Assessing the dimensionality and local independence of WASSCE 2018 Mathematics objective tests scores in Lagos State, Nigeria

O. M. Alade, S. Aletan, B. S. Sokenu
Department of Educational Foundations (with Educational Psychology),
University of Lagos, Nigeria

Abstract
This research work is a descriptive survey design type aimed at investigating the extent to which the 2018 West African Senior Secondary Certificate Examination Mathematics objective tests items fulfil the unidimensional and local independence assumptions. All the year three Senior Secondary School students (SS3) of the 1,647 senior secondary schools that registered for the 2020 WASSCE in Lagos State formed the population for the study. The sample for this investigation comprised 1334 (SS3) students from 18 schools which were randomly selected from the six educational districts of Lagos State (three schools from each educational district). Two research questions were raised and analysed using Stout's test of essential unidimensionality, conducted in DIMTEST version 1.2.0 and Yen Q3 statistics. Findings indicated that the 2018 WASSCE (SC) Mathematics objective items did not meet the assumption of unidimensionality and that there is more than one dimension that accounted for the variation observed in examinees' responses to the Mathematics test items. In addition, some items violated the assumption of local independence because their correlation residuals are greater than >0.2. Hence, test developers and the external examining bodies are encouraged to be open to the use of IRT and standard statistical methods for estimating test parameters to ensure the quality of the assessment instrument.
Keywords: Dimensionality, Local Independence, WASSCE, IRT, Mathematics

Introduction
Assessment is a fundamental practice in every teaching and learning situation that happens in school and outside the school system. It is a process that helps instructors make decisions about the learners' progress and to improve instructional decisions. Appraisal or an assessment is the way toward gathering and deciphering proof to make judgment about a student's learning, (National Research Council NRC, 2001). When an assessment is properly carried out, it influences learning and when made authentic, it provides feedback which is revised to improve learning. According to Odumosu (2012), assessment can be made valid, fair, ethical, feasible and an efficient tool for learning using multiple measures.

In assessment practice in Nigeria, different tests are developed by teachers and are used to determine the ability, proficiency or curriculum-related achievements of students, (Agah, 2013). A test, according to Nworgu (2011), is a structured situation comprising a set of questions to which an individual is expected to respond, and each question in the test has a preferred answer. Nworgu (2011), further noted that the behaviour of an individual is quantified based on his responses to the questions. Onuka and Durowoju (2011) also described a test as an instrument which can be used in distinguishing a few characteristics, traits, qualities, attributes, etc possessed by a person, an object or a thing. Based on the above explanations, it can be deduced that a test is a single occasion, undimensional, timed exercise, usually in structured response or free response item format. It is used quantitatively and qualitatively to ascertain the magnitude of the construct one possesses. It is one of the instruments used for the measurement and evaluation of students' educational achievements.

Olonode (2016), opined that the present Nigerian educational framework is by all accounts a long way from accomplishing the ideal instructive objectives and goals as there are recognizable confirmations of decrease in standard as proved by the fluctuating outcomes in public examinations. Proof of this is noticeable in schools and public examinations particularly in a subject like Mathematics. Aliyu (2015), defined Mathematics as the solid

foundation of Science and Technology. Adeyegba in Ojerinde, (2013) observed that there is not really any area of science that doesn't utilize Mathematical ideas to clarify its own concepts, theories or models. Actually, the most punctual progress of humanity came through scientific controls including the utilization of numbers. In spite of the important role

Mathematics plays in everyday life, it has remained one of the least successful subjects in Nigerian schools. The trend of students' performance in Mathematics in the West African Senior Secondary Certificate Examination (WASSCE) affirms this assertion. The performance is fluctuating and not stable as revealed in table 1.

Table 1: Statistics of Students' Performance in WASSCE Mathematics between 2008 and 2018

SUBJECT MATHEMATICS	YEAR	TOTAL ENTRY	TOTAL SAT	TOTAL CREDIT 1-6	TOTAL PASS 7-8	FAIL
	2008	1292890	1268213 (98.09)	726398 (57.28)	302266 (23.83)	218618 (17.24)
	2009	1373009	1348528 (98.22)	634382 (47.04)	344635 (25.56)	315738 (23.41)
	2010	1331374	1306535 (98.13)	548065 (41.95)	363920 (27.85)	355382 (27.20)
	2011	1540141	1508965 (97.98)	608866 (40.35)	474664 (31.46)	421412 (27.93)
	2012	1695878	1658357 (97.79)	838879 (50.58)	478519 (28.86)	298742 (18.01)
	2013	1686990	1656527 (98.19)	1897655 (54.18)	462176 (27.90)	245263 (14.80)
	2014	1655794	1632377 (98.59)	1011608 (61.97)	357555 (21.90)	211941 (12.98)
	2015	1602362	1581420 (98.69)	901845 (57.02)	425628 (26.91)	219759 (13.89)
	2016	1552169	1536643 (98.99)	1056045 (68.72)	309526 (20.14)	131755 (8.57)
	2017	1565106	1550348 (99.05)	1276782 (82.35)	160623 (10.36)	44874 (2.89)
	2018	1576465	1563457 (99.17)	1208457 (77.29)	220804 (14.12)	77009 (4.92)

Source: The West African Examinations Council (WAEC), Test Development Division, Ogba, Lagos

A quick glance at Table 1 revealed that there have been fluctuating performances of students in the WASSCE between 2008 and 2018; Given that a credit pass in Mathematics is a compulsory requirement for admission into Nigerian tertiary institutions for all courses, it is certain that a large number of candidates who completed secondary school between 2008 and 2018 were not eligible for admissions into tertiary institutions.

The unstable performance may be linked to a number of factors such as difficulty of the subject, students' attitude to the subject and even the quality of items used in the assessment of West African Senior Secondary Certificate Examination (WASSCE) (Chukwudi, 2019). Another area of challenge as identified by Olonode, (2016), is the observation that

examiners are often faced with great challenges of how to assess students to obtain scores with fair and equitable treatment and reduce item difficulty especially in Mathematics. As indicated by Rover (2005), a fair and equitable test is one which gives all the testees equal chance to showcase the abilities and information which they have gained and which are pertinent to the inspiration driving the test. This matter of test quality therefore brings to the front burner, the subject of test dimensionality and local independence.

Test dimensionality refers to the number of traits underlying a test that accounts for variation in examinees' test performance (Oguoma, Metibemu & Okoye, 2016) A test is termed unidimensional if there is only one trait (dimension) accounting for variation in

examinees' test performance (Yu, Popp, DiGangi & Jannasch-Pennel, 2007). When there are two or more constructs accounting for variation in examinee's test performance, the test is termed multidimensional (Oguoma, Metibemu & Okoye, 2016). According to Tate (2003), the appraisal of the quantity of constructs emanating from the interaction of examinees with test items ought to be a significant piece of the advancement, assessment, and upkeep of public examinations. Oguoma, Metibemu & Okoye, (2016), assessed the appropriateness of scoring 2014 WASSCE Mathematics multiple choice tests using Classical Test Theory (CTT). Their findings suggest that 2014 WASSCE Mathematics test was multidimensional. Awopetu & Afolabi (2016) also conducted a study using factor analysis; they concluded that the 2011 NECO SSCE Mathematics items were unidimensional. Mfonobong & Umobong, (2017), conducted a research on the dimensionality of National Examinations' Council's Biology Examinations: The discoveries of their investigation uncovered that most of the items measured unidimensional constructs; and were valid and reliable. Abd-El-Fattah, Al-Sinami and Shourbagi (2014) completed an exploration on the utilization of Rasch examination in analyzing the number of constructs and differential item functioning of the Arabic edition of the apparent physical ability scale for pupils. Their findings revealed that the mean scores of male and female participants are different. Another variable of interest in this study is local independence of test items.

Local item independence guarantees that the probability of a testee acquiring the right response to a test item must not rely upon the appropriate response given to different items in the test (Okwilagwe and Ogunrinde, 2017). According to Okwilagwe & Ogunrinde (2017), the relationship between examinee's abilities and local item independence ensures that the likelihood of an examinee getting a test item right should not be subject to the response given to different items in the test. An investigation directed by Joshua, Ubi & Abang (2011) on 'Item Local independence in selection examination in Nigeria: implications for

assessment of regional education', utilising Yen Q3 statistics, found that the UTME mathematics items for 2000, 2001, 2002 and 2003 years were, as it were, locally independent.

A variety of tests are used in education but the use of multiple choice tests is a trend all over the world. There are basically two types of tests for evaluating or measuring the students' abilities and these are essay type and objective type tests. Among the various forms of objective type test, multiple choice item is of much concern to this study. Multiple choice items are hugely flexible and easily create items that exactly match a learning outcome which may emphasize factual knowledge, stress understanding or application of knowledge.

Multiple choice test items usually have many options for each question, the testee is supposed to select the best choice among a set of four options (Udeani, 2011). Multiple choice test items are most often used to assess knowledge of a particular topic and they typically appear on achievement tests which measure simple memorization. Multiple choice questions are used for many reasons, but most importantly, the multiple choice questions lend themselves to a thorough quantitative analysis using the item analysis. It is essential that in order for scores to accurately represent the degree to which a student has attained an educational objective, the test should be reliable and valid.

This study is anchored majorly on Item Response Theory (IRT) by Lord (1980). IRT focuses on how specific test items function in surveying constructs. The theory makes it possible to scale test items for difficulty, to construct parallel types of tests and to accommodate versatile computerised testing. The central idea of IRT rests upon the individual items of a test rather than upon the aggregate of the item responses, for example, a test score. It expresses that the likelihood of getting a question right is a function of a latent trait or ability. People with lower ability have less of a chance, while people with high ability are very likely to answer many questions correctly. The theory is based on three assumptions: (a) unidimensionality of the trait to be measured (b) local independence of items (c) item

characteristics curve (ICC). That is, the response of an examinee to an item can be displayed by a mathematical item response function (IRF). Thus, IRT entails an item analysis procedure which ensures that questions are of appropriate standard before selection for test. This theory is relevant to this research work because it is used for test construction, test calibration and in addressing test- item biasness. It additionally empowers the examiners to put their examinees in the right capacity level on the grounds that the IRT analysis can describe the items and the abilities of the examinees.

Statement of the problem

Literature reveals that several factors determine students' Mathematics academic achievement, among which are; school quality, teacher quality, teaching methodology, school environment, student-related factors, etc. The effort at improving students' achievement, performance and skill is on the increase with lesser effort concentrated on improving test items performance.

However, despite the solutions proffered or implemented by researchers, the situation is yet to improve. An aspect which research in Mathematics education has not focused much on is how assessment practices in terms of test dimensionality and local independence affect students respond to test items and their ultimate scores. This shows that the procedures and frameworks for test development and how test items are scored can have impact on students' performance. It is as a result of this that research must shift the emphasis towards the appraisal techniques being adopted by classroom teachers and public examination bodies. This is because the assessment practices may be one reason why students are not performing up to expectation in Mathematics. No doubt, poorly-worded test items with many constructs may be confusing to test takers and if tests are not properly scored, examinee's final scores in a test may not reflect their actual ability. It is based on this that this study is assessing the dimensionality and the local independence of the 2018 WASSCE (SC) Mathematics items in Lagos State.

Research Questions

In doing this investigation, the accompanying questions guided the study.
1. What exact degree did the 2018 WASSCE (SC) Mathematics objective items obey the assumption of unidimensionality under the IRT framework?
2. To what extent did the 2018 WASSCE (SC) Mathematics objective items obey the assumption of Item local independence under the IRT framework?

Research Methodology

Research Design

The design used was descriptive survey design.

Population

All the SS3 students in public and non- public secondary schools in Lagos State form the population for this investigation.

Sample and Sampling Techniques

Multistage sampling process was used to select 1,334 students from eighteen schools, (three schools from each of the six educational districts in Lagos State) comprising public, private, urban and rural areas.

Instrumentation

The research instrument used for the study adopted 50 multiple-choice 2018 WASSCE Mathematics items. Each of the items has options A to D, with one correct option and three distracters. The test content covered SS3 Mathematics scheme in Lagos State. The items were deemed reliable and valid because they have been validated and standardized by the West Africa Examination Council (WAEC).

Analysis of Data

To assess dimensionality of test data, two types are evident. At the point when a score is generated for a test, there is a verifiable presumption that the items share a typical essential constructs; it implies that unidimensionality is tenable otherwise, multidimensionality is evident. For this study, Stout's test of essential unidimensionality, conducted in DIMTEST version 1.2.0., was used to establish assumption of dimensionality of the 2018 WASSCE (SC). So as to accomplish this, test items were partitioned into two subtests. The

first of these subtests, known as Assessment Subtest (AT), is picked to such a degree that the items contained inside the portion are homogenous to one another, yet as dimensionally unmistakable from the rest of the items. The subsequent subtest is called Partitioning Subtest (PT), which contains all items not used in Assessment subtest (AT).

Moreover, items that might form a secondary dimension, the Assessment Subtest, were chosen empirically, utilizing the HCA/CCPROX cluster procedure and DETECT statistic in DIMTEST, and the examinee cluster was tested to see if it was dimensionally distinct from the remainder of the test. A random sample of 30% of the examinees was used to choose the Assessment Subtest, and the rest of the sample was used for the dimensionality test (Demars, 2010). The null and alternative hypotheses tested by DIMTEST are given by (Stout et al., 1996).
They are: H_0: AT $\cup$ PT satisfies essential unidimensionality (d = 1)
H_i: AT $\cup$ PT fails to satisfy d = 1
H_0 posits that the AT and PT partitions assess the same dominant underlying dimension, while the alternative hypothesis (H_0 suggests that the items in the AT partition are best represented by a dimension that is distinct from that driving responses to the PT items. Likewise, after

remedy for bias, the test statistic, T, is thought to be normally distributed. The null hypothesis for T is that the responses are unidimensional, hence, acceptance of the null hypothesis implies unidimensionality. If otherwise, then multidimensionality is present (Demars, 2010).

Item local independence of 2018 WASSCE (SC) Mathematics objective items were estimated using correlation residual with Yen Q3. The items with correlation residual values equal or greater than 0.2 were considered locally dependent and should be deleted (Demars, 2010).

Results and Interpretation
Research Question 1: To what extent did the 2018 WASSCE (SC) Mathematics objective items obey the assumption of unidimensionality under the IRT framework?

In order to answer this research question, Stout's test of essential unidimensionality, conducted in DIMTEST version 1.2.0., was used to establish the assumption of dimensionality of the 2018 WASSCE (SC).

Table 2 presents the result of the DIMTEST statistic of unidimensionality of the 2018 WASSCE (SC) items.

Table 2: Stout's test of essential unidimensionality Statistics of 2018 WASSCE (SC) Mathematics objective items

TL	TGbar	T	P-value
13.6623	2.1999	11.41	0.0000

The result above, showed that the null hypothesis was rejected (T = 11.41, p = 0.000). This is because the test statistic is greater than the critical value of 1.96 from the normal distribution. This leads to the conclusion that the AT and the PT are dimensionally distinct. Consequently, the test is inherently multidimensional in nature. The implication is that more than one trait accounted for the observed variation in the examinees' performance in the test.

Research Question 2: To what extent did the 2018 WASSCE (SC) Mathematics objective items obey the assumption of Item local independence under the IRT framework?
To answer this question, item local independence of the 2018 WASSCE (SC) Mathematics objective items were estimated using correlation residual with Yen Q3. Table 3 presents (an abridged) Yen Q3 Statistics for the test items.

Item	sc1	sc2	sc3	sc4	sc5	sc6	sc7	sc8	sc9	sc10
sc1	1.00									
sc2	-0.04	1.00								
sc3	0.02	0.20	1.00							
sc4	0.00	-0.04	0.07	1.00						
sc5	-0.06	-0.08	-0.12	0.05	1.00					
sc6	0.03	-0.19	-0.07	0.04	-0.08	1.00				
sc7	0.04	0.12	-0.15	0.06	0.02	0.13	1.00			
sc8	-0.03	0.14	0.13	0.07	-0.17	-0.12	-0.07	1.00		
sc9	-0.03	-0.05	-0.02	-0.08	-0.01	0.05	0.07	0.05	1.00	
sc10	-0.29	0.03	-0.20	-0.05	-0.03	-0.02	-0.08	-0.08	0.03	1.00
+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+
sc49	-0.02	-0.08	-0.03	-0.02	0.01	0.00	0.04	-0.08	-0.01	-0.08
sc50	0.00	-0.06	-0.03	0.01	-0.03	0.03	0.03	-0.03	-0.04	-0.06

Item	sc11	sc12	sc13	sc14	sc15	sc16	sc17	sc18	sc19	sc20
sc11	1.00									
sc12	0.03	1.00								
sc13	-0.22	-0.12	1.00							
sc14	-0.03	0.13	0.01	1.00						
sc15	0.02	0.21	-0.20	0.11	1.00					
sc16	-0.05	0.06	-0.14	-0.11	0.17	1.00				
sc17	-0.16	0.09	-0.08	0.20	-0.11	0.00	1.00			
sc18	0.18	0.01	-0.13	-0.08	0.21	0.02	-0.14	1.00		
sc19	-0.14	0.01	0.13	-0.07	-0.15	-0.12	-0.22	-0.09	1.00	
sc20	-0.09	-0.20	0.18	-0.10	-0.10	-0.27	-0.07	-0.07	0.18	1.00
+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+
sc49	0.05	0.04	-0.05	-0.03	0.08	0.01	-0.08	0.02	0.01	-0.07
sc50	0.05	0.01	0.02	-0.01	0.02	0.02	-0.05	0.03	0.00	-0.04

Item	sc21	sc22	sc23	sc24	sc25	sc26	sc27	sc28	sc29	sc30
sc21	1.00									
sc22	-0.16	1.00								
sc23	-0.24	0.01	1.00							
sc24	0.10	0.28	-0.21	1.00						
sc25	0.11	-0.01	-0.01	0.08	1.00					
sc26	-0.11	0.09	-0.03	0.07	0.19	1.00				
sc27	-0.01	0.07	0.05	0.20	-0.09	0.17	1.00			
sc28	0.05	0.09	-0.09	0.08	0.27	0.15	0.15	1.00		
sc29	-0.08	0.15	0.14	0.18	0.01	0.15	0.19	0.12	1.00	
sc30	0.04	-0.13	-0.06	-0.11	0.23	-0.11	-0.14	0.02	-0.24	1.00
+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+
sc49	0.00	-0.08	-0.03	-0.03	0.06	-0.03	-0.06	-0.01	-0.09	0.07
sc50	-0.03	-0.04	-0.02	-0.05	0.03	-0.01	-0.05	-0.01	-0.03	0.05

Item	sc31	sc32	sc33	sc34	sc35	sc36	sc37	sc38	sc39	sc40
sc31	1.00									
sc32	-0.02	1.00								
sc33	-0.10	-0.04	1.00							
sc34	-0.13	-0.16	-0.15	1.00						
sc35	0.36	-0.06	-0.14	0.09	1.00					
sc36	0.19	0.02	0.03	-0.11	-0.04	1.00				
sc37	-0.02	0.07	-0.08	-0.18	-0.29	0.15	1.00			
sc38	-0.01	-0.11	0.31	-0.09	0.00	0.09	0.07	1.00		
sc39	-0.02	-0.01	-0.06	0.01	-0.07	-0.25	0.16	-0.04	1.00	
sc40	0.13	0.02	0.13	-0.28	0.02	0.26	0.03	0.07	0.04	1.00
+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+
sc49	0.00	-0.04	0.01	-0.04	-0.02	-0.07	-0.01	-0.02	0.02	-0.02
sc50	-0.01	-0.01	0.00	-0.01	-0.03	-0.07	-0.02	-0.01	0.00	0.00

Item	sc41	sc42	sc43	sc44	sc45	sc46	sc47	sc48	sc49	sc50
sc41	1.00									
sc42	-0.31	1.00								
sc43	0.25	-0.06	1.00							
sc44	-0.14	-0.03	0.28	1.00						
sc45	-0.33	0.52	-0.54	-0.21	1.00					
sc46	-0.21	-0.33	0.11	0.36	-0.01	1.00				
sc47	-0.04	0.38	0.12	0.18	0.05	-0.31	1.00			
sc48	0.42	0.45	0.23	-0.30	0.09	-0.64	0.31	1.00		
sc49	-0.83	0.27	-0.19	0.20	0.38	0.35	-0.09	-0.44	1.00	
sc50	-0.56	0.18	0.04	0.25	0.11	0.32	-0.31	-0.27	0.64	1.00

Table 2 depicts the correlation item residual of item local independence of the 2018 WASSCE (SC) Mathematics objective with the use of Yen Q3 statistics. From the item correlation residual table (table 3), the following pairs of items violate the assumption of local item independence. Thus: item 2 and 3, 8 and 17; 12 and 15, 5 and 19, 4 and 20, 3 and 22 and so on. This is because the correlation residual (values) of these items are equal or greater than 0.2. Therefore, about 28 paired items violated the assumption of item local independence while 1176 items are locally independent in measuring the underlying trait among the examinees used for the study. This implies that the paired items depend on one another as a result of their correlation item residual that is equal or greater than 0.2.

Discussion of Findings
The results of the 2018 WASSCE (SC) Mathematics multiple choice test reveal an infringement of the unidimensionality assumption implicit in the IRT. The findings collaborated the finding of Oguoma, Metibemu and Okoye (2016), who assessed the dimensionality of 2014 West African Secondary School Certificate Examination mathematics objective test items in Imo State, they found that the test is multidimensional, however, the finding negates that of the study of Awopetu and Afolabi (2016) who, after using factor analysis, concluded that the 2011 NECO SSCE Mathematics items were unidimensional. Their contradicting findings could be attributed to the different statistical tools employed in the studies.

Furthermore, the result from the findings revealed that 28 out of 50 items of the 2018 WASSCE (SC) Mathematics multiple choice test violated the assumption of local independence. This result negates the finding of Ubi & Abang (2011) on item Local independence on selection examination in Nigeria, using Yen Q3 statistics. They

discovered that the UME Mathematics items for 2000, 2001, 2002 and 2003 years were locally independent.

Conclusion
Based on the results obtained in the course of this study, it was concluded that the 2018 WASSCE (SC) Mathematics multiple choice test violated the assumptions of unidimensionality and local independence.

Recommendations
It is therefore, recommended that unidimentionality and local independence of standardized Mathematics Achievement tests should not be assumed. The test items should be subjected to both item analyses based on IRT theory to ensure that quality items are selected (items that are particularly unidimensional and have local independence).

References
Abd-El-Fattah, S. M., Al-Sinani, Y. & Shourbagi, S. E. (2014). The use of Rasch analysis in examining the dimensionality structure and differential item functioning of the Arabic version of the perceived physical Ability Scale for children. Australian Journal of Educational & Developmental Psychology, 14, 29-44.

Agah, J. J. (2013). Relative efficiency of test scores equating methods in the comparison of students continuous assessment measures. Education Research Network for West and Central Africa, 3(2)340-348.

Aliyu, R. T. (2013). Development and validation of Mathematics achievement test using the Rasch Model. An Unpublished Ph.D Thesis in Delta State University, Abraka, Nigeria.

Awopeju, O.A. & Afolabi, E.R. (2016). Comparative analysis of classical test theory and item response theory based, item parameter estimates of senior school certificate mathematics examination. Journal of Educational Research, 96(6), 340-349.

Chukwudi, D. U., (2019). Differential Item Functioning and Item Analysis of Basic Education Certificate Examination Mathematics Multiple Choice Test Items in Enugu, Nigeria. (Unpublished P. hd thesis), Department of Educational Foundations University of Lagos.

Lord, F.M. (1980). Application of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Mfonobong, E. U, &Udeme, E. T. (2017). Dimensionality of National Examinations' Council's Biology Examinations: Assessing Test quality in modern trend approach. African Journal of Theory and Practice of Educational Assessment, 5(1), 14-30

National Research Council (NRC)(2001). Retrieved from http://www:nrcnigeria.org.html.

Nworgu, B. G. (2011). Differential item functioning: A critical issue in regional quality assurance. Journal of Educational Assessment in Africa, 6(3), 112-123.

Odumosu, J. N. (2012). Application of three-parameter logistic model in the calibration of a mathematics achievement test. Journal of Educational Assessment in Africa, 7(3), 162-172.

Oguoma, C. C., Metibemu, M. A. & Okoye, R. O. (2016). An assessment of the dimensionality of 2014 West African Secondary Certificate Examination Mathematics Objective Test Scores in Imo State, Nigeria. African Journal of Theory and Practice of Educational Assessment, 4(5), 18-33.

Ojerinde, D. (2013). Classical test theory (CTT) versus item-response theory (IRT): An evaluation of comparability of item analysis result. A paper presented at the Institute of Education University of Ibadan 23rd May, 2013 lecture series.

Okwilagwe, E. A. &Ogunrinde, M. O. (2017). Assessment of unidimensionality and local independence of WAEC and NECO 2013 Geography achievement tests. African Journal of Theory and Practice of Educational Assessment, 5(1), 30-45.

Olonode, P.O. (2016). Equating 2014 senior school certificate mathematics examinations of West African Examinations Council and National Examinations Council in Lagos State. Nigeria. (Unpublished P. hd thesis), International center for Educational Evaluation University of Ibadan.

Onuka, O. A. & Durowoju, O. E. (2011). Curtailing examination fraud for improving quality assurance in the African examination system. A paper presented at the 29th Conference of the Association for Educational Assessment in Africa, Nairobi, Kenya 1-5 August.

Rover, C.(2005). 'That's not fair!' Fairness, bias and differential item functioning in language testing. Retrieved from http://www2.hawaii.edu/rover/brownbag.

Tate, R. L. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. Applied Psychological Measurement, 27, 159-203.

Ubi, A. T & Abang B. (2011). Calibration of mathematics and geography items for joint command schools promotion examinations for Nigerian Army Education Corps in Nigeria. African Journal of Theory and Practice of Educational Assessment, 2(2), 40-60.

Udeani, U. (2011). Test development: a practical guide for teachers. Lagos: Sibon