ISSN 2705-2559
E-ISSN 2705-2567

ASSESSING THE INTERNAL STRUCTURE OF MATHEMATICS ACHIEVEMENT TEST FOR JUNIOR SECONDARY SCHOOL STUDENTS IN OSUN STATE

BY

Taiwo O. Ajeigbe (PhD): Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Osun State Nigeria; E-mail: taiaje215@gmail.com

&

Ogunsakin, I. B.: Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Osun State Nigeria; E-mail: sakinbamikole@yahoo.com

Abstract

The study assessed the internal structure of mathematics achievement test*. A descriptive survey research was adopted for the study. The population consisted of all Junior Secondary School 3 (JSS 3) Students in Osun state. The purposive sampling technique was used for the study because of the homogeneity characteristics of interest to the researcher. Hence, Osun central senatorial district was selected for the study. 5 Local Government areas were randomly selected from the 10 Local Government Areas in Osun central senatorial district, from each of the 5 selected Local Government Areas, 2 schools were selected from each of the Local Government area making total of 10 Junior Secondary School in Osun state. The sample size comprised 600 students, from each of the 10 schools, an intact class was used. 60 students OMR sheet were randomly selected from the intact class, from the selected schools given total of 600 students. The instrument for this study was (2012) multiple – choice in mathematics of Basic Education Certificate Examination for Junior Secondary School in Osun State titled Mathematics Achievement Test (MAT). The data collected were analyzed using Stout's Test of Essential Unidimensinality (STEU) and Correlation Residual with Yen Q_3 Statistics. The result showed that more than one dimension was accounted for based on the variation observed in examinees responses to the test items, it further revealed that 48 (80%) out of 60 items of 2012 Multiple – Choice in Mathematics of Basic Education Certificate Examination in Osun State violated the assumption of item Local Independence. The study recommended that for quality items to be guaranteed, the test items should be subjected to unidimensional and local independence assumptions of IRT frame work.*

Keywords: Internal structure, Unidimensionality, Local independent and Mathematics achievement tests

Introduction

Internal structure of test items has to do with the latent trait of a test item. The term latent is used to emphasize that discrete item responses are manifestations of hypothesized traits, constructs, or attributes, not directly observed, but must be inferred from the manifest responses. However, test dimensionality can be described as the number of traits underlying a test that accounts for variation in students' test performance. The student's performance, in an extremely set of n-dimensional latent space are often described by a vector of ability scores as (θ1, θ2, θ3…. θn). Item response models that presume many latent traits which accounts for students' test performance is known as multidimensional item response model. Unidimensional IRT models rely on a strong assumption that each test item is designed to measure some facet of the same underlying ability or so-called unified latent trait. It is necessary that a test intending to measure one certain trait should not be affected by other traits, especially when only the overall test scores are reported and used as an assessment criterion for various ability levels (Sheng and Wikle 2007). Most test score interpretations are based on the idea that a single test score indicates a particular hidden psychological feature. Nonetheless, even in tests designed to measure a single concept or feature, unidimensionality is frequently questioned.

In the field of educational measurement all over the world, had in the last four decades experienced changes to satisfy expanding needs for valid interpretations of individual scores from educational tests. The Classical Theory (CTT) and Item Response Theory (IRT) are two extensively utilized measurement frameworks for evaluating students' test performance (IRT). IRT that considers the test data's dimensionality (uni-dimensional or multi-dimensional), local dependence, and item attributes curve, and models students' performance based on the number of visible dimensions. However, this measurement framework is distinct from Classical Test Theory (CTT) framework which assumed that scoring of students' performances can only be done within the confine of unidimensional assumption. That is, when a single trait is enough to explain the variation observed in the student's performance. In order to achieve the

ISSN 2705-2559
E-ISSN 2705-2567

analysis of dimensinality of test items, numerous procedures for the assessment of unidimensionality have been reported and examined by difference scholars (Ojerinde &Ifewulu 2012; De Champlain & Tang, 1993; Hambleton &Rovinelli, 1986; Hattie, 1984, 1985; Hattie, Krakowski, Rogers, & Swaminathan, 1996). Traditionally, the assessment of unidimensionality of dichotomous data has been implemented through factor analytic techniques: techniques based on linear factor analysis of tetrachoric correlations as stated by (Hambleton &Rovinelli, 1986; Roznowski, Tucker, & Humphreys, 1991); Normal Ogive Harmonic Robust Moment (NOHARM) II (Fraser, 1988), based on polynomial factor analysis developed by McDonald (1967) and associated procedures (De Champlain, 1992; De Champlain & Gessaroli, 1991); TESTFACT (Bock, Gibbons, & Muraki, 1988), applied by Zwick (1987); and LISCOMP (Muthén, 1984, 1987).

In educational testing, it is often important and necessary to investigate the underlying structure of a test through dimensionality assessment endeavors at the test level. Such efforts provide the researchers and practitioners with evidence regarding test validity, shed light on the relation between different domains, and help check the tenability of the pivotal assumption of unidimensionality in item response theory (IRT). Violation of this assumption can result in biased item and ability parameter estimates (Ackerman, 1989; Kirisci, Hsu, & Yu, 2001), which can further negatively affect IRT equating (e.g., Dorans& Kingston, 1985) and cause incorrect classification of examinees into different proficiency groups (e.g., Zhang, 2010). One of the cornerstones of standard IRT models is the assumption of local item independence (LII) because multiple items are connected together to a common passage, items within a unit are not likely to be conditionally independent, which means that the independence assumption might be violated. Local independence as an assumption of IRT is an element to be considered when dealing with the internal structure of test items apart from the dimensions of the items, local independence reveals the relationships between test items in an examination question

The items that are put to Rasch analysis are required to be independent of each other. That is, a correct or wrong reply to one item should not lead to a correct or wrong reply to another item. This means that there should not be any correlation between two items after the effect of the underlying trait is conditioned out, i.e., the correlation of residuals should be zero. The items should only be correlated through the latent trait that the test is measuring (Lord and Novick, 1968). If there are significant correlations among the items after the contribution of the latent trait is removed, i.e., among the residuals, then the items are locally dependent or there is a subsidiary dimension in the measurement which is not accounted for by the main Rasch dimension (Lee, 2004). In other words, performance on the items depends to some extent on a trait other than the Rasch dimension which is a violation of the assumptions of local independence and unidimensionality. If the assumption of local item independence is violated, any statistical analysis based on it would be misleading. Specifically, estimates of the latent variables and item parameters will generally be biased because of model misspecification, which in turn leads to incorrect decisions on subsequent statistical analysis, such as testing group differences and correlations between latent variables.

The standard unidimensional IRT model requires LII (Embretson&Reise, 2000; Lord & Novick, 1968). This show that there is an interplay between unidimensional and LII, in such models, the probabilities that an examinee will provide a specific response to an item are a function of two components: 1. The test-taker's location on q, that is, his or her ability; and 2. One or more parameters (difficulty parameter, discrimination parameter, and guessing parameter) describing the relationship of the item to q. For instance, according to the one-parameter Rasch model (Rasch, 1960), the probability that a person i will successfully answer an item j, given the person's ability, qi , and the item's difficulty, d, is equal to: qi , and the item's difficulty, d, is equal to

$$P(X_{ij} = 1 \mid \theta_i \; \delta_j) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)} \quad \text{……………………..………Equation 1.0}$$

Because the likelihood of success depends only on the person's ability and on item characteristics, this means that the response to any item is unrelated to any other item given the latent trait q.

In other words, the unidimensionality assumption means that although the items may be highly intercorrelated in the test as a whole, this situation is a function that rests solely on the ability of the test-takers. When the trait level is controlled, local independence implies that no relationship remains between the items (Embretson&Reise, 2000) If two items are locally independent, then success or failure on one

ISSN 2705-2559
E-ISSN 2705-2567

item does not affect the probability of succeeding on the other item, given ability. Mathematically, if item j_1 and item j_2 are locally independent, then:
$P(X_{ij1} = X_1 \; and \; X_{ij2} = X_2 \mid \theta_i) = P(X_{ij1} = X_1 \mid \theta_i) \, P(X_{ij2} = X_2 \mid \theta_i)$ …. Equation 2.0
where $X_1 \; and \; X_2$ are equal to 0 or 1. Given the trait level, θ the conditional probability of achieving any pattern of scores on independent items is the product of the probabilities for the distinct items. The violation of the LII assumption can have substantial consequences on test parameter estimates and on proficiency estimates. Research studies show that statistical analysis of data with LID is misleading (Chen & Thissen, 1997; Chen & Wang, 2007; Junker, 1991). Tuerlinckx and De Boeck (2001) mathematically and empirically demonstrated the impact of LID on difficulty and discrimination item parameters. They showed that if negative LID is not modelled, the discrimination parameters of the interdependent items are underestimated. They also showed that the discrimination parameter (α_j) depends on the difficulty of the item it interacts with, but not on the difficulty of the item itself. Due to its effect on the discrimination parameter, **the** negative LID deflates the item information (as a function of the square of α_j) and the standard error of measurement is underestimated.

Additionally, other causes of LID cited by Yen (1993) relate to the content of items, namely, item chaining (items organized in steps) and explanation arising out of previous answer and stimulus dependence. This stimulus-LID can be produced by an examinee’s unusual level of interest in or background knowledge about the common stimuli or by the fact that information used to answer different items is interrelated in the stimulus. Chen and Thissen (1997) define this category of dependence as “underlying local dependence” because it assumes a separate trait common to each set of locally dependent items. These separate traits can therefore be regarded as minor dimensions existing beside the unique essential latent dimension q. In order to generate precise and accurate parameter estimates.IRT assumptions must be satisfied. Local item independence (LII) is a key assumption because it is directly related to the parameter estimation process. To be specific, the formation of the likelihood function, which is used to find the most likely estimates of item and ability parameters, relies upon the local independence assumption explicitly during the past two decades, extensive research was conducted to examine LID related problems (Yen, 1984, 1993).

The common findings from these studies are that moderate- to high-level LID may lead to an overestimation of the slope parameter, information value, and test reliability. Facing the problems of LID, some researchers have attempted to build new models to account for LID so that it might be allowed to occur (Gibbons &Hedeker, 1992; Hoskens& De Boeck, 1997; Jannarone, 1986; Tuerlinckx& De Boeck, 1998; Wainer, Bradlow, & Wang 2007; Wilson & Adams, 1995). In addition, the literature contains numerous approaches for detecting and/or modeling LID. Chen and Thissen (1997) investigated the X^2 statistic, the G^2 statistic, the standardized coefficient difference, and the standardized log odds ratio difference as potential detection markers of local dependence for pairs of items. These four statistics are typically used to look at the covariation of two-way contingency tables like the expected 1 and observed contingency tables in this case. Chen and Thissen (1997) compared the Pearson's X^2 statistic and the likelihood ratio G^2 statistic to the Q3, a pairwise measure of correlation of the residuals from IRT models. They found that the X^2 and G^2 indices are slightly less potent than Yen's Q3 for detecting local independence. Despite the abundance of LID studies, almost all of them are limited to the educational field examining standardized achievement tests. One major reason for overlooking LID problems is probably due to the fact that the dimensionality assumption and local item independence are closely related. However, any dimensionality test is a probability test and thus no guarantee can be made as to whether conclusion is absolutely right or wrong.

The conclusion of any dimensionality test is always a matter of degree rather than a yes/no statement. Since no instrument could strictly satisfy the dimensionality assumption as required in the IRT method, what we are testing is the degree to which the violation is small enough to be insensitive in the estimation process. The extent to which assessment techniques can effect students' performance is yet to adequately draw attention of researcher in Nigerian mathematics education. Without a doubt, if a test has items that measure more than one latent trait or factor, the examinees' performance will be negatively impacted if the test is scored using a measurement framework that does not support modelling tests with multiple traits. It is against this background that the Internal Structure of Mathematics Achievement Test for Junior Secondary School Students in Osun State was assessed**.** Hence this study.

ISSN 2705-2559
E-ISSN 2705-2567

Research Questions

(i). To what degree did the 2012 Mathematics Objective Test of Osun State Junior Secondary Certificate Examination obey assumption of unidmensinsality under the IRT frame work?

(ii.) To what degree did the 2012 Mathematics Objective Test of Osun State Junior Secondary Certificate Examination obey assumption of local independence under the IRT frame work?

Methodology

The research design for this study employed survey method. The population consisted of all Junior Secondary School 3 (JSS 3) Students in Osun state. Osun State has three senatorial zones namely: Osun Central, Osun East and West. The purposive sampling technique was used for the study because of the homogeneity characteristics of interest to the researcher. Hence, Osun central senatorial district was selected for the study, in Osun central senatorial districts, 5 Local Government areas were randomly selected from the 10 Local Government Area in Osun central senatorial district, from each of the 5 selected Local Government Area, 2 schools were selected from each of the Local Government area making total of 10 Junior Secondary School in Osun State. The sample size comprised 600 students, from each of the 10 schools, an intact class was used. 60 students OMR sheet were randomly selected from the intact class, from the selected schools given total of 600 students. The instrument for this study was (2012) Multiple – Choice in Mathematics of Basic Education Certificate Examination for Junior Secondary School in Osun State titled Mathematics Achievement Test (MAT). The data were analysed using stout's test of essential unidimensinality (STEU), tetrachoric correlation coefficient among pairs of item and full information of item factor analysis with mirt package of R Core team (Chalmers, 2012), the items with correlation residual with Yen Q_3Statistics with correlation residual values equal or greater than (≤ 0.299) were considered locally dependent.

Results

Research Question 1

(i). To what degree did the 2012 Mathematics Objective Test of Osun State Junior Secondary Certificate Examination obey assumption of unidmensinsality under the IRT frame work? To analyse for dimensionality, the responses of examinees from 2012 mathematics objective test of Osun state Junior Secondary Certificate Examination items were subjected to Stoust's Test of Essential Unidimensionality(STEU implemented in DIMTEST 2.0 package) (Stout, 2005). This was done by separating the test in to two subtests, the Assessment Subtest (AT) and the Partitioning test (PT). The AT are the items chosen as those that measure best along adominant trait. They are chosen so that they measure best in the direction most opposite to that of the PT items. The Assessment Subtest (AT), was selected empirically, using the HCA/CCPROX cluster procedure and DETECT statistic in DIMTEST, and this item cluster was tested to see if it was dimensionally distinct from the remainder of the test. A random sample of 30% of the examinees responses was used to select the Assessment Subtest, and the remaining 70% of the examinees responses (PT) was used for the dimensionality test.

Table 1: Unidimensionality of 2012 multiple – choice in mathematics of basic education certificate examination under were analysed using stout's test of essential unidimensionallity (STEU)

TL	TGbar	t	p-value
11.7900	8.5674	3.2066	0.0007

It can be seen from the Table 1.0 that the AT were dimensionally distinct from the remaining items of the test (t = 3.2066, p-value = 0.0007, one-tailed); therefore, the assumption of unidimensionality was rejected. This result shows that more than one dimension accounted for the variation observed in examinees responses to the test items. Hence, the 2012 mathematics objective test of Osun state Junior Secondary Certificate Examination items violated unidimensionality assumption.

Research Question 2

(ii.) To what degree did the 2012 Mathematics Objective Test of Osun State Junior Secondary Certificate Examination obey assumption of unidimensionality under the IRT frame work? To answer this research question, item local independent of 2012 Multiple – Choice in Mathematics of Basic Education Certificate Examination were estimated using the correlation residual with Yen Q_3 statistics. Table 2.0 presents an abridged Yen Q_3statistics for the test items

ISSN 2705-2559
E-ISSN 2705-2567

Item	It1	It2	Itm3	It4	It5	It6	It7	It8	It9	It10	It11	It12	It13	It14	It15	It16	It17	It18	It19	+	+	It59
It1	1.00																					
It2	-0.07	1.00																				
It3	0.18	0.13	1.00																			
It4	0.04	-0.07	-0.1	1.00																		
It5	-0.01	-0.05	-0.05	0.02	1.00																	
It6	-0.03	0.06	-0.04	-0.02	0.09	1.00																
It7	0.06	-0.01	-0.07	-0.02	0.02	**0.37**	1.00															
It8	0.02	0.06	-0.04	-0.01	0.05	0.02	-0.08	1.00														
It9	-0.02	-0.05	-0.19	0.03	0.09	**0.30**	**0.36**	-0.03	1.00													
It10	-0.03	0.01	-0.07	0	0.13	-0.16	-0.05	0.05	-0.07	1.00												
It11	-0.11	-0.03	0.00	0.02	0.15	0.11	0.11	0.03	0.12	0.06	1.00											
It12	0.02	0.07	0.03	-0.05	0.14	0.15	-0.03	**0.20**	0.05	0.13	0.13	1.00										
It13	0.14	-0.24	0.03	0.04	0.00	0.1	-0.08	0.12	-0.11	0.01	0.04	-0.01	1.00									
It14	0.06	-0.04	0.06	0	-0.05	0.02	0.04	-0.11	-0.06	-0.11	-0.03	-0.09	-0.01	1.00								
It15	-0.01	-0.09	-0.02	0.11	-0.08	-0.02	0.05	-0.08	0.11	-0.22	0.11	-0.02	-0.12	-0.05	1.00							
It16	0.06	-0.04	0.09	-0.08	-0.13	0.01	0.09	-0.3	-0.02	-0.01	-0.02	-0.11	-0.09	0.07	0.03	1.00						
It17	-0.18	-0.01	-0.15	0.14	-0.08	0.04	0.02	-0.01	0.13	-0.1	0.11	-0.08	0.15	-0.04	0.21	0.16	1.00					
It18	-0.03	0.02	0.05	-0.17	-0.09	-0.08	-0.12	**0.28**	-0.04	0.06	-0.26	0.11	-0.05	-0.22	-0.09	0.02	-0.18	1.00				
It19	-0.07	-0.16	0.03	-0.11	0.03	-0.01	0.01	-0.01	-0.04	-0.11	0.11	-0.08	0.03	-0.08	-0.01	0.02	0.06	-0.09	1.00			
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
It59	+	0.03	0.09	-0.14	0.06	0.02	0.5	0.08	0.07	0.09	0.1	0.13	0.15	0.06	0.11	-0.03	-0.15	0.24	-0.12	+ +	+ +	1.00

Table 2: Inter-correlation matrix of the residual of 2012 multiple – choice in mathematics of basic education certificate examination

ISSN 2705-2559
E-ISSN 2705-2567

Results of analysis in Table 2.0 are correlation coefficients among items showing the extent of local independence of the items. Each item correlated perfectly with itself give value of 1.00 but correlated at different extents with others. It can be deduced that the pairs of items that were locally dependent were 48 in number representing 80 % of the total items while 12 pairs were not locally dependent representing 20% of the total items. The implication of this result is that 48 out of 60 items of 2012 multiple – choice in mathematics of basic education certificate examination in Osun State violated the assumption of item local independence.

Discussion

Internal structure of test items is considered very pertinent. This has to do with the latent trait of test items, which can be assessed using some basic item response theory assumptions such as unidimensionality, local independence, and item characteristics curve. However, in assessing the latent trait of test items via the application of IRT model, unidemsionality should be considered as the first step in assessing the viability and fitness of the test items, since it has direct relationship with other statistics emended in item analysis. It can be observed that wrong unidimensionality results estimation affects other assumptions of IRT models and vice versa. It is thus argued that given the role of dimensionality assessment in supporting a variety of psychometric endeavors, assessing dimensionality should be a prerequisite to applying most commonly used item response theory (IRT) models in social sciences (e.g., Childs & Oppler, 2000; Jang & Roussos, 2007; Seraphine, 2000). The results of this study showed that more than one dimension were accounted for based on the variation observed in examinees responses to the test items. Hence, the 2012 mathematics objective test of Osun state Junior Secondary Certificate Examination items violated unidimensionality assumption. This is in tandem with Oguoma, Metibemu and Okoye (2016) who assessed the dimensinality of 2014 west African examination council mathematics objective test in Imo state. They found that test is multidimensional.

Furthermore, the results from this finding showed that 48 (80%) out of 60 items of 2012 multiple – choice in mathematics of basic education certificate examination in Osun State violated the assumption of item local independence. This result is in consonance with the result of AladeAletan and Sekonu (2020) who assessed the dimensionality and local independence of WASSCE 2018 mathematics objectives test scores in Lagos State, their study revealed that the mathematics objective items did not meet the assumption of local independence because their correlation residual are greater than 0.2. Moreover, this result contradicts the finding of Ubi and Abang (2011) on item local independence on selection examination in Nigeria using Yen Q_3 statistics. They discovered that UME Mathematics item for 2000, 2001, 2002 and 2003 years are locally independent.

Conclusion

Based on the study findings, it is concluded that the 2012 mathematics objective test of Osun state Junior Secondary Certificate Examination items revealed an infringement of the unidimensionality and Local independence assumption implicit in IRT.

Recommendations

1. Prior to adopting a measuring framework for estimating test results, the assumption of unidimensionalityshould be tested, and the number of dimensions should be evaluated.
2. To guarantee that quality items are selected, the test items should be subjected to item analysis based on IRT (item that are unidimensional and locally independence)
3. The Osun State Ministry of Education should form a committee of psychometricians to oversee the test development process in order to ensure that test items for public examinations in the state are of high quality.

References

Alade, O. M., Aletan, S., & Sokenu, B. S. (2020) Assessing the dimensionality and local independence of WASSCE 2018 Mathematics objective tests scores in Lagos State, Nigeria. *AJB-SDR2(1)* 8-16

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied psychological measurement*, *12*(3), 261-280.

ISSN 2705-2559
E-ISSN 2705-2567

Chen, C. T., & Wang, W. C. (2007). *Effects of ignoring item interaction on item parameter estimation* and detection of interacting items. Applied Psychological Measurement, 31, 388–410.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289.

Chids, R. A. & Oppler, S. H. (2000). Implications of test dimensionality for unidimensional IRT scoring: an investigation of a hight-stakes testing program. *Educational and Psychological Measurement*, *60* (6), 939-955

Comrey, A. L. & Lee, H. B. (1992). A first course in, factor analysis (second edition). Hillsdale, NJ: Lawrence Erlbaum.

De Champlain, A. F. (1992). *Assessing test dimensionality using two approximate chi-square statistics*. University of Ottawa (Canada).

De Champlain, A. F. (1996). The effect of multidimensionality on IRT true-score equating for subgroups of examinees. *Journal of Educational Measurement*, *33*(2), 181-201.

De Champlain, A., & Tang, K. L. (1993). The effect of nonnormal ability distributions on the assessment of dimensionality. In *annual meeting of the National Council on Measurement in Education, Atlanta, GA*

Dorans, N. J., & Kingston, N. M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE verbal scale. *Journal of Educational Measurement*, *22*(4), 249-262.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.
Erlbaum

Fraser, C. (1988). NOHARM II: A Fortran program for fitting unidimensional and multidimensional normal ogive models of latent trait theory. Armidale, N.S.W.: University of New England, Centre for Behavioral Studies.

Gibbons, R. D., &Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423-436.

Hambleton, R. K., & Swaminathan, H. (1985). A Look at Psychometrics in the Netherlands.

Hambleton, R. K., &Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. *Applied psychological measurement*, *10*(3), 287-302.

Hattie, J., Krakowski, K., Jane Rogers, H., & Swaminathan, H. (1996). An assessment of Stout's index of essential unidimensionality. *Applied psychological measurement*, *20*(1), 1-14.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, 1523-1543.

Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological methods*, *2*(3), 261.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika*, *51*(3), 357-373.

Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika*, *56*(2), 255-278.

Kirisci, L., Hsu, T. C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied psychological measurement*, *25*(2), 146-162.

Lee, Y.-W. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test, Language Testing 21, 74-100). Princeton, NJ: Educational Testing Service.

Masters, G. N. (1988). Item discrimination: When more is worse. Journal of Educational Measurement, *25 (*1), 15-29

McDonald, R. D. (1967). Nonlinear factor analysis. Psychometric Monographs, No. 15, 32(4, Pt. 2)

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115-132.

Nandakumar, R., & Stout, W. (1993). Refinements of Stout’s procedure for assessing latent trait unidimensionality. *Journal of educational statistics*, *18*(1), 41-68.

ISSN 2705-2559
E-ISSN 2705-2567

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. &Ifewulu, B.C. (2012). Item unidimensionality using 2010 unified tertiary matriculation examination mathematics pre-test. International Conference of IAEA (Kazakhstan) Psychologica

R-Tate, Y. (2003) A comparison of selected empirical methods for assessing the structure of Responses to test items," *Appl. Psychol. Meas.27* (3) 159–202

Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement*, *15*(2), 109-127.

Seraphine, A. E. (2000). The performance of DIMTEST when latent trait and item difficulty distributions differ. *Applied Psychological Measurement*, *24*(1), 82-94.

Sheng, Y., & Wikle, C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, *67*(6), 899-919.

Sireci, S. G., Thissen, D., &Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, *28*(3), 237-247.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*(4), 589-617.

Stout, W. (2005). DIMTEST (Version 2.0) [Computer Software]. Champaign. IL: The William Stout Institute for Measurement

Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, *26*(3), 247-260.

Tuerlinckx, F., & De Boeck, P. (1998). Modeling local item dependencies in item response theory.

Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological methods*, *6*(2), 181.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.

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, 5–27

Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika*, *60*(2), 181-198.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. Applied Psychological Measurement, 8(2), 125-145.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of educational measurement*, *30*(3), 187-213.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement*, *24*(4), 293-308.