ISSN 2705-2559
E-ISSN 2705-2567

CONFORMITY OF NATIONAL EXAMINATIONS COUNCIL MATHEMATICS MULTIPLE CHOICE TEST ITEMS TO THE ASSUMPTION OF ITEM RESPONSE THEORY

BY
Adediwura Alaba Adeyemi: Department of Educational Foundations and Counselling, Faculty of Education, ObafemiAwolowo University, Ile-Ife, Nigeria; E-mail: yemtoy20002000@gmail.com
&
Jimoh Kasali: Department of Educational Foundations and Counselling, Faculty of Education, ObafemiAwolowo University, Ile-Ife, Nigeria; E-mail: jimoh.bukola@yahoo.com

Abstract
The study established if 2016 NECO Mathematics multiple choice test items satisfy unidimensionality and local independence assumptions of Item Response Theory; and it determined the nature of item characteristics curve in the 2016 NECO mathematics items. Ex-post facto design was adopted for the study. The population for the study comprised candidates who enrolled and sat for June/July SSCE 2016 NECO Mathematics Examination. The sample comprised 276,338 candidates from three purposively selected Geo political zones in Nigeria. The results showed that the 2016 NECO Mathematics test was essentially unidimensional. However, nine of the items are interdependent on each other. Also, the results revealed that 12 items do not produce a monotonically increasing function in the form of forward S shape for the probability of correct response recommended for good item under IRT.
Keywords: Item characteristic curve, Item local independence, Item response theory and Items unidimensionality

Introduction

Mathematics is known to be a required subject in Nigerian schools particularly at primary and secondary levels. This is as a result of its significance as stated in the National Policy on Education (NPE); all science, technological and vocational subjects need it. Universally, Mathematics is generally considered to be among the crucial subjects in the classroom environment. Mathematics is regarded as a subject with a close connection or relationship to other subjects such as research and innovation Federal Republic of Nigeria (FRN, 2013). As far as learners can tell, it is still one of the most complex subjects in the classroom. The preceding five years continuously saw the poor decline of the test takers' performances. The negative performances in NECO particularly in Mathematics become a matter of urgent concern by the educators, policy planners and all stakeholders including the students who are the primary victims. The poor performances caused a 'blame game' within the country; accusation and counter accusation flooded the National Dailies. Who was to blame; the teachers or the students? Policy makers on educational reforms, the parents, the government or NECO body itself? In the year 2013, only 708,314 students, who constituted 68.5% out of 1,034,264 that sat for the NECO examination, were able to pass Mathematics at credit level. By the following year (i.e., 2014) there was continuous decline of students' performance in the general examination with heavy casualties in the subject of Mathematics as only 667,529 (68.1%) passed with credit within 979,886 candidates who sat for the Mathematics test. There was however, a substantial improvement in the year 2015 as 776,378 (80%) passed Mathematics test at credit level out of 969,490 students who sat for the mathematics examination.

Gegbe and Koroma (2014) discovered that students' helpless achievements in General and Further Mathematics can be traced back to a variety of factors, including teaching methods, educators' attitudes toward the subject, and teachers' lack of understanding of the fundamental concepts in Further Mathematics, among others. Odili and Asuru (2011) believed that senior high school students are lacking in all concepts of Further Mathematics, including indices and logarithms, and that the majority of the candidates did not cover the syllabus before taking the exam. Although several factors had being identified to be responsible for the poor and dwindling students' performance in mathematics, research have not paid adequate attention on the conformity of test items with assumptionsof Item Response Theory (IRT). Item Response Theory ((IRT) is the most significant development in psychometrics. It 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. According to Ojerinde, et. (2012), Item response Theory is based on the latent trait theory, which incorporates measurements assumptions about examinee, item and test performance and how this performance relates to knowledge as measured by individual items on a test. Item response Theory attempt to model the relationship between an unobservable variable referred to as that examinee's

ISSN 2705-2559
E-ISSN 2705-2567

ability or trait and the possibility of the examinee correctly responding to any particular test item (Lord, 1980).

Item Response Theory (IRT) models are commonly used to model the latent traits associated with a set of items or questions in a test or survey. In education, testing is an inherent part of the curriculum as an assessment tool to evaluate students' subject matter proficiency and skill development. IRT attempts simultaneously to examine the appropriateness of the questions in terms of measuring what they are designed to measure and the proficiency of the respondents. In making use of IRT to perform all these activities, test items must conform with the three assumptions of IRT. They are; unidimensionality, local independence and item characteristic curve. Unidimensionality is one of the assumptions of item response theory. The assumption of unidimensionality is that a set of items or a test measure only one latent ability (Kyung, 2013). This implies that the achievement of each test-taker is assumed to be governed by one latent trait, referred to as ability. Since peoples' intellectual and individual attributes as well as the effect of test achievement cannot always be regulated, it is not always possible to meet up with this assumption. One can then discuss the unidimensionality of a test only when it evaluates one dominant ability (Hambleton, Swaminathan & Rogers, 1991).

To fulfill this assumption of unidimensionality, one can employ any of these eleven methods of testing for unidimensionality as cited by Ojerinde and Ifewulu (2012): Cronbach analysis test, Factor analysis, Eigenvalue test, Random baseline test, Biserial test, Factor loading test, Congruence test, Part/Whole test, Communality test, Vector frequency test and Confirmatory Factor Analysis (F.A) and Structural Equation Modelling (SEM) test, using the Statistical Package for Social Science (SPSS). A support for the unidimensionality of the items in the scale is provided when the model fits the data well and there are no noteworthy residual relationships (no such relationship higher than or equal to 0.20) (Ojerinde, 2013). Any violation of this assumption would result in inadequacy of the model in describing the data, thereby leading to an unreliable estimation of the examinee's ability. Therefore, the correct specification of the number of the latent dimensions is directly tied to the construct validity of the test (Rijn, Sinharay, Haberman& Johnson, 2016).

Local independence which is the second basic assumption of IRT, means that, a correct or wrong answer to one achievement test item should not lead to a correct or wrong answer to another item. This means that there should not be any correlation between two items after the effect of the underlying trait of testees is partialized. The items should only be correlated through the latent trait that the test is measuring (Lord & Novick, 1968). If there are significant correlations among the items after the contribution of the latent trait is partialized, then the items are locally dependent or there is a subsidiary dimension in the measurement which is not accounted for by the main Rasch model-based dimension (Lee, 2004). In other words, performance on the items then depends, to some extent, on a trait other than the Rasch model dimension; this means a violation of the assumptions of local independence and unidimensionality. If the assumption of local independence is violated, any statistical analysis based on it would be misleading. In addition, it is not clear what constructs the item responses reflect; consequently, it is not clear how to combine those responses into a single test score, whether item response theory is being used or not (Wang, Cheng & Wilson, 2005).

Local independence also leads to little errors of estimate. This may be a serious problem in computerised adaptive testing, where standard error estimation are the criteria for terminating the test. It has the potential to cause the test to end prematurely (Zenisky, Hambleton & Sireci, 2003). A simple examination of individual response to such national examinations in Nigeria reveals the nature of bias item. A test should be unidimensional in order to be free of errors. The assumption of unidimensionality is designed to measure a single property or skill for all examinees. The assumption of unidimensionality is the most complex and most restrictive assumption of item response theory. There are possibilities for this assumption to be violated. For example, two items designed to assess multiplication skill in mathematics could be as follows: (i). what is 3×7? (ii). what is the product of three and seven? Item (i) requires only knowledge in mathematical operations, while item (ii) requires for its solution, a specific amount of reading competence as well as knowledge of mathematical operations. When different attributes are being measured as in item (ii), the issue of item bias enters into consideration if such item is administered to two different groups and the responses of one of the groups are dependent on the secondary skill. This type of item measures different types of skills among different groups.

ISSN 2705-2559
E-ISSN 2705-2567

The Item Characteristic Curve (ICC) is a mathematical function that connects the probability of success on an item to the ability level by the item set or test that contains it. The item characteristics curve provides a sensible separation among different item response theory (Baker, 2001). It is a crucial design square of item response theory; the wide scope of different creates of the theory rely on this curve (Baker, 2001). The second distinct property is item discrimination, which describes how well it can distinguish between examinees who have ability underneath the item location and those who have ability over the item location. The steepness of the item's characteristic curve in its middle section is reflected by this property.According to Momanyi, Too, and Simiyu (2015), many variables play a role in students' academic performance with little empirical effort on suitability of the items for the measurement of intending latent traits. Based on this, this study examined the conformity of 2016 NECO Mathematics multiple choice questions with IRT assumptions.

Purpose of the Study

The broad objective of this study is to establish 2016 NECO Mathematics multiple choice items conformity with the IRT assumptions. Specifically, the objectives of the study are to:

1. establish if 2016 NECO Mathematics multiple choice test items satisfy unidimensionality;
2. ascertain2016 NECO Mathematics multiple choice test items conformity with local independence assumption of Item Response Theory (IRT);
3. determine the nature of item characteristic curve in the 2016 NECO Mathematics items;

Research Questions

1. To what extent does the 2016 NECO Mathematics multiple choice test items conform to the assumption of unidimensionality of Item Response Theory?
2. To what extent does the 2016 NECO Mathematics multiple choice test items conform to the assumption of Local independence of Item Response Theory?
3. What is the nature of item characteristics curve in the 2016 NECO Mathematics items?

Methodology

The study employed the ex-post facto design method. The population for the study comprised all 1,022,474 candidates who enrolled and took the National Examinations Council's (NECO) Senior School Certificate Mathematics Examination in June/July 2016.The sample consisted of 276,338 students' selected using multistage sampling procedure.A total of three zones (South-west, South-east and North-west) were randomly selected from the six geopolitical zones of Nigeria and from each of the zones, three states were randomly selected. Ogun, Ondo and Osun were selected from the South-west while Anambra, Enugu and Imo states were selected from the South-east and Kano, Sokoto and Zamfara states were selected from the North-west. Every candidate that registered and sat for the examination in the nine selected states made up the sample. The research instruments used for the study were Optical Marks Record Sheets for the National Examinations Council (NECO) June/July 2016 SSCE Mathematics objectives items. The OMR sheets contained the responses of examinees of the NECO June/July 2016 Mathematics objective items. The examination had 60 items in the objectives format and were scored dichotomously.The Stout test of unidimensionality and Yen Q3 statistics were used to analyse collected data.

Results

Research question 1: To what extent does the 2016 NECO Mathematics multiple choice test items conform to the assumption of Unidimensionality?

To answer this research question, the responses of the sampled examinees to the 2016 NECO Mathematics multiple choice was subjected to test Stout's test of essential unidimensionality (Stout, 1987). It also implemented the Supplementary Item Response Theory Models, sirtpackage (Robitzsch, 2019) of R Language and Environment for statistical computing (R Core, 2019) was According to Jang & Roussos, 2007; Zhang, 2007 the dimensionality of a test data judged based on the following:

Strong multidimensionality	DETECT > 1.00
Moderate multidimensionality	.40 < DETECT < 1.00
Weak multidimensionality	.20 < DETECT < .40
Essential unidimensionality	DETECT < .20
Maximum value under simple structure	ASSI=1 RATIO=1

ISSN 2705-2559
E-ISSN 2705-2567

Essential deviation from unidimensionality ASSI > .25 RATIO > .36
Essential unidimensionality ASSI < .25 RATIO < .36.
The result of the unidimensionality assessment of the NECO Mathematics test is presented in Table 1.

Table 1 Unidimensionality assessment of 2016 NECO Mathematics test

	Unweighted	Weighted
DETECT	-0.284	-0.284
ASSI	-0.242	-0.242
RATIO	-0.266	-0.266

Table1 showed the essential unidimensionality assessment of the 2016 NECO Mathematics multiple choice test. The table showed that the 2016 NECO mathematics test was essentially unidimensional (maximum DETECT value = -0.284 (< .20), ASSI = -0.242 (< 0.25) and RATIO = -0.266 (< 0.36)). The result showed that the 2016 NECO Mathematics test was essentially unidimensional. The implication of the result is that the 2016 NECO Mathematics multiple choice test was unidimensional and that one dominant ability accounted for the variations observed in the performance of the candidates that sat for 2016 NECO Mathematics multiple choice test.

Research Question 2: To what extent does the 2016 NECO Mathematics multiple choice test items conform to the assumption of item local independence?

Assessment of model-data fit involves two stages: 1. fitting the data test to the four IRT models and thereafter, the fitness of the four models to the data set are compared. The model that produced the best fit to the data is adjudged the model that fit the data. To achieve this feat, several measures can be applied. According to Oguoma, Metibemu and Okoye (2016) and (Finch & French, 2015), prominent among the measures include Chi-square difference test and information indices. Information indices are simply measures of variance not explained by a model, with an added penalty for model complexity. Among the most popular of these indices are the Akaike information criterion (AIC; Akaike, 1973), the Bayesianinformation criterion (BIC; Schwarz, 1978), and the sample-size-adjusted BIC (SBIC; Enders &Tofihi, 2008). These information indices are computed using the -2loglikelihood chi-square value and is interpreted in such way that the model with the lower value exhibits a better fit to the data. In addition, the chi-square and likelihood ratio goodness of fit tests the null hypothesis that two nested models provide the same fit to a set of data. A statistically significant likelihood indicates a difference in the models under examination. Table 2 presents the result of the model-data fit assessment.

Table 2: Model-data fit assessment of 2016 NECO Mathematics test items

PL model	AIC	AICc	SABIC	HQ	BIC	logLik	X^2	df	p
	Comparing 1 and 2 PL model								
1	18109920	18109920	18110368	18110107	18110562	-9054899	820636.1	59	0
2	17289402	17289402	17290282	17289769	17290664	-8644581			
	Comparing 2 and 3PL models								
2	17289402	17289402	17290282	17289769	17290664	-8644581	169158.5	60	0
3	17120304	17120305	17121624	17120854	17122196	-8559972			
	Comparing 3 and 4PL models								
3	17120304	17120305	17121624	17120854	17122196	-8559972	216159.2	60	0
4	16904265	16904266	16906025	16904998	16906788	-8451893			

Table 2 showed the model-data fit assessment of 2016 NECO Mathematics test items.The table showed that when the fitness of 1PL and 2PL model data were compared, the result showed that the 2PL had AIC =17289402, SABIC =17290282, BIC =17290664values that were less than the AIC =18109920, SABIC =18110368, BIC =18110562values of the 1PL. In addition, the Likelihood ratio test revealed that 2PL fitted the data better than 1PL was statistically significant (χ^2 (59) = 820636.1, p < 0.05). These results showed that the 2PL model fitted the data better than the 1PL model. In search for a better model for the test data, the fitness of 2PL model to the 2016 NECO Mathematics test itemswere in turn compared to the fitness of 3PL model to the test data. The result showed that the 3PL model fitted the data better than the 2PL model (3PL model's AIC = 17120304, SABIC = 17121624, BIC = 17122196 values were

ISSN 2705-2559
E-ISSN 2705-2567

respectively less than the 2PL model's AIC =17289402, SABIC =17290282, BIC =17290664; the Likelihood ratio test that 3PL model fitted the data better than 2PL model was statistically significant, (χ^2 (60) = 169158.5, p < 0.05)). Furthermore, in search for a better model for the test data, the fitness of 3PL model to the 2016 NECO Mathematics test items was in turn compared to the fitness of 4PL model to the test data. The result showed that the 4PL model fitted the data better than the 3PL model (4PL model's AIC = 16904265, SABIC = 17121624, BIC = 17122196 values were respectively less than the 3PL model's AIC = 17120304, SABIC = 16906025, BIC = 16906788; the Likelihood ratio test that 4PL model fitted the data better than 3PL model was statistically significant, (χ^2 (60) = 216159.2, p < 0.05)). The result revealed that unidimensional four-parameter logistic model fitted the 2016 NECO Mathematics test items. Thus, the test was calibrated using four-parameter logistic model.The result of the ILD assessment based the 4-PL that was found to fit the test data is presented in the table that follows.

Table 3: Item Local Independence of 2016 NECO Mathematics Test Items

	it21	it25	it27	it28	it29	it31	it36	it42	it44	it45	it47	it48	it58	it60
it9			0.3					0.3						
it19	0.3													
it22		0.3												
it25						0.3								
it27				0.3										
it31							0.3							
it33								0.3	0.3			0.3		0.3
it40										0.3				
it42									0.4	0.4				
it44										0.5				
it45														
it51													0.3	

Table 3 showed the abridged inter-item correlation of the items coefficient existing among the residuals of pairs items of the NECO test data (Yen Q3 statistic). The table showed that pair of items respectively showed correlation coefficient greater than the bench mark (i.e., 0.2). The implication of this is that one of the paired of items violated the item local independence. The result showed that item 9 and 27, item 9 and 43, item 19 and 21, item 22 and 25, item 25 and 31, item 28 and 27, item 31 and 36, item 33 and 42, item 33 and 44, item 33 and 48, item 33 and 60, item 40 and 43, item 42 and 44, item 42 and 45, item 44 and 45 and item 51 and 58 had inter-residual correlation greater than the benchmark, 0.2. The implication of the result is that one of the item pairs violated the item local independence. Therefore, the result showed that 9 items depended on other items in the 60-item test. The implication of the result is that item 9, 19, 22, 31, 28, 33, 44, 45 and 58 violated item local assumption.

Research Question 3: What is the nature of item characteristics curves of the 2016 NECO Mathematics items?

Figure 1 to 12 showed the item characteristics of the items of the NECO test.

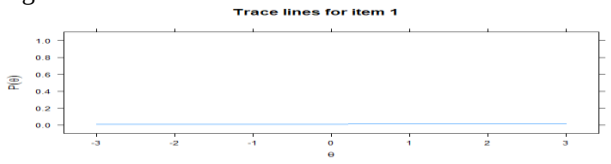

Figure 1: Item Characteristics Curve for Item 1

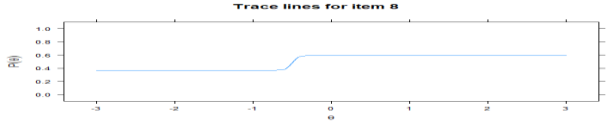

Figure 2: Item characteristics curve for item 8

ISSN 2705-2559
E-ISSN 2705-2567

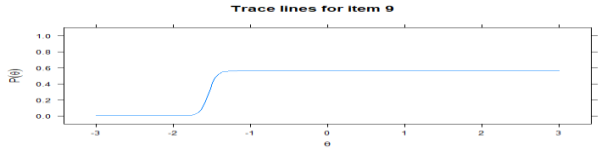

Figure 3: Item characteristics curve for item 9

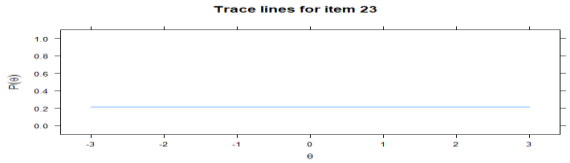

Figure 4: Item characteristics curve for item 23

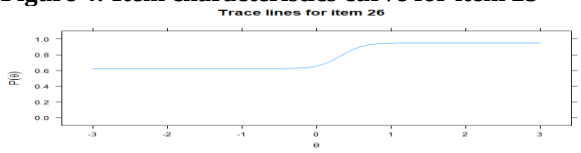

Figure 5: Item characteristics curve for item 26

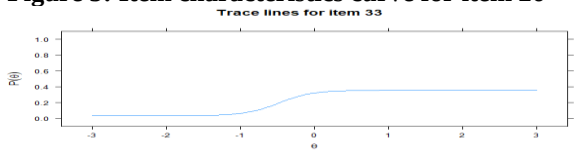

Figure 6: Item characteristics curve for item 33

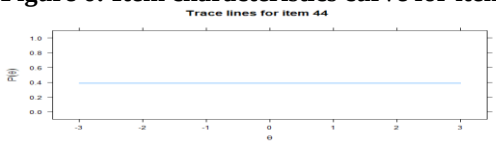

Figure 7: Item characteristics curve for item 44

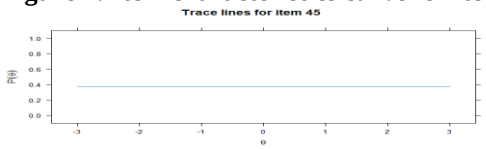

Figure 8: Item characteristics curve for item 45

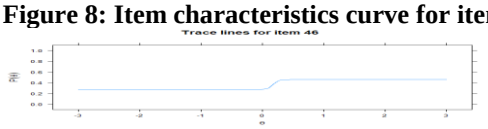

Figure 9: Item characteristics curve for item 46

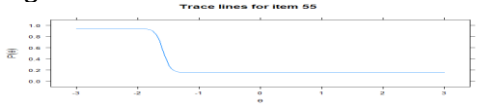

Figure 10: Item characteristics curve for item 55

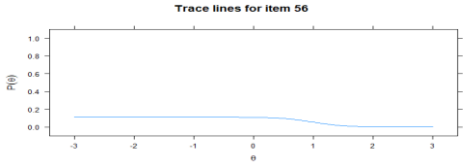

Figure 11: Item characteristics curve for item 56

ISSN 2705-2559
E-ISSN 2705-2567

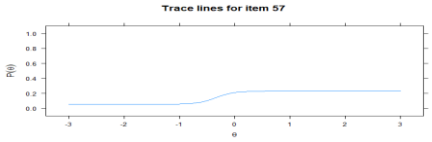

Figure 12: Item characteristics curve for item 57

Figure 1 to Figure 12 showed item characteristics curves of 12 items of the 2016 SSCE NECO Mathematics multiple choice test that failed to satisfy De Mars (2010) assertion. On each of the item characteristics curves, there is a line. The line represents the probability of examinees responding correctly to an item as a function of the latent trait denoted by θ underlying performance on the item. According to De Mars (2010), the line should be a monotonically increasing function which is in form of forward S shape. The result presented in Figure 1 to 12 showed that 12 items (1, 8, 9, 23, 26, 33, 44, 45, 46, 55, 56, and 57) do not produce a monotonically increasing function in the form of forward S shape for the probability of correct response recommended for good item under IRT. This implies that12 of the 2016 SSCE NECO 60 Mathematics multiple choice test items were poor items.

Discussion

The result of the analysis of the research question one indicated that 2016 NECO Mathematics test was essentially unidimensional. The implication of the result is that the 2016 NECO Mathematics multiple choice test was that one dominant ability accounted for the variations observed in the performance of the candidates that sat for 2016 NECO Mathematics multiple choice test. This finding was similar to finding Ojerinde and Ifewulu (2012) regarding the item unidimensionality, using 2010 of unified tertiary matriculation examination Mathematics that the 2010 UTME Mathematics complied with the assumption of unidimensionality. This finding also agrees with the findings of Bejar (1980) who observed that a single dimension of an item is sufficient enough to explain how test takers perform on such item. For mock Mathematics, this dimension is, thus, intelligence. Furthermore, the finding from this study also corroborated the findings of Orim (2015) who investigated the unidimensionality of WAEC and NECO Biology items and found unidimensionality in 47 out 50 items for WAEC and 39 out 50 for NECO with the aid of chi-square analysis. He further assumed strongly that WAEC, given its extent of compliance, is a better measuring instrument than NECO test in terms of underlying assumption of unidimensionality.

Another result of the study showed that nine items depended on other items in the 60-item test. The implication of the result is that one of the item pairs violated the item local independence (i.., items 9, 19, 22, 31, 28, 33, 44, 45 and 58 violated item local assumption). The result is in tandem Lee's (2004) findings which showed local dependence among itemswithin passages in a test of English as aforeign language. The results furthermore support, to some extent, the finding of Cai (2010) who performed analysis on a subset of PISA 2000 data (Booklet 8) using a random effect model, that there was the residual dependence between items within units.

Conclusion

Arising from the study, it can be concluded that the 2016 NECO Mathematics tests conform with the unidimensional and item characteristic curve assumption of item response theory.

Recommendations

1. There are items that did not conform with the local independence assumption of IRT.
2. Items local independence should always be ensured at the development stage

References

AKAIKE, H. (1973). *A NEW LOOK AT THE STATISTICAL MODEL IDENTIFICATION* [IEEE TRANSACTIONS ON AUTOMATED CONTROL]. (PP. 716-723).

BAKER, F. B. (2001). *THE BASICS OF ITEM RESPONSE THEORY* (2ND ED.). UNITED STATES OF AMERICA: ERIC CLEARING HOUSE ON ASSESSMENT AND EVALUATION.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75, 581–612.

Enders, C. K., &Tofighi, D. (2008). The impact of miss-specifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15*, 75–95.

FEDERAL REPUBLIC OF NIGERIA. (2013). *NATIONAL POLICY ON EDUCATION*. LAGOS: NERDC PRESS.

ISSN 2705-2559
E-ISSN 2705-2567

FINCH. W. H., & FRENCH, B. F. (2014). THE IMPACT OF GROUP PSEUDO-GUESSING PARAMETER DIFFERENCES ON THE DETECTION OF UNIFORM AND NON-UNIFORM DIF. *PSYCHOLOGICAL TEST AND ASSESSMENT MODELING, 56*(3), 25-44.

Gegbe B. & Koroma. J., (2014). Students and Teachers' Perception of the Causes of Poor Academic Performance in General and Further Mathematics in Sierra Leone: A Case Study of BO District Southern Province. *Int J Eng Res Gen Sci.* 2(5):240–253.

Grotjahn, R. (Ed.). (2010). *Der C-Test: Beiträgeaus der aktuellenForschung/The C-Test: Contributions from current research.* Frankfurt am Main: Lang.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park CA: Sage Publication.

JANG, E. E., & ROUSSOS, L. (2007). AN INVESTIGATION INTO THE DIMENSIONALITY OF TOEFL USING CONDITIONAL COVARIANCE-BASED NONPARAMETRIC APPROACH. *JOURNAL OF EDUCATIONAL MEASUREMENT, 44*(1), 1-21.

LEE, K. (2004). PSYCHOMETRIC PROPERTIES OF THE HEXACO PERSONALITY INVENTORY. *MULTIVARIATE BEHAVIOURAL RESEARCH, 39*(2), 329–358.

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

LORD, F. M., & NOVICK, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Momanyi, J. M., Too, J. & Simiyu, C. (2015). Effect of Students' Age on Academic Motivation and Academic Performance among High School Students in Kenya. *Asian Journal of Education and e-Learning* 3(5), 337 – 342

Odili, G. A. and Vincent A. Asuru V. A. (2011). Impact evaluation of Further Mathematics curriculum in Nigeria. Educational Research and Reviews 6(20), 997-1004. Available online at http://www.academicjournals.org/ERR DOI: 10.5897/ERR10.158.

OGUOMA, C. C., METIBEMU, M. A., & OKOYE, R. (2016). AN ASSESSMENT OF THE DIMENSIONALITY OF 2014 WEST AFRICAN SECONDARY SCHOOL CERTIFICATE EXAMINATION MATHEMATICS OBJECTIVE TEST SCORES IN IMO STATE, NIGERIA. *AFRICAN JOURNAL OF THEORY AND PRACTICE OF EDUCATIONAL ASSESSMENT, 4*(2), 43-45.

OJERINDE, D., &IFEWULU, B. C. (2012). *ITEM UNIDIMENSIONALITY USING 2010 UNIFIED TERTIARY MATRICULATION EXAMINATION MATHEMATICS PRE-TEST* [PAPER PRESENTATION]. 2012 INTERNATIONAL CONFERENCE OF IAEA, KAZASTAN.

OJERINDE, D., POPOOLA, O., &OYENNEHO, P. (2012). *INTRODUCTION TO ITEM RESPONSE THEORY: PARAMETER MODELS, ESTIMATION AND APPLICATION*. GOSHEN PRINT MEDIA LTD.

Ojerinde, A. (2013). Classical Test Theory (CTT) VS Item Response Theory (IRT): An evaluation of the comparability of item analysis results. A guest lecture presented at the Institute of Education, University of Ibadan on 23rd May.

Orim, S. N. (2015). *Unidimensionality and classical test parameters of 2007 WAEC and NECO biology tests in Cross River State, Nigeria* (Unpublished master's thesis). University of Calabar, Calabar, Nigeria.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

RIJN, P. W., SINHARAY S., HABERMAN S. J., & JOHNSON M. S. (2016). ASSESSMENT OF FIT OF ITEM RESPONSE THEORY MODELS USED IN LARGE-SCALE EDUCATIONAL SURVEY ASSESSMENTS. *LARGE SCALE ASSESSMENTS IN EDUCATION, 4*(10).

ROBITZSCH, A. (2019). *SIRT: SUPPLEMENTARY ITEM RESPONSE THEORY MODELS, R-PACKAGE, VERSION 3.3-26*. RETRIEVED FROM HTTPS://CRAN.R-PROJECT.ORG/PACKAGE=SIRT.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrical, 11*(52), 589-617.

Wang, W. C., Cheng, Y. Y., & Wilson, M. R. (2005). Local item dependence for items across tests connected by common stimuli. Educational and Psychological Measurement, 65(1), 5-27.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2003). *Effect of local independence on the validation of IRT item, test and ability statistic.* Amherst: University of Massachusetts.

ZHANG, J. (2007). CONDITIONAL COVARIANCE THEORY AND DETECT FOR POLYTOMOUS ITEMS. *PSYCHOMETRIKA, 72*(1), 69-91. doi:10.1007/s11336-004-1257-7

Momanyi, J. M., Too, J.& Simiyu, C. (2015). Effect of Students' Age onAcademic Motivation and Academic Performance among High SchoolStudents in Kenya. *Asian Journal of Education and e-Learning 3*(5), 337 – 342