Gender Differential Item Functioning of Test Items of 2015 Joint Senior Secondary II Mathematics Promotion Examination in Ondo State, Nigeria

By Oloda Festus Sunday Smart

Abstract- The research is aimed at finding out whether the multiple-choice mathematic items of Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) administered in 2015 function differentially in terms of the gender (male and female) of the examinees and also investigate the general performance of students in the OSJSSPE multiple–choice mathematics items administered in 2015. The study employed an ex-post facto research design. A sample of 3,135 examinees was selected from a population of 52,922 examinees who sat for the examination using two-stage random sampling techniques. One research question was raised, and one hypothesis formulated and tested for significance at $p<0.05$ level. The analyses revealed that the general performance of the students in the OSJSSPE multiple– choice mathematics items administered in 2015 was high. However, the examinees of equal ability from both male and female students had a different probability of answering some items correctly; thus, the multiple-choice mathematics items of OSJSSPE administered in 2015 functioned differentially on male and female examinees.

Keywords: differential item functioning, item biased, joint senior secondary ii promotion examination, latent trait, mathematics.

GJHSS-G Classification: FOR Code: 139999

GENDERDIFFERENTIALITEMFUNCTIONINGOFTESTITEMSOF2015JOINTSENIORSECONDARYIIMATHEMATICSPROMOTIONEXAMINATIONINONDOSTATENIGERIA

Strictly as per the compliance and regulations of:

Gender Differential Item Functioning of Test Items of 2015 Joint Senior Secondary II Mathematics Promotion Examination in Ondo State, Nigeria

Oloda Festus Sunday Smart

Abstract- The research is aimed at finding out whether the multiple-choice mathematic items of Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) administered in 2015 function differentially in terms of the gender (male and female) of the examinees and also investigate the general performance of students in the OSJSSPE multiple–choice mathematics items administered in 2015. The study employed an ex-post facto research design. A sample of 3,135 examinees was selected from a population of 52,922 examinees who sat for the examination using two-stage random sampling techniques. One research question was raised, and one hypothesis formulated and tested for significance at $p<0.05$ level. The analyses revealed that the general performance of the students in the OSJSSPE multiple–choice mathematics items administered in 2015 was high. However, the examinees of equal ability from both male and female students had a different probability of answering some items correctly; thus, the multiple-choice mathematics items of OSJSSPE administered in 2015 functioned differentially on male and female examinees. It was recommended that examination bodies, test experts, and people charged with developing, validating, and administering tests need to carry out differential item functioning analysis for all items before administration of a test.

Keywords: differential item functioning, item biased, joint senior secondary ii promotion examination, latent trait, mathematics.

I. Introduction

The importance of mathematics in the manpower and technological development of a nation cannot be overemphasized. According to Oloda and Fakinlede (2017), mathematics is a discipline that has various areas of studies which relate with other discipline or subjects such as Basic Science, Basic technology and others which plays important role in the security, sustainability and technological development of any nation. In other words, mathematics is the linchpin in the task of technological development of any nation. Ale & Adetula (2010) highlights the intricate link of mathematics to science, and technology and opined that without the knowledge of mathematics, there would be no science, and without science, there will be no technology and without technology, there is no modern society.

Despite the importance accorded mathematic, it has been observed that students still perform poorly at both internal and external examinations. The poor performance of students at the secondary school level has been a concern to the public in Nigeria. This is usually noticed when the yearly West African Senior School Certificate Examination (WASSCE) and Senior School Certificate Examination (SSCE) conducted by National Examination Council (NECO) results are released. This is evident in the performance of Nigerian candidates in general mathematics in the West African Secondary School Certificate Examination (WASSCE) from 2006 to 2011, which revealed that the percentage of students that passed mathematics at credit level was between 39% and 47%, except for a little improvement of 57.27% in 2008.

One of the factors attributed to the poor performance of students in mathematics achievement tests in schools as observed in literature is gender. There is the general belief that male students seem to perform better than female students in mathematics and mathematics-related subjects. According to Smith & Walker (1988) and Popoola & Ajani (2011), male students perform better than female students in mathematics. Etukudo (2002) opined that boys generally perform better than girls even though they are put into the same classroom situation. However, Cronbach (1977) was of the opinion that boys and girls do not differ much on ability measures. This is not unconnected with the fact that opportunities for development are much more the same until the school leaving age.

Multiple-choice is one of the various types of objective tests that can be used to measure learning outcomes; the score of the multiple-choice item is independent of the subjective influence of the marker or the examiner. Thus, the individual examiner doing the scoring is not required to make the judgment; the score is consistent regardless of the prejudice of the examiner. According to Barret (2001), multiple-choice tests are generally biased towards males, while the female students experience more difficulties with questions involving numerical, spatial, or high reasoning skills. It is generally believed that multiple-choice tests are prone to guessing. Thus, guesswork in most multiple-choice achievement tests has become the order of the day in most institutions. Ojerinde (1985) reported that in Nigeria, most students who do not have a flare for

Author: Ph.D, National Mathematical Centre, Sheda-Kwali, Abuja, Nigeria. e-mail: fssmartol@gmail.com

mathematics would result in guesswork at the secondary and tertiary levels. This is not unconnected with students testing their luck and thus gets out of the examination hall as quickly as possible. However, no statistics exist in this regard as to why students resort to guesswork, but if critically viewed within the test theories, it is assumed to be due to some psychological and situational factors.

Also, Lee, as cited in 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 a combination of both. Conversely, the favored groups are the advantaged group during promotion and admission or selection into science-based courses in the high institutions while, the disadvantaged groups, on the other hand, are disallowed due to some factors tagged extraneous and irrelevant variables that interfere with the measurement of the underlying psychological construct being measured. These factors relate to the group like gender, socio-economic status, location has significant influence on the examinees' response to the item. This implies that the test is multidimensional or measures more than one trait.

Consequently, Smith (1985) referred to the above as measurement disturbance which was classified by Smith (1985) into three categories:

1. Disturbances as a result of characteristics of the person that is independent of the items such as fatigue, boredom, illness, cheating, among others.
2. Disturbances as a result of interaction between the characteristics of the person and the properties of the item, such as item content, item type, guessing, item bias, among others.
3. Disturbances as a result of the properties of the items which are independent of the characteristics of the person, such as the person's ability and item difficulty.

Many research findings have shown that there are differences in the academic performance of separate groups since the inception of testing. An item is biased if it discriminates between members of different groups who have the same ability on what is being measured. Put differently, members of different groups who have the same trait level differ in their score on the item. In his contribution, Plake, as cited in Adebule (2009), defines bias in tests as a situation when items in an achievement tests are found to favour one group over another for reasons not explainable by differences in achievement level between groups. Generally, bias in test items is regarded as a systematic error in measurement. Item bias is the degree to which items that comprise a measurement scale are systematically related to various exogenous variables (e.g., age, gender, location, race, and so on) after conditioning on the latent variable of interest. Items that

show bias in any measuring instrument may affect the properties of the measuring instrument; for instance, test items that are for a group of equal ability should not have statistical differences between the groups. However, if differences exist between the groups, then the validity of the test items is threatened.

A test is gender-biased if men and women with comparable ability levels tend to obtain different scores. Thus, we can say that the test contained items that measure different traits for male and female examinees with comparable abilities. Test fairness is very important in test development; a test is said to be fair if systematic errors (biases) are not present. A systematic error occurs when the construct measured by a test contains some irrelevant elements that can threaten the validity of the test. A test should enable all examinees to have an equal chance to demonstrate personal skills and knowledge vital to the purpose of the test. Items that show bias in any measuring instrument may affect the properties of the measuring instrument. Thus, score 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.

Item response theory (IRT), also known as the latent trait theory or strong true score theory, is a family of latent trait models that are used to establish psychometric properties of items. According to Ojerinde, Popoola, Ojo, & Onyeneho (2012). Item response theory connotes and theoretically assumes that there exist a relatively common trait or characteristic that can be used to determine an individual's ability to succeed with a particular task. Such tasks may be in terms of the individual's response by thinking (cognitive), feeling (affective), and acting (psychomotor). This theory is considered to be one of the most important developments in psychological testing in recent times. Item response theory (IRT) model was designed to solve the problems of classical test theory (CTT); it has been the only theory that gives the valid measurement in terms of test construction and interpretation for assessment of test taker's ability in psychometric analysis. Many of the recent development in testing have their origin in the concept of IRT, such as tailored testing, adaptive computer testing, and improved equating of test forms. According to Pine (1977), Item bias based on the Item response theory (IRT) concept is classified as unbiased if all individuals having the same underlying trait (ability) have an equal probability of getting the item correct regardless of sub-group membership.

Differential item functioning (DIF), also referred to as measurement bias, exists when persons with different group membership but identical overall test scores have different probabilities of solving a test item correctly or giving a certain response on a

questionnaire. According to Kamata and Vaughn (2004), Differential Item Functioning (DIF) occurs whenever people of the same ability level but different groups have different probabilities of endorsing an item. The focus of DIF analysis is on differences in performance between groups that are matched concerning ability, knowledge or skill of interest. Lee (1990) opined that to investigate bias at the item level, developers of large-scale assessments usually conduct a differential item functioning (DIF) analysis. However, not all cases of DIF necessarily have to be interpreted as item bias that will jeopardize the fairness of the test. Instead, DIF is a 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, then 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, students of equal ability level from urban areas display a higher probability of answering an item correctly more than students from rural areas of equal ability level because the content in the test is biased against students from rural areas, then we say the item exhibit DIF and should be considered for modification or removal from the test items.

The Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) was introduced as an intervention measure to reduce the poor performances of students in public examinations. It is only those students that passed the OSJSSPE both in public and private secondary schools that the state government would pay their registration fees and also allow them to sit for the West African Senior School Certificate Examination (WASSCE) and the Senior School Certificate Examination (SSCE) conducted by NECO in the Ondo State. The senior secondary II students that take these examinations are expected to have been exposed to the same course content at the same time frame within the same number of periods. Thus, they are supposed to be of equal probability of success irrespective of gender, location, and type of school.

The 2014 Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) has undergone criticism from stakeholders in the education sector due to the failure rate, which was as bad as having less than half of the senior secondary II students in some schools promoted. Some critics believed that the state government, through the state functionaries in the ministry of education, influenced the result of the examination of the public schools to reduce the number of students in the Senior Secondary School III (SSS3), which will, in turn reduce the total amount to be paid to WAEC or NECO for registration by the state government through the ministry of education. The study investigated the differential item functioning (DIF) of all items in mathematics multiple-choice items of the 2015 Ondo State Joint Senior Secondary II Promotion Examination concerning the gender (male and female). The study was guided by the research question: What is the general performance of the students in the OSJSSPE multiple-choice mathematics items administered in 2015? Also, a Null hypothesis was postulated to guide the study; that is, the Ondo state Joint Senior Secondary II Promotion Examination (OSJSSPE) multiple-choice mathematics items administered in 2015 will not function differentially between males and female examinees.

II. Research Method

The study adopted an ex-post-facto research design. For the ex-post facto design, the researcher started with the observation of the dependent variable and then studied the independent variables in retrospect for their possible relation to an effect on the dependent variable(s). This design is therefore relevant to this study because the researcher does not have direct control over the independent variables since the manifestations had directly occurred and the analysis would be performed on existing data. The population for the study consisted of 52,922 of male and female students in the senior secondary II that responded to the 50 multiple-choice items in mathematics of Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) administered in 2015. The total sample for the study consisted of 3,135 senior secondary II students that responded to the 50 multiple-choice items in mathematics of OSJSSPE administered in 2015 as contained in the Optical Mark Recorder (OMR) sheets from twenty-four selected senior secondary schools in Ondo State, Nigeria, using two-stage sampling techniques. In the first stage, two local government areas (LGA) in each of the three senatorial districts of Ondo State were selected using the purposive sampling technique. In stage two, two public schools (one each from rural and urban areas) and two private schools (one each from rural and urban areas); thus, 12 public schools and 12 private schools were selected using a stratified sampling technique. A total of 24 schools were used for the study. The instruments used for the study were the responses of all the sampled students to the 50 multiple-choice items in mathematics of the OSJSSPE administered in 2015 in the selected schools as contained in the Optical Mark Recorder (OMR) sheets. The items were already subjected to the processes of validation and standardization by the examination Department of Ondo State Ministry of Education. Thus, they were already valid and reliable instruments.

The 50 multiple-choice items in mathematics which the students responded to and were used for this study was constructed, conducted, and administered by the Examination Department of the Ministry of Education

in Ondo State. The OMR sheets comprised of section A and section B. Section A contains the demographic data of the respondents, while section B consisted of all items whose differential item functioning was determined. Descriptive statistics like percentages was used to answer the research question, while inferential statistics like the Welch t-test was used to test the hypothesis at a 0.05 level of significance.

III. RESULTS

The results of the analysis are presented below.
What is the general performance of students in the OSJSSPE multiple-choice mathematics items for 2015?

Table 1: General performance of the students in the OSJSSPE multiple – choice mathematics items for 2015?

Academic Performance	2015	
	N	%
Fail (0 -39)	678	21.6
Pass (40 – 49)	192	6.1
Credit (50 – 69)	579	18.5
Distinction (70 -100)	1686	53.8
Total	3135	100.0

Table 1 shows that 1686 (53.8%) of the students who sat for OSJSSPE had Distinction in Mathematics, 579 (18.5) had credit, 192 (6.1%) passed, while 678 (21.6%) failed. This implies that the General performance of the students in the OSJSSPE multiple-choice mathematics items for 2015 is high. The performance of the students in the OSJSSPE multiple-choice mathematics items for 2015 is further depicted in the figure below.
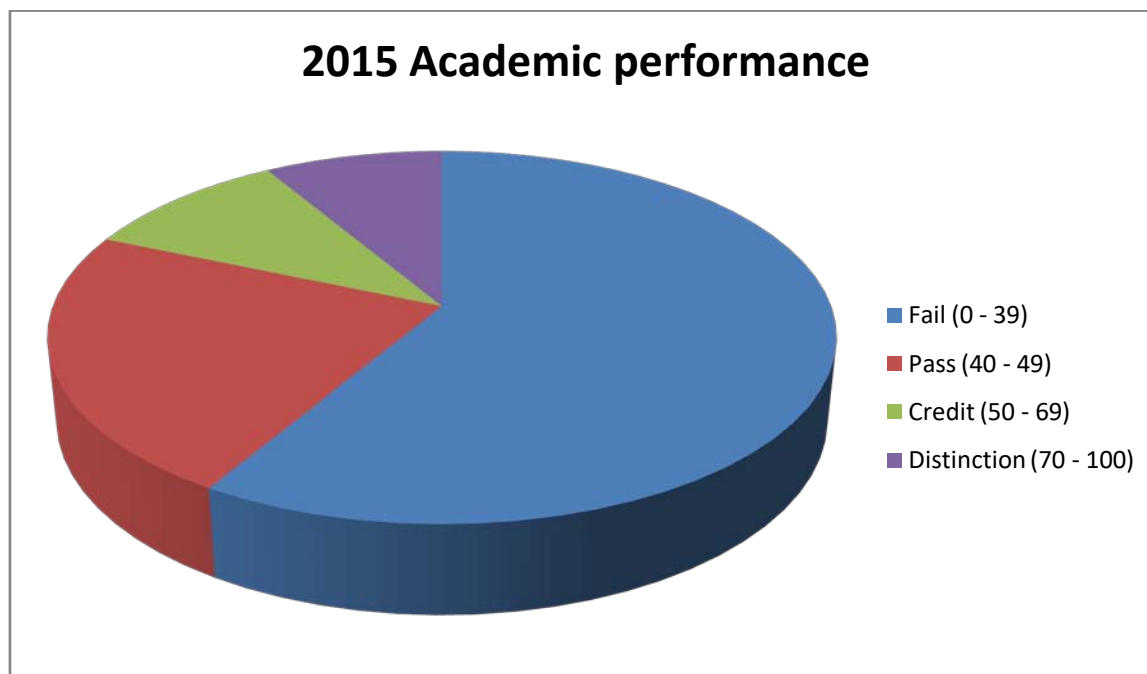

Figure 1: General performance of the students in the OSJSSPE multiple-choice mathematics items for 2015?

Testing the Null hypothesis: The Ondo state Joint Senior Secondary II Promotion Examination (OSJSSPE) multiple – choice mathematics items administered in 2015 will not function differentially between male and female examinees.

Table 2: Summary of DIF Class Specification of 2015 OSJSSPE Based on Gender   male = 1   female = 2

Item Numb	Item Name	Joint S E	DIF Contras	t	df	Prob	Bias Against	Decision
1	I0001	.11	.19	.19	INF	.0729		No DIF
2	I0002	.10	-.03	-.03	INF	.7534		No DIF
3	I0003	.11	.00	.00	INF	1.000	Female	No DIF
4	I0004	.10	-.29	-.29	INF	.0045		DIF
5	I0005	.10	.00	.00	INF	1.000		No DIF
6	I0006	.10	.00	.00	INF	1.000	Male	No DIF
7	I0007	.10	.15	.15	INF	.1605	Female	No DIF
8	I0008	.10	.27	.27	INF	.0077		DIF
9	I0009	.09	-.46	-.46	INF	.0000		DIF
10	I0010	.09	.00	.00	INF	1.000	Female	No DIF
11	I0011	.10	.00	.00	INF	1.000	Female	No DIF
12	I0012	.10	-.18	-.18	INF	.0703	Female	No DIF
13	I0013	.09	-.48	-.48	INF	.0000	Male	DIF
14	I0014	.10	-.21	-.21	INF	.0340	Male	DIF
15	I0015	.10	-.25	-.25	INF	.0117		DIF
16	I0016	.09	.21	.21	INF	.0177	Female	DIF
17	I0017	.09	.36	.36	INF	.0000	Male	DIF
18	I0018	.10	-.06	-.06	INF	.5396	Female	No DIF
19	I0019	.09	-.54	-.54	INF	.0000	Male	DIF
20	I0020	.10	.23	.23	INF	.0235		DIF
21	I0021	.10	-.25	-.25	INF	.0087	Male	DIF
22	I0022	.09	.59	.59	INF	.0000	Female	DIF
23	I0023	.10	.00	.00	INF	1.000		No DIF
24	I0024	.10	-.12	-.12	INF	.2269		No DIF
25	I0025	.09	.32	.32	INF	.0004		DIF
26	I0026	.09	-.76	-.76	INF	.0000		DIF
27	I0027	.10	-.06	-.06	INF	.5236	Male	No DIF
28	I0028	.10	.12	.12	INF	.2319	Male	No DIF
29	I0029	.09	.07	.07	INF	.4365	Male	No DIF
30	I0030	.10	-.18	-.18	INF	.0642	Female	No DIF
31	I0031	.10	.00	.00	INF	1.000	Male	No DIF
32	I0032	.09	.14	.14	INF	.1286	Female	No DIF
33	I0033	.10	.44	.44	INF	.0000		DIF
34	I0034	.10	.42	.42	INF	.0000	Female	DIF
35	I0035	.09	.92	.92	INF	.0000		DIF
36	I0036	.10	-,22	-,22	INF	.0268	Male	DIF
37	I0037	.09	.36	.36	INF	.0001	Female	DIF
38	I0038	.10	-.27	-.27	INF	.0062	Female	DIF
39	I0039	.09	.00	.00	INF	1.000	Female	No DIF
40	I0040	.09	-.20	-.20	INF	.0257	Female	DIF
41	I0041	.10	.12	.12	INF	.2148	Male	No DIF
42	I0042	.10	.52	.52	INF	.0000		DIF
43	I0043	.09	-.57	-.57	INF	.0000	Male	DIF
44	I0044	.09	-.22	-.22	INF	.0131		DIF
45	I0045	.10	-.84	-.84	INF	.0000		DIF
46	I0046	.10	-.49	-.49	INF	.0000		DIF
47	I0047	.09	.62	.62	INF	.0000		DIF
48	I0048	.10	-.02	-.02	INF	.8247		No DIF
49	I0049	.09	-.14	-.14	INF	.1065		No DIF
50	I0050	.10	.61	.61	INF	.0000		DIF

P < 0.05 (Significant)

A cursory look at Table 2 above shows that 28 (56%) items were flagged as DIF against both male and female examinees since their P value were less than 0.05. 13 items were flagged as DIF items against male examinees. The flagged items against male examinees are 8,16,17,20,22,25,33,34,35,37,42,47 and 50. This is because their p – values are less than 0.05 in each case. Similarly, 15 items were flagged as DIF items against female examinees because their p-values are less than 0.05 in each case. The items that flagged as DIF items against female examinees are as follows: items 4,9,13,14,15,19,21,26,36,38,40,43,44,45 and 46 respectively. The study showed that 28 items listed above are statistically functioning differentially between males and females at the significance level of 0.05. The null hypothesis is rejected. This implies that the OSJSSPE multiple-Choice Mathematics items administered in 2015 function differentially between males and females examinees.

IV. Discussion

The result showed that the general performance of the students in the OSJSSPE multiple-choice mathematics items administered in 2015 was high. The findings also revealed that 28 (56%) of the OSJSSPE multiple-choice mathematics items administered in 2015 displayed DIF based on the gender of the examinees using a statistically significant level of P-value of 0.05 (P < 0.05) (Linacre, 2010a). 13 items favoured male examinees while 15 items favoured the female examinees. The null hypothesis is rejected. This implies that the OSJSSPE multiple-choice mathematics items administered in 2015 functioned differentially between male and female examinees. This finding conforms with the finding of Adedoyin (2010) that mathematics items were gender-biased. Also, the result agrees with the findings of Banabas (2012) that mathematics questions used by WAEC contained items that measure different things for male and female examinees with the same mathematics ability. Abedalaziz (2010) and Madu (2012) also reported that mathematics items exhibit differential item functioning in favour of male examinees. However, the study contradicts Adebule (2013) that mathematics items did not function differentially based on the gender of the examinees.

V. Conclusion

The study investigated items that exhibit differential item functioning in the 2015 Joint Senior Secondary II Mathematics Promotion Examination in Ondo State, Nigeria. Based on the findings, it was concluded that the general performance of the students in the OSJSSPE multiple-choice mathematics items administered in 2015 was high. Also, examinees of equal ability from both male and female students had a different probability of answering some items correctly; thus, the multiple-choice mathematics items of OSJSSPE administered in 2015 functioned differentially on male and female examinees.

VI. Recommendations

Based on the findings, the following recommendations are made:

1) Examination bodies, Test experts, and people charged with developing, validating, and administering tests need to carry out differential item functioning analysis for all items before administration of a test.
2) There is a need for teachers, officials of the examination department in the ministry of Education to be trained by experts on item writing. This would acquit them with the processes of finding the psychometric properties and the detection of DIF of each item, which will, in turn, improve the quality of students' assessment in our schools.
3) Detection of DIF has been investigated using OSJSSPE multiple-choice mathematics items in this study. There is a need for replication using other cognitive subjects in OSJSSPE.

Conflict of Interest Statement

The author declares no conflicts of interest regarding the article entitled "Gender Differential Item Functioning of Test Items of 2015 Joint Senior Secondary II Mathematics Promotion Examination in Ondo State, Nigeria".

References Références Referencias

1. Abedalaziz, N. (2010). Agender-related differential item functioning of Mathematics test items. *International Journal of Educational and Psychological Assessment,* 5(1), 101.
2. Adebule, S. O. (2009). Identification of biased items in school certificate mathematics achievement test. International *Journal of Research in Education*, 1 (2).
3. Adebule, S. O. (2013). A study of differential item functioning in Ekiti State Unified Mathematics Examination for Senior Secondary Schools. *Journal of Education and Practice*, 4 (17).
4. Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations: a case study from Bostwana Junior certificate examination in mathematics. *Educational Research and Review Academic Journal,* 5(7).
5. Ale, S.O & Adetula L.O (2010). The National Mathematical Centre and the mathematics improvement project in nation building. *NMC Journal of Mathematical Sciences Education,* 1(1).
6. Berret, S. (2001). Differential item functioning: A case study of first year economics students. *International Education Journal,* 2 (3).
7. Cronbach, L.J (1977): *Educational Psychology*. New York: Harcourt Brace Jovanorichs.

8. Etukudo, U.E (2002). The effect of computer assisted instruction on gender and performance of junior secondary school students in mathematics ABACUS Journal of Mathematical Association of Nigeria, 27(1.)
9. Kamata, A & Vaughn, B.K. (2004). An introduction to differential item functioning analysis. Learning Disabilities: *A Contemporary Journal,* 2.
10. Linacre, J.M (2010a). A users guide to WINSTEPS. Retrieved December, 2016 from *http://www.win steps.com*
11. 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).
12. Ojerinde, A. (1985). Application of Ziller's index of tendency to guess on mathematics test. A paper presented at 2nd*Annual Association of Nigerian Psychological Association*. University of Ife, Ile-Ife, May 8-10.
13. Ojerinde, Dibu. Popoola, Kunmi, Ojo, Ranti. & Onyeneho, Patrick (2012). *Introduction to item response theory-parameter models, estimation and application*, Lagos: Goshen Print Media Limited.
14. Oloda F.S and Fakinlede G.O (2017). Self–assessment in mathematics as correlate of performance of students in basic technology and basic science in ondo state, nigeria: implication for curriculum development. *ABACU The Journal of Mathematical Association of Nigeria,* 42 (2).
15. Pine, S.M. (1977). Applications of item characteristics curve theory to the problems of test bias. In Weis. D. J (Ed) Application for computerized adaptive testing. Mineapolis, Minn: *University of Minnesota Psychometric Methods Program.*
16. Popoola, B.A & Ajani, T.O (2011). Influence of pupils' gender, environment and school ownership status on mathematics performance of secondary school students in Yewa North, Ogun State. *Abacus Journal of Mathematical* Association *of Nigeria*, 36(1).
17. Smith, R. M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational and Psychological Measurement*, 45.