement with Mathematics background who examined the items. This was done for proper scrutiny and vetting. Each of the items of the instrument was equally matched with the general questions and research hypotheses in order to determine whether the instrument measure what it purport to measure. The instrument was adjudged to have both face and content validity for data collection.

The instrument was trial-tested using 120 SSS3 students in three secondary schools outside the sampled schools in Ekiti State. Their responses were scored and analyzed using Kuder-Richardson (KR-20) formula to determine the internal consistency (reliability) of the instrument. The Kuder Richardson formula 20 (KR_{20}) was used to established a reliability coefficient of 0.75 for the objective test, which showed it was reliable The data collected in this study were analyzed using a two-parameter logistic model (2PL) implemented in BILOG-Mg version 3.3.0 software to answer the general questions.

Hypotheses were tested using Bilog M.G. software statistical analysis which was used to generate IRT Item Characteristic Curve (ICC) for each WAEC examination items to show whether items function differentially. Students' responses to WASSCE Mathematics items were calibrated using a two-parameter logistic model (2PL) implemented in BILOG-Mg version 3.3.0 software. This is one of the numerous software available for Item Response calibration of test items in educational assessment. , an independent sample t-test was implemented using R programming language software version 4.1.1. was used to test hypotheses at 0.05 level of significance.

Results

Hypothesis 1

Items in WASSCE multiple-choice Mathematics administered by WAEC in2019 do not function significantly between male and female examinees.

Table 1 presented the outputs of the item response theory (IRT) technique to analyse the DIF of 2019 mathematics items with respect to the Gender of the examinees.

Items in WASSCE multiple-choice Mathematics administered by WAEC in 2019 do not significantly function differentially across the school location.

Table 1: IRT analysis of DIF with respect to Gender

	Gender				
Item	Male	Female	Difference	Decision	Remarks
1	1.667	1.416	0.252	NO DIF	
2	1.315	1.638	-0.323	NO DIF	
3	2.270	2.122	0.149	NO DIF	
4	0.776	0.925	-0.149	NO DIF	

5	0.439	1.327	-0.887	DIF	Favour male students
6	1.466	1.595	-0.129	NO DIF	
7	1.577	2.278	0.700	DIF	Favour male students
8	2.193	1.565	0.628	DIF	Favour female students
9	1.972	1.988	0.016	NO DIF	
10	1.900	1.648	0.252	NO DIF	
11	1.533	1.794	0.261	NO DIF	
12	1.466	1.751	0.285	NO DIF	
13	3.401	3.273	0.127	NO DIF	
14	2.763	3.544	0.781	DIF	Favour male students
15	3.997	2.486	-1.511	DIF	Favour female students
16	3.895	3.363	-0.532	DIF	Favour female students
17	1.690	2.617	0.927	DIF	Favour male students
18	2.482	2.388	0.094	NO DIF	
19	2.851	2.681	0.170	NO DIF	
20	2.472	3.839	1.367	DIF	Favour male students
21	2.379	3.475	1.096	DIF	Favour male students
22	2.482	2.282	0.200	NO DIF	
23	2.564	2.531	0.033	NO DIF	
24	3.099	2.823	0.275	NO DIF	
25	1.035	3.200	2.165	DIF	Favour male students
26	2.882	3.852	0.970	DIF	Favour male students
27	0.776	0.923	-0.148	NO DIF	
28	2.169	2.304	-0.136	NO DIF	
29	1.971	3.879	1.908	DIF	Favour male students
30	2.564	2.786	-0.222	NO DIF	
31	1.876	1.923	-0.047	NO DIF	
32	-0.432	-0.117	-0.315	NO DIF	
33	1.782	2.424	0.642	DIF	Favour male students
34	2.455	2.946	0.491	NO DIF	
35	0.147	0.992	0.845	DIF	Favour male students
36	1.839	1.124	-0.715	DIF	Favour female students
37	1.020	1.722	0.702	DIF	Favour male students
38	1.555	1.597	-0.042	NO DIF	
39	3.734	2.825	-0.909	DIF	Favour female students
40	1.600	1.701	0.101	NO DIF	
41	0.341	0.978	0.637	DIF	Favour male students
42	0.108	0.275	0.166	NO DIF	
43	2.564	2.707	0.144	NO DIF	
44	2.194	3.870	1.676	DIF	Favour male students
45	3.003	3.928	0.925	DIF	Favour male students
46	2.705	2.446	0.259	NO DIF	
47	0.051	0.060	-0.010	NO DIF	
48	2.897	1.131	-1.766	DIF	Favour female students
49	0.584	-0.065	0.649	DIF	Favour female students

50	2.676	2.569	0.107	NO DIF

Table 2a: Descriptive statistics of items that function differentially between male and female examinees

	Gender	Mean	Std. Deviation	Std. Error Mean
Adjusted Difficulty Indices	Male	3.08	0.90	0.34
	Female	1.71	0.94	0.24

Table 2b. Independent sample t-test of items that function differentially between male and female examinees

			Adjusted Difficulty Indices	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		0.15	
	Sig.		0.70	
t-test for Equality of Means	T		-3.22	-3.27
	Df		20.00	12.29
	Sig. (2-tailed)		0.00	0.01
	Mean Difference		-1.37	-1.37
	Std. Error Difference		0.43	0.42
	95% Confidence Interval of the Difference	Lower	-2.25	-2.28
		Upper	-0.48	-0.46

Table 3a presented the descriptive statistics of items that function differentially between male and female examinees. It was shown that items that function in favour of male examinees had ($\overline{X}$= 3.08, SD = 0.90) while items that favour female examinees had ($\overline{X}$= 1.71, SD = 0.94). This result implies that, on average, male examinees had more mathematics test items favouring them than their female counterparts' number of items. Moreover, the mean difference was further confirmed using an independent sample t-test (see Table 5b). The statistics showed that the mean difference was statistically significant (t = -3.22, df = 20, p = 0.00 (i.e p < 0.05)). Consequently, the null hypothesis, which stated that items in WASSCE multiple-choice Mathematics administered by WAEC do not significantly function differentially between male and female examinees, was therefore rejected.

This implies that the male examinees like numerical subjects such as mathematics and have a better understanding (easy) than female contemporaries. This

submission has been established in the literature as well. In addition, female examinees found the items somehow difficult compared to their male counterparts. Also, WAEC must always establish psychometric properties of their test items to conclude that items that constitute their test are devoid of item bias across sub-population of examinees.

Hypothesis 2

Gender of the students will not significantly influence item performance as it relates to difficulty of WASSCE Mathematics administered by WAEC

To answer this hypothesis, item performance indices (see Table 4; columns 2 & 3) for all the items across the Gender of the examinees were used to conduct the analysis. To achieve this feat, an independent sample t-test was implemented using R programming language software version 4.1.1. The results are presented in Tables 8a and 8b as follows;

Table 4a: Descriptive statistics of items performance between male and female examinees

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Item Performance	Male	50	1.93	1.02	0.14
	Female	50	2.15	1.06	0.15

Table 4b. Independent sample t-test items performance between male and female examinees

			Item Performance Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		0.15	
	Sig.		0.70	
t-test for Equality of Means	T		-1.02	-1.02
	Df		98.00	97.89
	Sig. (2-tailed)		0.31	0.31
	Mean Difference		-0.21	-0.21
	Std. Error Difference		0.21	0.21
	95% Confidence Interval of the Difference	Lower	-0.63	-0.63
		Upper	0.20	0.20

Table 8a presented the descriptive statistics of items performance between male and female examinees. It was shown that test item performance (difficulty/threshold) for male examinees had ($\overline{X}$= 1.93, SD = 1.02) while items performance for female examinees had ($\overline{X}$= 2.15, SD = 1.06). This result implies that female examinees found the mathematics test items harder than their male counterparts on average for all the items. Moreover, the mean difference was further confirmed using an independent sample t-test (see Table 8b). The statistics showed that the mean difference was not statistically significant (t = -1.02, df = 98, p = 0.31 (i.e p > 0.05)). Consequently, the null hypothesis, which stated no significant difference in the item performance levels of WASSCE Mathematics items based on Gender, was therefore not rejected. This implies that overall, the level of difficulty of the test items was similar across male and female examinees. However, observed few items across the Gender that displayed the presence of item biasness still needs to work on by the public examining body. Thus, the body must ensure in the near future that all the items in the test give equal opportunity to all the examinees irrespective of their demographic profiles.

Discussion

Based on the finding the result showed that items that function in favour of male examinees had ($\overline{X}$= 3.08, SD = 0.90) while items that favour female examinees had ($\overline{X}$= 1.71, SD = 0.94). This result implies that, on average, male examinees had more mathematics test items favouring them than their female counterparts' number of items. Moreover, the mean difference was further confirmed using an independent

sample t-test (see Table 3b). The statistics showed that the mean difference was statistically significant (t = -3.22, df = 20, p = 0.00 (i.e p < 0.05)). The findings revealed that items that behave in favour of urban examinees had ($\overline{X}$= 2.89, SD = 0.69) while items that favour rural examinees had ($\overline{X}$= 1.85, SD = 0.82). The implication is that, on average, urban examinees had more mathematics test items favouring them than the number of items of their female counterparts. The result showed that items that behave in favour of urban examinees had ($\overline{X}$= 2.89, SD = 0.69) while items that favour rural examinees had ($\overline{X}$= 1.85, SD = 0.82). The implication is that, on average, urban examinees had more mathematics test items favouring them than the number of items of their female counterparts. The findings is in conformity with the findings of Ogbebor & Onuka (2013). However, the study contradicts Adebule (2013) that mathematics items did not function differentially on the basis of the school location of the examinees.

The mean difference was further verified using an independent sample t-test, as shown in Table 5b. The result remarked that the mean difference was statistically significant (t = -3.16, df = 20, p = 0.01 (i.e p < 0.05)). This implies that those examinees from urban areas had a better chance in mathematics than their contemporaries in the rural environment.

Also, WAEC must embrace the item response theory framework of establishing the psychometric properties of their test items before finalising the set of items that make up their test to remove item bias across sub-population of examinees from different school locations.

Conclusion

The findings revealed that the multiple-choice mathematics items administered by WAEC in 2019 function differentially on gender.

Recommendations

Based on the findings of the study, the following recommendations were made:

1. females students should be encouraged and motivated in studying Mathematics.
2 government should pay more attention to schools in rural areas by providing learning materials

References

Aborisade O. J. (2020) A comparative analysis of psychometric properties of mathematics items constructed by WAEC and NECO in Nigeria using item response theory approach. *Academic Journal.* 15(1), 1-7, DOI:10,5897/ERR2019.3850

Adegoke, B.A. (2012) Statistical methods for behavioural and social sciences research. 2nd ed. Ibadan: Evergreen Printing Ventures.

Adegoke, B.A. (2013) comprarisn of item statistics of aphysics achievement test using classical test and item response theory frameworks. *Journal of Education and practice,* 4(22), 87-96.

Amajuoyi, I. J. Joseph, E. U., & Udoh, N. A. (2013). Content validity of May/June West Africa Senior School Certificate Examination (WASSCE) Questions in Chemistry. Journal of Education and Practice, 4(7), 15-21

Anagbogu G. E., Akpan S. M., Ashibi N.I (2011). Analysis of item Difficulty parameters on Item PCharacteristic Curves as a Function of Changes in WAEC and NECO Examination Instruments and Students Ability Parameters in Mathematics Objective Test in Cross River State, Nigeria. *African Journal,* 3(1). 123-134

Awopetu, O.A. & Afolabi, e. r. i. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal,* 12(28), 263-284.

Bandele S.O., & Adewale, A. E. (2013). Comparative Analysis of the difficulty level of WAEC, NECO and NABTEB. *Mediterran Journal of Social Sciences*, 7(4), 761-764.

Battuaz, M., (2015). Wald's test on differential item functioning detection method. Retrieved on 6/4/2019 from http:/www.iiste.org.

Federal Republic of Nigeria (2004). National Po;icy on Education . 4th edition. Lagos: NERDC.

Joseph, C. C. Jason, J.I. & Ron. D. H. , (2015). Overview of classical test theory and item response theory for quantitative assessment of item in developing patient-reported outcome measures. *HHS Public Access*, 36(5). 648-662.

Kolawole, E. B. (2011), Principle of Test Construction and Administration. 2nd Revised Edition. Lagos: Bolabay poblications(Nig). Louis Cohen, Lawrence

Kolawole, E. B. (2002), Assessment of West African Examination Council (WAE) and National Examination Council (NECO) result in both mathematics and English Language in Ekiti State secondary schools Examination. *Mathematics Science Education,* 152-165

Lee, S. H. (2015). Lord Wald test for detecting DIF in multidimensional IRT model: A comparism of two approaches. Unpublished doctoral dissertation. State School Universityof New Jersey, New Jersey

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

Nasiru, S. & Ali A. A. (2019). Gender and School Type Related Item Bias of 2014 NECO English Language Examination in Kano State, Nigeria. *FUDMA Journal of Educational Foundation (FUJEF),* 2(1), 231-240

Ogbogo S. & Opara, I. M. (2019). Differential Item Functioning in English Language Test Using Iytem Response Theory for Ethnic Groups. *Journal of Economics and Sustainable Development.* ISSN 2222-1700 (paper) ISSN 2222-2855