European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available on-line at: www.oapub.org/edu

DOI: 10.46827/ejes.v8i2.3607 Volume 8 │ Issue 2 │ 2021

LOCATION DIFFERENTIAL ITEM FUNCTIONING OF TEST ITEMS OF 2015 JOINT SENIOR SECONDARY II MATHEMATICS PROMOTION EXAMINATION IN ONDO STATE, NIGERIA

Oloda, Festus Sunday Smart[i]
National Mathematical Centre,
Sheda-Kwali, Abuja,
Nigeria

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 location (rural and urban) of the examinees and also investigate if the examination is unidimensional with a view to improving the quality of test items in the examination. 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 multistage random sampling techniques. One research question was raised, and one hypothesis formulated and tested for significance at p<0.05 level. Results of the analyses revealed that the Ondo State Joint Senior Secondary II Promotion Examination multiple – choice mathematics items administered in 2015 was unidimensional. However, the result revealed that some of the OSJSSPE multiple – choice mathematics items administered in 2015 displayed Differential Item Functioning (DIF) based on the school location of the examinees. It was recommended that teachers, officials of the Examination Department in the Ministry of Education should be trained on item writing by test developers, which will in turn improve the quality of students' assessment in secondary schools.

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

1. Introduction

Mathematics is a very desirable tool in all spheres of human endeavour, such as science, engineering, social science and arts. It is the bedrock to which the technological development of any nation is hinged. According to Adetula (2010), mathematics is the

---

[i] Correspondence: email fssmartol@gmail.com

linchpin in the task of national capacity building in science and technology and therefore, any short comings in this subject constitute drawbacks to the achievement of our science and technology objectives. In light of the above, the education policy makers in Nigeria made mathematics a compulsory subject from primary through secondary school and candidates are expected to pass it as part of the basic requirements for tertiary education. Also, a credit pass in mathematics among others is required to gain admission into Nigeria universities.

Mathematics is an indispensable tool for development. The great recognition given to mathematics as a result of its contribution to the development of the society is expected to translate to a satisfactory students' performance in the subject rather; the reverse is the case in Nigeria society. Students still perform poorly at both internal and external examinations; this is usually noticed in our society when the yearly external examinations such as West African Senior School Certificate Examination (WASSCE) and Senior School Certificate Examination (SSCE) conducted by National Examination Council (NECO) results are released. The poor performance of students is often the result of the inability on the part of the Secondary School students to pass external examinations. This is supported by the performance of Nigeria candidates in General Mathematics in the West African Secondary School Certificate Examination (WASSCE) from 2008 to 2016 which showed that the percentage of students that passed mathematics at credit level and above was between 40% and 50% except a little improvement recorded in 2013, 2014 and 2015 where the percentage was 54.27%, 61.97% and 52.17% respectively.

There are many factors attributed to the poor performance of students in mathematics achievement tests in schools. It has been observed in literature that such factors include the environment in which a school is located. According to Popoola & Ajani, (2011) students from urban schools performed better than those from rural schools, the reason is that urban schools are better staffed, with better facilities, students are exposed to good study habits, and they are highly motivated to study under a conducive learning environment. Onah (2011) and Owoeye & Yara (2011) indicated that students from urban schools perform better in science than their counterparts in rural schools. In light of the above, psychologists and test experts have tried to select test contents that would give neither location an advantage nor favour any group of examinees over the other.

There are two theories of measurement, the Classical Test Theory (CTT) and Item Response Theory (IRT). Classical test theory is a theory of measurement which was developed from a simple assumption made by test theories since the inception of testing. It is called theory of true and error scores; this theory of measurement was the foundation for test development for over hundred years ago. The true score is the score that the examinee would have made if there was no error of measurement. Put differently, true score is the difference between the observed score and the error score. According to Lord & Novick (1968), the error score is a random and normally distributed variable that is "*a disturbance that is due to a composite of multiple of factors not controlled in the measurement process*". Classical test theory has dominated the measurement of cognitive abilities in the educational system since the inception of testing. However, there are shortcomings of CTT among which is the items statistics that represent item difficulty and item discriminating power are sample dependent that is, they could vary depending on the sample of the testees that take a particular test. Thus, it is difficult to compare testees results between different tests that are designed to measure same ability hence CTT-based estimate of a testee's ability is not invariant across tests designed to measure that ability. Classical test theory assumes all items are equal; thus, test scores are derived by adding up the number of items scored correctly without taking into cognizance the differences in difficulty, discrimination and vulnerability to guessing while generating test score. In other words, all items are mapped to the same position on a latent continuum that represents the trait under measurement. For instance, a test taker who gave correct answers to the twenty most difficult items has the same score of 20 as the one that answered correctly the twenty easiest items. The implication of this is that the trait or ability level of individual test taker is determined only by the quantity and not the quality of items. Another important shortcoming is lack of invariance in its estimate of the trait level of the item. For instance, the heaviness of a piece of crushed rock is an intrinsic and constant property of the piece, thus, this property should not differ significantly when lifted by people of different weight-lifting ability if otherwise such measurement is faulty. It should be of note that the piece may appear lighter to a person with high weight-lifting ability. However, the intrinsic weight of the piece will not change. Also, the two statistics (item difficulty and item discrimination) that form the basis of many classical test theory analyses are sample dependent.

The emergence of the item response theory (IRT) was to overcome the above limitations. Item response theory (IRT) also known as strong true score theory, the latent trait theory or modern mental test theory. Item response theory assumes that there exists a relatively common trait or characteristic that can be used to determine an individual's ability to perform a particular task. Oloda, Oluwaniyi and Fakinlede (2018) opined that such a task 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. Many of the recent development in testing have their origin in the concept of IRT such as tailored testing, computer adaptive testing and so on.

IRT models are mathematical functions that state explicitly the probability of a discrete outcome, such as a correct response to an item, in terms of person and item parameters. Item parameters include difficulty (location), discrimination (slope) and Pseudo-guessing (lower asymptote). There are currently three operational IRT models; a primary distinction among the models is the number of parameters used to describe items. The operational IRT models are:

i) 1- parameter logistic model (IPLM Rasch model),
ii) 2- parameter logistic model (2PLM), and
iii) 3- parameter logistic model (3PLM).

In item response theory, the two statistics (item difficulty and item discrimination) are technical properties that are used to describe item characteristics curve (ICC), item discrimination describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. An item characteristic curve represents the probability of a correct answer to an item expressed as a function of ability. Thus, the probability of getting an item correct is a function of the amount of ability, it can be observed that: A person with a higher ability has a higher probability of getting an item correct and two persons with the same ability have an equal probability of getting an item correct.

Test is an instrument used to elicit a sample of behaviour from which the general behaviour is inferred. There are various kinds of tests that can be used to measure learning outcomes; these are in the form of essay or objective. Oloda and Adebule (2017) defined an objective test as a type of test which the score is independent of the subjective influence of the examiner or the marker, that is, the individual marker doing the scoring is not required to make any judgment; the score is consistent regardless of the prejudice of the marker. There are also different forms of objective tests such as true or false, yes or no, filling the gap, matching type and multiple-choice. Among the various objective test formats, the multiple-choice format is the most widely used in school-based assessments, entrance examinations and standardized tests. It is generally believed that multiple-choice tests are prone to guessing, thus guesswork in most multiple-choice achievement tests have become the order of the day in most institutions. Test items should only discriminate among examinees based on the differences in the ability under measurement. Ogbebor (2012) opined that, examinees of the same latent trait should respond to test items correctly irrespective of their gender, school location and school type. A test should enable all examinees to have an equal chance to demonstrate personal skills and knowledge that are vital to the purpose of the test. Thus, items that show bias in any measuring instrument may affect the properties of the measuring instrument.

Nworgu (2011) opined that current research data have implicated tests used in national and regional examinations as functioning differently with respect to different subgroups. That is, students' scores in such examinations are determined largely by the group to which a student belongs and not by students' ability. The 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. 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 Oloda and Adebule (2017), Differential Item Functioning (DIF) occurs whenever examinees of the same ability level but different groups have different probabilities of answering any item correctly. The focus of DIF analysis is on differences in performance between groups that are matched with respect to ability, knowledge or skill of interest.

Differential item functioning (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, boys display higher probability of answering any item correctly more than girls of equal ability level due to the fact that the content in the test is biased against girls, then we say the item exhibit differential item functioning (DIF) and should be considered for modification or removal from the test item.

The Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) was introduced during the 2012/2013 academic session 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 would be allowed to sit for the West African Senior Certificate Examination (WASSCE) and the Senior School Certificate Examination (SSCE) conducted by NECO. The Ondo state government usually pays the WASSCE registration fees for the students in public secondary schools. The senior secondary II students that takes 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 location.

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 with respect to the location (urban and rural) of the examinees. The study was guided by the research question: Are the mathematics multiple-choice items of the 2015 Ondo state Joint Senior Secondary II Promotion Examination unidimensional? 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 on the school location of the examinees.

2. 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 dependent variable(s). The population for the study consisted of 52,922 of male and female students in the senior secondary II located in both urban and rural areas 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 purposive sampling technique (where at least one public and one private school are located at both rural and urban areas in each LGA). 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 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. Winsteps 3.91 statistical analysis software of the Rasch model was used for the appraisal of the unidimensionality of measure while inferential statistics like Welch t-test was used to test the hypothesis at 0.05 level of significance.

3. Results

The results of the analysis are presented below.

Research Question: Are the mathematics multiple-choice items of the 2015 Ondo state Joint Senior Secondary II Promotion Examination unidimensional?

Table 1: 2015 Standardized Residual
Variance in Eigenvalue Units = Item Information Units

Observations	**Eigenvalue**	**Observed**	**Expected**
Total raw variance in observations	75.4736	100%	100%
Raw variance explained by measures	25.4736	33.8%	33.9%
Raw variance explain by person	15.0321	19.9%	20.0%
Raw variance explain by items	10.4415	13.8%	13.9%
Raw unexplained variance (total)	50.0000	66.2%	66.1%
Unexplained variance in 1st Contrast	3.5625	4.7%	
Unexplained variance in 2nd Contrast	2.2406	3.0%	
Unexplained variance in 3rd Contrast	2.0383	2.7%	
Unexplained variance in 4th Contrast	1.8773	2.5%	
Unexplained variance in 5th Contrast	1.8494	2.5%	

Rasch assumes the unidimensionality of the measure and not of the data (ie the person's response). A cursory look at table 1 shows that the total raw variance explained by measure was 33.8% while the total raw variance unexplained by the measures was 66.2%

(noise). This is in contrast with the submission of Wright, 1996 and Linacre, 2009 that if the difference between the variance explained by Rasch dimension and the noise is considerably high, the unidimensionality of the test is supported. Also, unexplained variance in 1st contrast eigenvalue 3.5625 and 4.7% recorded negates the condition that if the first contrast has "units" (eigenvalue) less than 3 (for a reasonable length test) then the test is probably unidimensional (Linacre, 2006). However, this case is supported by one of the Rule of Thumb which states that variance explained by measures greater than 60% is good and unexplained variance in 1st contrast (size) less than 3.0 is good and unexplained variance in the 1st contrast less than 5% is good. However, the first contrast eigenvalue 3.56 implies that the contrast between opposing factors has the strength of about 4 items. It is more likely we have a "fuzzy" "broad" dimension since we are dealing with mathematics which includes arithmetic, algebra, geometry, word problems and so on. Rasch measures are always unidimensional and linear, their concurrence with the data is never perfect, it is always approximate (Linacre, 2006) Thus, the Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) multiple – choice mathematics items administered in 2015 did not violate unidimensionality assumptions.

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 on the school location of the examinees.

Table 2: Summary of DIF Class Specification of 2015 OSJSSPE based on Location (urban = 1; rural = 2)

Item Number	Item Name	Joint S E	DIF Contrast	Rasch – Welch t	df	Prob	Bias Against	Decision
1	I0001	.12	-.17	-1.35	INF	.1781		No DIF
2	I0002	.11	.13	1.15	INF	.2503		No DIF
3	I0003	.12	.00	.00	INF	1.000		No DIF
4	I0004	.12	-.41	-3.45	INF	.0006*	Rural	DIF
5	I0005	.11	-.08	-.77	INF	.4444		No DIF
6	I0006	.12	-.11	-.94	INF	.3464		No DIF
7	I0007	.12	-.53	-4.31	INF	.0000*	Rural	DIF
8	I0008	.11	-.37	-3.26	INF	.0011*	Rural	DIF
9	I0009	.09	.03	.33	INF	.7430		No DIF
10	I0010	.09	.30	3.23	INF	.0013*	Urban	DIF
11	I0011	.11	-.16	-1.42	INF	.1548		No DIF
12	I0012	.11	.09	.83	INF	.4040		
13	I0013	.10	-.30	-2.94	INF	.0033*	Rural	DIF
14	I0014	.12	-.37	-3.11	INF	.0019*	Rural	DIF
15	I0015	.11	-.32	-2.83	INF	.0047*	Rural	DIF
16	I0016	.09	.54	5.93	INF	.0000*	Urban	DIF
17	I0017	.09	.23	2.51	INF	.0122*	Urban	DIF
18	I0018	.11	.02	.22	INF	.8245		No DIF
19	I0019	.10	.04	.42	INF	.6778		No DIF
20	I0020	.11	-.06	-.49	INF	.6274		No DIF
21	I0021	.10	.09	.86	INF	.3880		No DIF
22	I0022	.09	.07	.76	INF	.4457		No DIF
23	I0023	.10	.10	.92	INF	3584		No DIF
24	I0024	.11	-.33	-2.90	INF	.0038*	Rural	DIF
25	I0025	.09	.00	.00	INF	1.000		No DIF

26	I0026	.10	-.42	-4.15	INF	.0000*	Rural	DIF
27	I0027	.11	.05	.42	INF	.6728		No DIF
28	I0028	.11	-.05	-.45	INF	.6523		No DIF
29	I0029	.10	-.40	-4.10	INF	.0000*	Rural	DIF
30	I0030	.10	.03	.24	INF	.8085		No DIF
31	I0031	.12	-.58	-4.91	INF	.0000*	Rural	DIF
32	I0032	.10	-.29	-3.02	INF	.0026*	Rural	DIF
33	I0033	.12	-.59	-5.04	INF	.0000*	Rural	DIF
34	I0034	.10	.05	.45	INF	.6544		No DIF
35	I0035	.09	.29	3.18	INF	.0015*	Urban	DIF
36	I0036	.11	.05	.44	INF	.6593		No DIF
37	I0037	.09	.16	1.74	INF	.0820		No DIF
38	I0038	.11	.12	1.16	INF	.2445		No DIF
39	I0039	.10	-.12	-1.27	INF	.2054		No DIF
40	I0040	.09	.11	1.14	INF	.2542		No DIF
41	I0041	.10	.48	4.75	INF	.0000*	Urban	DIF
42	I0042	.10	.72	7.50	INF	.0000*	Urban	DIF
43	I0043	.10	-.28	-2.72	INF	.0066*	Rural	DIF
44	I0044	.09	.29	3.10	INF	.0020*	Urban	DIF
45	I0045	.11	.08	.72	INF	.4714		No DIF
46	I0046	.10	.61	5.94	INF	.0000*	Urban	DIF
47	I0047	.10	.08	.86	INF	.3924		No DIF
48	I0048	.11	.00	.00	INF	1.000		No DIF
49	I0049	.09	.29	3.19	INF	.0014*	Urban	DIF
50	I0050	S.10	-.31	-2.98	INF	.0029*	Rural	DIF

P < 0.05 (Significant)

Table 2 shows the Differential Item Functioning (DIF) statistics of the Rasch model for each of the 50 items for location. There is an incidence of DIF if the probability is less than 0.05, thus there is an incidence of DIF in 23 items at 0.05 level of significance. That is, 46% of the 2015 OSJSSPE multiple-choice mathematics items functioned differentially for urban and rural locations. The DIF items against urban are 10, 12, 16, 35, 41, 42, 44, 46 and 49 while the DIF items against rural are 4, 7, 8, 13, 14, 15, 24, 26, 29, 31, 32, 33, 43, and 50. The study showed that the above-listed items are statistically significant by functioning differentially between urban and rural at significant level of 0.05. The null hypothesis is rejected. This implies that the Ondo state Joint Senior Secondary II Promotion Examination (OSJSSPE) multiple-Choice Mathematics items administered in 2015 function differentially on the school location of the examinees.

4. Discussion

Rasch assumes the unidimensionality of the measure and not of the data (ie the person's response). The total raw variance explained by measure was 33.8% while the total raw variance unexplained by the measures was 66.2% (noise). This is in contrast with the submission of Wright, 1996 and Linacre, 2009 that if the difference between the variance explained by Rasch dimension and the noise is considerably high, the unidimensionality of the test is supported. However, this case is supported by one of the Rule of Thumb which states that variance explained by measures greater than 60% is good and unexplained variance in 1st contrast (size) less than 3.0 is good and unexplained variance in the 1st contrast less than 5% is good. However, the first contrast eigenvalue 3.56 implies that the contrast between opposing factors has the strength of about 4 items. It is more likely we have a "fuzzy" "broad" dimension since we're dealing with mathematics which includes arithmetic, algebra, geometry, word problems and so on. Rasch measures are always unidimensional and linear, their concurrence with the data is never perfect, it is always approximate (Linacre, 2006) Thus, the OSJSSPE multiple–choice mathematics items administered in 2015 did not violate unidimensionality assumptions.

The findings showed that 23 (46%) of the 2015 Ondo state Joint Senior Secondary II Promotion Examination (OSJSSPE) multiple-choice mathematics items flag off DIF with respect to school location of the examinees using a statistically significant level of P-value of 0.05 ($P < 0.05$) (Linacre, 2010a). Nine (9) items favoured examinees from rural locations while 14 items favoured those from urban locations. The null hypothesis is rejected. This implies that the OSJSSPE multiple-choice mathematics items administered in 2015 functioned differentially on the school location of the examinees. This finding is in conformity with the finding of Ndifon, Umoinyang & Idiku (accessed, 21 March, 2017) that there is existence of differential item functioning in mathematics between students from urban and rural areas. Also, the result agrees with the findings of Ogbebor & Onuka (2013) that the 2010 NECO Economics questions were biased in relation to school location of the examinees. The finding is also in consonance with the result of the study carried out by Patrick and Bright (2018) that 40% of the 2014 Basic Education Certificate Examination (BECE) Social Studies multiple-choice items function differentially for urban and rural examinees. However, the study contradicts Adebule (2013) that mathematics items did not function differentially on the basis of the school location of the examinees. Also, the findings of this study disagree with Lee and Mclritire (2001) whose findings revealed that there is no significant difference between performance of rural students and urban students.

5. Conclusion

The study investigated items that exhibits 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 multiple-choice mathematics items of OSJSSPE administered in 2015 is unidimenssional. Also, examinees of equal ability from both urban and rural schools had different probability of answering some items correctly; thus, the multiple-choice mathematics items of OSJSSPE administered in 2015 functioned differentially on the school location of the examinees.

5.1 Recommendations

Based on the findings, the following recommendations are made:

1) There is the need for teachers, officials of the examination department in the ministry of Education to be trained on item writing by test developers. 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 schools.

2) Test experts and developers should explore the use of the Rasch model to detect items that flag off differential item functioning (DIF) in various subject examinations.
3) Examination bodies, test experts and people charged with the responsibility of developing, validating and administering tests need to carry out differential item functioning analysis for all items before administration of test.

Conflict of Interest Statement

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

About the Author

Oloda, Festus Sunday Smart (PhD) is a Research Fellow I in the Mathematical Sciences Education Programme at the National Mathematical Centre (NMC), Abuja, Nigeria. He obtained a Certificate in Education (NCE, Economics/Mathematics) from Lagos State College of Education, Ijanikin, Lagos. He bagged Bachelor of Science Education (Mathematics) degree from University of Lagos, Akoka-Lagos. He obtained his Master in Education (Tests and Measurement) and Doctor of Philosophy (Tests and Measurement) from Ekiti State University, Ado-Ekiti, Nigeria. He has conducted extensive research in self-assessment of students in mathematics, Differential item functioning (DIF) of test items in mathematics, and Determinants of student's performance in mathematics. Dr. Oloda, F.S.S is a member of Ethno- Mathematics Research group, a joint project of the National Mathematical Centre (NMC) and Prof. Nkechi Madona Agwu, a Carnegie African Diaspora Fellow in the USA. Also, he is a member of Mathematical Sciences Education Research group at the National Mathematical Centre, Abuja, Nigeria. He is a registered member of the Teacher's Registration Council of Nigeria (TRCN) and Mathematical Association of Nigeria (MAN).

References

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.

Adetula, O. L. (2010). Mathematics improvement project (NMC-MIP): a project transforming the Mathematics performance of students. *International Journal of Contemporary Educational Research,* 2(2), 104-117

Canice, E. O. (2000). *Introduction to educational and psychological measurement.* Lagos: Behenu Press & Publishers.

Ebel, R. L. (1965). *Measuring Educational achievement Englewood Cliff*, New Jersey: Prentice Hall. Cited in: Canice, E. O. 2000. *Introduction to educational and psychological measurement.* Lagos, Shomolu: Behenu.

Lee J. and Mclritire W. (2001). Interstate variation in mathematics achievement of rural and non-rural students. Journal of research education 16 (3)

Linacre, J. M. (2010a). A users' guide to WINSTEPS. http://www.winsteps.com [Accessed December, 2016]

Lord, F. M. & Norvik, M. R. (1968). *Statistical theories of mental test scores, Reading,* MA: Addison- Wesley.

Ndifon, B. O., Umoinyang, I. E & Idiku, F. O. (2010). Differential item functioning of 2010 Junior Secondary School Certificate Mathematics Examination in southern educational zone of Cross River State, Nigeria. www.iaea.info/documents/paper_3fc7465c.pdf. [Accessed, 21st March, 2017]

Nworgu, B. G. (2011). Differential item functioning: A critical issue in regional quality assurance. *A Paper Presented in NAERA Conference*.

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

Oloda, F. S. S. & Adebule S. O. (2017). Differential item functioning of test items of 2014 Joint Senior Secondary II Mathematics Promotion Examination in Ondo State, Nigeria. *EKSUJOE Journal of Education* 7(3), 40-47

Oloda, F. S. S., Oluwaniyi, S. D. and Fakinlede, G. O. (2018). An introduction to a paradigm shift from classical test theory (CTT) to item response theory (IRT). *ABACUS, Journal of the Mathematics Association of Nigeria,* 43(1)

Onah, E. F. (2011). Influence of sex and school location on students' achievement in agricultural science. *African Journal of Science, Technology and Mathematics Education (AJSTME),* 1 (1) 96-102

Owoeye, J. S. and Yara, P. O. (2011). School location and academic achievement of secondary school in Ekiti State, Nigeria. Asian social science (ASS) 7(5), 170-173.

Patrick U. Osadebe and Bright Agbure (2018). Assessment of differential item functioning in social studies multiple choice questions in Basic Education Certificate Examination. European Journal of Education studies 4(9), 236-256.

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), 56-65.