UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.18 (2);
July-December, 2024 https://unijerps.org
2023 Impact Factor:* 6.834;** 2022 ***Impact Factor: 5.156; ***2021 Impact Factor: 5.641;***
2020 Impact Factor: 5.355. *Indexing and Impact Factor URL: http://sjifactor.com/passport.php?id=21363*

DIMENSIONALITY AND LOCAL INDEPENDENCE OF TEST ITEMS OF MATHEMATICS ACHIEVEMENT TEST

Isaac Anayo Onyejiekwe
isaaconyejiekwe@yahoo.com
Department of Educational Foundations, Faculty of Education, Nnamdi Azikiwe University, Awka
&
Prof. Romy O. Okoye
Department of Educational Foundations, Faculty of Education, Nnamdi Azikiwe University, Awka

ABSTRACT
This study examined the dimensionality and local independence of test items of mathematics achievement test. It was carried out in Rivers State. Two research questions guided the study. Descriptive survey research design was adopted .The instrument for data collection was a 120 – item draft of the Mathematics Achievement Test (MAT), developed by the investigators. The population of the study consisted of 2022/2023 SS3 Mathematics students in the two hundred and seventy-eight (278) public secondary schools in Rivers State, Nigeria, numbering 59,223.The researchers used a samples of 1, 812 SS3 students obtained through a combination of multi stage and cluster sampling technique . The analysis of data was done using Mplus and DIMPack software while the results of the analysis of Research Question 1 was obtained using DIMTEST statistic in DIMPACK 1.0 software and YEN Q3 statistic was used to answer Research Question 2. The results of the major findings in this study were as follows: the mathematics achievement multiple choice tests were multidimensional. The study also indicated that the local independence assumption of the IRT was not grossly violated using Yen's Q3 to screen items for local dependence, 88% of item residual correlations were below absolute value of 0.2. The major recommendation was that test developers and the external examining bodies were encouraged to be open to the use of item response theory and standard statistical methods for estimating test parameters to ensure the quality of the assessment instrument. It was as well recommended that test practitioners should endeavour to perform dimensionality assessment analysis of their tests before administration; this will help them to generate quality test items that will subsequently ensure correct interpretation of test.

Keywords: Dimensionality, Local Independence, Mathematics and Achievement test

Introduction

In the classical test theory framework, achievement tests are often developed and used, on the assumption that each of them measures a single construct. This assumption may not be true all the time. When this assumption is not true, any interpretation given to scores obtained with

such instruments may be misleading. It is for this reason that experts who adopt the item response theory framework try to ascertain, during instrument development, whether or not an instrument measures one or more traits (constructs). This is often referred to as test of dimensionality, which is one of the assumptions tested during instrument development. One of the assumptions that are also used in this framework is that of item local independence.

As in other achievement tests, scores from mathematics achievement tests are often treated as measures of single traits. It is therefore necessary to ascertain the dimensionality of such tests, that is, determining whether they measure single or multi- traits .It is also necessary to ascertain whether the items are locally independent.

Dimensionality and local independence are among the two major assumptions of item response theory in which this study anchores on and both of them are the thrust of this study.

Dimensionality simply refers to the number of traits measured by an instrument. It is commonly assumed that one ability is being measured by a set of items in a test. Of course, this assumption may not be strictly met because there are always many cognitive, personality, and test-taking factors which impact on test performance, at least to some extent. These factors might include level of motivation, test anxiety, ability to work quickly, knowledge of correct use of answer sheets, and other cognitive skills in addition to the dominant trait .An instrument could be unidimensional or multidimensional .

A test is said to be uni-dimensional when it is assumed that it measures one trait (e.g.ability). According to Ziegler and Hagemann (2015), a uni-dimensional test assesses a single, underlying trait or construct. It anchors on the assumption that all items in the test have one common ground. The assumption of uni-dimensionality means that a test can only measure one trait at time .It assumes that the performance in a test can only be explained by a single latent trait. Therefore, it assumes that items in a test are expected to measure only one latent trait and that any other factors that influence item scores are treated as random errors or nuisance to the models. Unidimensionality scale can be evaluated by performing an item-level factor analysis, designed to evaluate the factor structure underlying the observed co-variation among item responses. The assumption can be examined by comparing the ratio of the first to the second Eigen value for each scaled matrix of tetra choric correlations. This ratio is an index of the strength of the first dimension of the data. Another indication of unidimensionality is that the first factor accounts for a substantial proportion of the matrix variance (Ogomaka, Onah and Amadi, 2017). For tests with many items, the assumption of unidimensionality may be unrealistic; however, Ojerinde, Popoola, Ojo, and Ariyo (2014) report that IRT models are moderately robust to departures from unidimensionality. What is required for this assumption to be met adequately by a set of test data is a “dominant” component or factor which influences test performance. This dominant component or factor is referred to as the ability measured by the

test. This is the ability on which examinees are being measured. All other contributing factors to test performance are defined as errors. Item response models in which a single ability is presumed sufficient to explain or account for examinee performance are referred to as uni-dimensional models.

If uni-dimensional assumption is violated a multidimensional model of IRT should be applied. A test is said to be multidimensional when it measures more than one construct. Multidimensional IRT (MIRT) models seek to explain an item response according to an individual's standing across multiple latent dimensions (Reckase, 2009). MIRT represents a broad class of probabilistic models designed to characterize an individual's likelihood of an item response based on item parameters and multiple latent traits. Multi-dimensional models assume that more than a single ability is necessary to account for examinee test performance.

The second assumption of item response theory is the local independence of items. This means that the probability of an examinee getting an item correct does not depend on the answers given to other items in the test. According to Ayozie (2015), there are two forms of this assumption, the strong and the weak forms. The strong form states that the test item responses of a given examinee are statistically independent. This means that an examinee's performance on one item does not affect his or her performance on the other items in the test. In effect, the test items measure single ability. Local independence is necessary for but does not guarantee dimensionality. The weak form of the assumption is met by a set of test items if pairs of items are uncorrelated for examinees of the same ability level. The distinction between the two forms is that for the strong form, items are statistically independent while for the weak form,items are uncorrelated. Local independence represents a restrictive assumption and so may not be strictly satisfied with many sets of mental test data. If the assumption of local independence is violated a testlet model of item response theory should be used for calibration of item parameters.

De Ayale (2010) states that Local item dependency is recognized as conditional independence. Ourya and Reza (2008) stated that the assumptions imply that the items in a test should not be related in any way. Hence it is the probability of an examinee rightly answering a test item that will be dependent on the testee's answers in other test items. This implies that an item should not provide a clue to answer another question rightly. Elias et.al (2006) stated that the principle of local independence requires that responses to different items are uncorrelated when θ (i.e. the person's ability) is fixed . If local independence is non-existent, a large correlation between two items can essentially affect the latent trait and hence causes lack of validity. It is important to note that, if two items are locally independent, the success or failure of the testee on one item is not affected by the probability of getting the other item right or wrong , when the ability level of the testee has been taken into consideration . Therefore, when there exist a possibility of a

test violating the assumption of local independence, it is called item dependence.

There are two ways or approaches to investigating Local independence namely: the stochastic and the statistical local independence. The test-taker's responses are not utilized to determine Local Stochastic Independence (LSI) . In line with this, Sijtsma and Molenaar (2002) ,state that the expected reaction to any given item is unconnected to the responses to other items . In yes –or- no- response item, it is assumed that one item must not be dependent on the other items. Local item independence is statistically determined using the test-takers' responses. Hence, it can be defined as a situation where each test-responses takers to multiple – choice item are independent to one another According to Ibrahim (2017) , the response to one item has no direct effect on the response to another . In agreement with this assumption, Embreston and Reise , (2000) opine that multiple – choice items on a test should not be related to one another .

Different authors and academicians have suggested a range of statistical approaches for measuring Local item independence. According to Wollenberg (1982) cited in, Fit measure Q2 is one of the methods of local item independence. Other statistical approaches for measuring local item independence include Tetrachoric correlation which is a biserial correlation, Pooled – Across Group (PAG) method, Partial Correlation, Yen Q_3, DIMTEST, etc. It is important to note that there are currently no well – documented claims that a single statistic is superior to another. Yen Q_3 was used in this study to determine the local independence of the Mathematics Achievement Test

Several studies on dimensionality and Local item independence have been carried out both within and outside Nigeria. For instance, a study was carried out by Anywale, Isaac-Oloniyo and Abayomi (2020) on assessment of dimensionality of Osun State Unified Mathematics Achievement Test. The study used causal comparative research approach. Results showed that Osun State Unified Mathematics Achievement Test items in 2020 was multidimensional in nature as the observed difference in the second eigen value of the observed data and that of second eigen value of the simulated data was statistically significant, p = 0.0099; Stout's test rejected the assumption of essential unidimensionality (t = 10.6260, $p<0.05$). Non-linear factor analysis and full information factor analysis revealed that four dimensions embedded in the test items and loadings of the items showed within – item multidimensionality.

Similarly, Oguoma, Metibemu and Okoye (2016) carried out a study on dimensionality of 2014 Mathematics achievement items of West African Senior Secondary Certificate Examination (WASSCE). To achieve this, the dimensionality of test was assessed. The study adopted causal comparative design. Results showed that the test and its items violate unidimensionality assumption (Stout's test rejected the assumption that the test was unidimensional, T = 7.3799, $P<0.001$, one tailed); BPMAT showed that the second eigen values was significantly

greater than that of the simulated data set, p = 0.0099). Furthermore, full information factor analysis showed that five dimensions underlie the test; the test items revealed within- item multidimensionality.

Alade, Aletan and Sokeum (2020) carried out a study on dimensionality and local independence of WASSCE 2018 mathematics objective test items in Lagos State in Nigeria. The study was a descriptive survey design type aimed at investigating the extent to which the 2018 West African Senior School Certificate Examination mathematics objective test items fulfilled the unidimensionality and local independence assumptions. The findings indicated that the 2018 WASSCE mathematics objective test was multidimensional. In addition, some items violated the assumption of local independence because their correlation residuals were greater than 0.2.

A study by Ubi, Umoinjang and Joshua (2012) was designed to assess the item local independence in University Matriculation Examination in Nigeria for the years 2000 to 2003. Data analysis using tertrachoric correlation analysis revealed that JAMB-UME mathematics test items were significantly locally independent. It is thus concluded that JAMB-UME mathematics tests for 2000 to 2003 met the assumptions of Item Response Theory (IRT) on local independence . Furthermore, Pila and Adewale (2020) compared dimensionality and local independence of test items of 2017 Basic Education Certificate Examinations (BECE) on National Values Education conducted by Benue State Examinations Board (BSEB) and National Examinations Council (NECO). The study adopted non-experimental design of ex-post-facto research type. Data were collected through personal visit to sampled schools. Two research questions were raised and answered using linear factor analysis and correlation residual with NOHARM and Yen Q3 software. Findings revealed that BSEB BECE 2017 National Values Education test fulfilled the assumption of unidimensionality, while two dimensions underlie NECO test, which accounted for variations in examinees' performance. Also, out of 45 items for each test, 6 items of BSEB and 12 items of NECO test violated local independence assumption (correlation residual values >0.2). Taiwo and Adebayo (2018) assessed and analyzed the dimensionality of the West African Examinations Council's Senior School Certificate 2013-2014 objective tests in Financial Accounting. Exploratory factor analysis research was adopted for the study. Data collected were analyzed using principal component analysis. The findings of this study revealed seven factors extracted from the PCA of 2013 West African Senior School Certificate Examination (WASSCE) Financial Accounting objective test with first and second factors having 37.848% and 26.306% of variance respectively. Nine factors were also extracted from the PCA of 2014 WASSCE Financial Accounting Objective test with only First factor having 28.907% of variance. Significant difference was observed between the two tests. It was concluded that the two tests were not the same in-terms of dimensionality. The study recommended that the West Africa Examinations Council (WAEC) should

intensify efforts towards improving the standard of the test items.

Results on the literature reviewed above show that most of the achievement tests in the studies were found to be multi-dimensional , while there was inconsistency regarding the results of test of local independence. The present study theregore set out to determine the dimensinality and local independence of Mathematics Achievement Test items . To that effect, the following research questions guided the study:

1.What is the dimensionality of the Mathematics Achievement Test?

2. How locally independent are the items of the Mathematics Achievement Test?

Research Method

Research Design. The design of the study was survey research. According to Nworgu as cited in Ukwuije and Adutchay (2016) , a survey research is one in which a group of people or items are studied by collecting and analyzing data from only a few people or items considered being representative of the entire group . Hence, this design was adopted in this study to ascertain the dimensionality and local independence of test items of mathematics achievement test.

Area of the study. The study was carried out in Rivers State of Nigeria.

Population of the study. The population of this study consisted of 2022/2023 SS3 mathematics students in the two hundred and seventy-eight (278) public secondary schools in Rivers State, Nigeria, numbering 59,223 (Planning , Research & Statistics (PRS) Department , Rivers State Senior Secondary Schools Board (RSSSB) , 2023). The choice of SS3 mathematics students for the study was because they had the background knowledge of the subject than other classes in the senior secondary school.

Sample and sampling techniques. The researchers used a sample of 1,812 SS 3 students. They used a combination of multi-stage and cluster sampling techniques. At the first stage, simple random sampling technique was used to obtain 8 Local Government Areas (this was 35% of the Local Government Areas in the State). In each of the obtained Local Government Areas, two schools were obtained through simple random sampling. This gave rise to 16 schools. Three arms of SS3 Students were obtained through simple random sampling from each of the schools. The entire students in the sampled arms were used.

Instrument for data collection. The instrument used for data collection was the Mathematics Achievement Test (MAT). A pool of 120 items was generated each with four response options A- E. The 120 items were spread over the content (topics) and their cognitive level as contained in the table of specifications. The initial instrument was given to two experts in the Department of Educational Foundations, Nnamdi Azikiwe University, Awka and one senior secondary school mathematics teacher. These experts were used to establish content validity of the instrument. At the end of the validation by experts, corrections were made, few questions were modified. None of the items was dropped, as all of them scaled through.

Method of Data Collection. The instrument was administered on a sample of 1,812 SS3 students. The administration was done through the help of mathematics teachers of the sampled classes.

Method of data analysis. The analysis of scores obtained from the instrument was done using Mplus and DIMPack software. The results of the analysis were presented according to the research questions that guided the study.

Research Question 1 was answered using DIMTEST statistics in DIMPACK 1.0 software. A p-value of DIMTEST statistic less than .05 level of significance level indicates unidimensionality (Reckase, 2009). Furthermore, if the difference between the number of items in Partitioning Subtest (PT) and the Assessment Subtest (AT) in a test is significant, there is evidence of multidimensionality (Anywale, Isaac-Oloniyo and Abayomi, 2020). Research Question 2 which is on local independence was answered using Yen Q3 statistics. A Yen Q3 statistic above . 2 indicates violation of local independence (Balazs and De Boeck, 2006).

Results and Discussion

Research Question 1: What is the dimensionality of the items of the mathematics achievement test**?**

The dimensionality assumption was investigated using DIMTEST statistics. Results obtained were presented in table 1 below**:**

Table 1: *Dimtest Statistics of Mathematics Achievement Test Items*

TL	TGbar	T	AT	PT	P-value
17.2871	3.9316	15.7880	38	82	0.0021

The result in Table 1 indicates that the mathematics achievement test is multidimensional since p-value is less than .05 level of significance. For this particular study, the difference between the AT and PT as shown in Table 1 is significant ($T=15.7880$, $p<0.05$). This led to the conclusion that the AT items were dimensionally distinct from the remaining items in PT. Therefore, multidimensionality is manifest in mathematics achievement test items. This implies that different aspects of latent trait in mathematics were adequately covered by the instrument.

The study indicated that the underlying latent ability of examinees responses to the instrument is multidimensional. The finding agreed with the study of Anywale, Isaac-Oloniyo and Abayomi (2020) on assessment of dimensionality of Osun State Unified Mathematics Achievement Test items. Findings from the work of Oguoma, Metibemu and Okoye (2016) on dimensionality assumption test on 2014 Mathematics achievement items of West African Senior School Certificate Examination (WASSCE) showed that the test items of WASSCE mathematics were inherently multidimensional in nature. Furthermore, Okwilagwe and Ogunrinde (2017) also found that fifty (50) items of 2013 WASSCE and sixty (60) items of National Examinations Council (NECO) Geography respectively violated assumption

of unidimensionality and that there were more than one dimension that accounted for the variation observed in examinees to the geography test items. The finding in the present study is however in contradiction with that of Esomonu and Eleje (2017) who found that economics quantitative diagnostic test for secondary school students based on IRT was unidimensional. Similarly, Latun (2011) also found physics diagnostic test for secondary school students based on IRT to be unidimensional.

Question 2. How locally independent are the items of the mathematics achievement?

The assumption of local independence of test items was investigated with the use of Yen's Q3 statistics. Based on this statistics, residuals for any pair of items should be uncorrelated, and generally close to zero. Residual correlations that are high indicate a violation of the local independence assumption, and this suggests that the pair of items have something more in common than the rest of the item set have in common with each other (Marais, 2013). For this study, using Yen's Q3 to screen items for local dependence, 88% of item residual correlations were below absolute value of 0.2. This indicates that the local independence assumption of the IRT was not grossly violated. The implication of the items being locally independent is that the item responses of the individual students do not affect the responses of any other students when the actual test was being administered. This further implies that within any given group of examinees (students) all characterized by same values, θ …..θ, the (conditional) distribution of the item scores of the mathematics achievement test are all independent of each other.

Conclusion

On the basis of these research findings, with the use of IRT framework, it was concluded that mathematics achievement test is multidimensional and the items are locally independent.

Recommendations

Based on the findings of this study, the following recommendations were:

1. Test developers and the external examining bodies should be encouraged to be open to the use of item response theory and standard statistical methods for estimating test parameters to ensure the quality of the assessment instrument.
2. Test practitioners should endeavour to perform dimensionality assessment analysis of their tests before administration; this will help them to generate quality test items that will subsequently ensure correct interpretation of test.

REFERENCES

Alade, O. M., Aletan, S. & Sokenu, B. S (2020). Assessing the dimensionality and Local independence of WASSCE 2018 mathematics objective test scores.*The African Journal of Behavioral and Scale development Research*, 2(1), 8-16.

Anywale, M. A., Isaac-Oloniyo, M. & Abayomi, F. R. (2020). Dimensionality assessment of binary response test items : A non-parametric approach of Bayesian item response theory measurement. *International Journal of Evaluation and Research in Education,* 9(2), 385-393.

Ayozie, C. N. (2015). *Application of IRT in development and standardization of chemistry achievement test for students.* Unpublished Masters Thesis, University of Nigeria Nsukka.

Balazs, D. M. & De Boeck, V.F.(2006). Estimation of compensatory and non-compensatory multidimensional item response models using Markov chain Montecarlo. A*pplied Psychological Measurement,* 27, 395-399.

De Ayala, R. J. (2010**).** *The theory and practice of item response theory. Journal of Educational Measurement 47(4)* .Available at https://onlinelibrary.wiley.com.https:/doi: 10.1111/j.1745-3984.2010.

Esomonu,N.P. M & Eleje,L.I.(2017).Diagnostic quantitative economics skill test for secondary schools: Development and validation using item response theory. *Journal of Education and Practice,8*(22),110-125.

Ibrahim, A. (2017). The Influence of item Parameters , Distracter functioning and Correct Response location of the Difficulty level of Multiple- Choice Questions . *Journal of Educational Studies , Institute of Education , university of Jos , Jos , Plateau state, Nigeria.* 17(1 **)** 490-**5**04. https://www. researchgate.com .net/*p*ublication/322042556

Latun , O.S. (2011). Develo*p*ment and validation of Diagnostic test in *p*hysics for secondary school students in Lim*popo P*rovince of sout Africa using item res*ponse theory . Unpublished M.ed Thesis Universsity of Pretora south Africa.*

Marais , I.(2013) . Local Dependence. In K.B. Christensen , S.,Kreiner & M. Mesbah (Eds), *Rasch models in health* .: JohnWiley & Sons Inc.

Ojerinde, D., Popoola,K.,Ojo, F. and Ariyo, A.(2014). *Practical applications of item response theory in large scale assessment.* Abeokuta: Marvelous Mike Press.

Ogomaka, M.C., Onah, F. E. & Amadi, N.C.(2017).The comparison of the development of chemistry achievement test using item response theory and classical test. *Nigerian Journal of Educational Research and Evaluation,16*(2),41-47.

Oguoma, C. C., Metibemu, M.A. & Okoye, R.O. (2016). An assessment of the dimensionality of 2014 West African secondary school examination mathematics objective test scores in Imo State, Nigeria, *African J. Theory Pract. Educ. Assessment*, 4, 18-33.

Okwilagwe, M. A. & Ogunrinde, E.A. (2017). Assessment of unidimensionality and local independence of WAEC and NECO

2013 Geography Achievement Tests. *African Journal. Theory Pract. Educ. Assess.*5,31-44.

Pila, N & Adewale, J. G. (2020).Assessment of dimensionality and local independence of state and national basic education certificate examination in national values education in Benue State.*The African Journ al of Behavioral and Scale Development Research*, 2(1), 27-38.

Popoola, K., Ojo, F. & Ariyo, A.(2014).*Practical applications of item response theory in large scale assessment.* Marvelous Mike Press.

Pourya, B.M. & Reza, P.(2008) . The effects of violation of local independence on the Pearson measures under Rasch model . *International Journal of Applied Logistics, 11(1) , 61-67*

Reckase, M.D.(2009).*Multidimensional item response theory*. London: Springer.

Sijtsma , K., & Molenaar, I.W. (2002). Introduction to non parametric item response theory. (5th ed.). SAGE Publications Inc.

Taiwo, B. Y. & Adebayo, A. D. (2018). Dimensionality of West African Senior School Certificate Examination June/July 2013/2014 objective tests in financial accounting. *Federal University Dutsin-Ma Journal of Educational Research, 3(1), 69-77.*

Taiwo, A. R. & Oyebola, A. S. (2016). Development and validation of mathematics aptitude test using Rasch model IRT. *International Journal of Education and Science Research,* 16(2), 2-15.

Ubi , I.O.,Umoinyang, I.E & Joshua, M.T (2011). Item local independence in selection examination in Nigeria . Education for Today, 7, (1) .

Ukwuije, R.P. J & Adutchay, V.O. (2016) . *Introductory research and methods and statistics in education.* Port Harcourt : Chadik Printing Press.

Wollenberg , A.L. (1982). Two new test statistic for Rasch model. Psychomerika , 47, 123 – 139.

Ziegler, M. & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfals and loopholes. *European Journal of Psychological Assessment 31(4*):*231-237*