

## **DIMENSIONALITY AND MODEL FIT ASSESSMENT OF 2022 NECO MATHEMATICS MULTIPLE-CHOICE QUESTIONS IN ANAMBRA STATE.**

**By**

**SOMTOO VICTOR NDUKA**  
*Email: ndukavictor6@gmail.com*  
*Contact: 07034321151*

**and**

**IKEANUMBA CHUKWUEMEKA BENEDICT**  
*Department of Educational Foundations, Faculty of Education*  
*NnamdiAzikiwe University, Awka, Anambra State*  
*Email: chukwuemekaikeanumba@gmail.com*  
*Contact: 07039370884*

### **Abstract**

Mathematics multiple-choice questions (MCQs) play an indispensable role in assessing student's knowledge level in the different aspects of mathematics. These test questions, to be valid and reliable, should have an infinitesimal variance between the scores it obtains and the student's ability. This study evaluates the reliability and validity of 2022 NECO mathematics multiple-choice questions by assessing the dimensionality and model fit of the test items. Two research questions guided the study. The study adopted a survey design, and the population for the study was made up of all the 26,421 SSIII students of the 2022/2023 academic session who enrolled for the NECO examinations in Anambra State, out of which a sample of 1,828 was selected by the use of purposive sampling technique. Data collected was analyzed using DIMTEST statistics for research question one, and Akaike Information Criterion (AIC) and Likelihood Ratio Test (LogLik) for research question two. The findings revealed that the 2022 NECO Mathematics MCQs do not fulfill the Item Response Theory assumption of Unidimensionality, while the data was best fitted into 4 Parameter Logistic Model (4PLM). It was recommended that examination bodies, and researchers involved in the use of IRT should ascertain the conformity of the test items to the IRT assumptions in order to avoid bias in item parameter estimation.

**Key word:** Item Response Theory, Dimensionality, Model Fit, NECO, Mathematics

## **Introduction**

Education plays an indispensable role in shaping individuals, building human capital, fostering economic growth, promoting social cohesion, and driving overall progress within a nation. This is evidently manifest in the Federal Government of Nigeria (2013) declaration in her National Policy on Education that education is an instrument "par excellence" for achieving national development.

In today's world, STEM (Science, Technology, Engineering and Mathematics) education is of utmost importance, as it provides the foundation on which the knowledge and critical thinking skills necessary for innovation, research, development, and problem-solving in various domains are acquired. Mathematics is an important aspect of STEM education. It is the study of numbers, quantities, structures, and patterns. Mathematics has various branches, including algebra, geometry, calculus, statistics, and probability, and serves as a foundation for many scientific and technological advancements. In Nigeria, it is one of the compulsory subjects offered in both primary and secondary schools, and a credit level pass, at the minimum, is the prerequisite to gaining admission into any tertiary institution.

To realize the goal of the overall development of the Nigerian society, there is a need to maintain high standard and quality STEM education delivery across the nation. The resulting effect of this delivery is quantified through measurement and evaluation. Omorogiuwa (2010) defined the concept of measurement as the assigning of numbers on a student's performance to show the extent to which a trait is present or absent, according to specified rules, using various tools. One of such tools is test. Others, as identified by Okoye (2015) include observation, rating scales, interviews, among others. Ukwuije (2012) defined test as an instrument administered to the respondents in order to determine some previously identified objectives in the individual. This means that through testing, the degree to which educational objectives have been achieved can be ascertained by a teacher or an examination body. The National Examination Council (NECO) is the Nigerian examination body that carries out this responsibility at the secondary school level.

Test items should have an infinitesimal variance between the scores it obtains and the student's ability, and this can only be accomplished if

such test item is of good quality. The quality of test items is judged worthwhile when it possesses appropriate psychometric properties (reliability and validity), and this is determined through item analysis. This inherent qualities of test items can be studied and evaluated from different perspectives and theories, and in educational measurement, there are two main test theories; the Classical Test Theory (CTT) and the Item Response Theory (IRT). These theories are based on different theoretical assumptions, and use different statistical approaches.

The Classical Test Theory (CTT) has been in use over the majority of the 20<sup>th</sup> century (Demirtash, 2002). CTT is based on the true score theory which views the observed score (X) as combination of the true score (T) and an error component (E) (Adedoyin, 2010). That is,  $X = T + E$ . The true score reflects the actual knowledge of the student, but this is not always obtainable as it is usually smirched by different sources of errors. The basic assumptions of the CTT are: the error scores and true scores from the same test are uncorrelated, and the error scores have an expected mean of zero in a population of respondents. (Lord cited in Ojerinde, 2013).

The Item Response Theory (IRT), in contrast to the CTT, focuses primarily on item-level information, and describes the relationship between an examinee's test performance and the latent traits assumed to underlie such performance (Hambleton, Robin, and Xing, 2000). For items that are dichotomously scored, IRT models the probability of a correct answer using four logistics functions. These are one-parameter logistic model (item difficulty), two-parameter logistic model (item difficulty, and item discrimination), three-parameter logistic model (item difficulty, item discrimination and guessing parameter), and the four-parameter logistic model (item difficulty, item discrimination, guessing parameter and upper asymptote parameter). The crucial benefits of IRT models are realized to the degree that the data fit the different models. The analysis of model-data fit is a check on the internal validity of the test items (Obinne, 2013). The basic assumptions of the IRT are: unidimensionality and local dependence. The concept of unidimensionality of test items according to Kyung (2013) implies that the test items measure only one latent trait. This means that performance is assumed to be governed by a single factor, and the concept of local independence of test items according to Ojerinde

(2013) implies that the probability of an examinee getting an item correctly is unaffected by the answer given to other items in the test.

The basic aim of standardizing test items is to produce tests of desired quality regardless of the test theory framework. It is therefore imperative that examination bodies, including NECO, employ the best practices in test construction in order to develop quality test items to efficiently measure student's achievement. This study evaluates the reliability and validity of 2022 NECO Mathematics multiple-choice questions by assessing the dimensionality and model fit of the test items.

### **Purpose of the Study**

The purpose of the study is to ascertain the inherent characteristics of the 2022 NECO Mathematics multiple-choice questions using the framework of the Item Response Theory.

Specifically, the study did:

1. Determine the extent to which the unidimensionality assumption of the item response theory was met by the 2022 NECO Mathematics multiple-choice questions.
2. Identify the IRT model that best fits the data of the 2022 NECO Mathematics multiple-choice questions.

The following research questions were raised to guide the study:

1. To what extent was the unidimensionality assumption of the IRT met by the 2022 NECO Mathematics multiple-choice test items?
2. What IRT model best fits the data of the 2022 NECO Mathematics multiple-choice test items?

### **Method**

Survey research design was adopted for this study. Generally, according to Nworgu (2015), a survey research is one in which a group of people or items is studied by collecting and analyzing data from only a few people or items considered to be representative of the entire group. A sample of 1,828, which was selected by the use of the purposive sampling technique, out of a population of 26,421 Senior Secondary School (SS3) students who enrolled for the 2022/2023 Senior School Certificate Examination in Anambra State was used in the study. The sample size was considered a good representation of the population. The instrument

for data collection was the May/June 2022 NECO Mathematics Paper III. This instrument was administered to the students of the sampled schools by the researcher, assisted by the Mathematics teachers of the school, under exam conditions. The data obtained were analyzed for the dimensionality assumption using the Dimtest statistics. The best fit model for the calibration of the IRT item parameter was determined using the Akaike Information Criterion (AIC) and Likelihood Ratio Test (LogLik).

## **Results and Discussion**

**Research Question 1:** To what extent was the IRT unidimensionality assumption met by the 2022 NECO mathematics multiple-choice test items?

**Table 1: Dimtest Statistics of 2022 NECO mathematics multiple-choice test items**

TL	TGbar	T	P-value
17.1871	3.9326	13.1887	0.0000

The result in Table 1 indicates that 2022 NECO mathematics multiple-choice test items are multidimensional since p-value is less than 0.05 level of significance. 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 (Anyawale, Isaac-Oloniyo and Abayomi, 2020). For this particular study, the difference between the AT and PT as shown in table 1 is significant ( $T=13.1887$ ,  $p<0.05$ ). This led to the conclusion that the AT items were dimensionally distinct from the remaining items in PT. Therefore, multidimensionality is manifest in 2022 NECO mathematics multiple-choice test items. This manifest multidimensionality implies that different aspects of mathematical abilities were measured by the instrument. The finding agreed with that of Anyawale, Isaac-Oloniyo and Abayomi (2020) who tried to ascertain the dimensionality of Osun State Unified Mathematics Achievement Test items, and found that they were multidimensional in nature. Findings from the work of Oguoma, Metibemu and Okoye (2016) on dimensionality assumption test on 2014 Mathematics achievement items of West African Senior Secondary Certificate Examination (WASSCE) also concluded that the test items of WASSCE mathematics were inherently multidimensional in nature.

Furthermore, Okwilagwe and Ogunrinde (2017) found that fifty (50) items of 2013 WASSCE and sixty (60) items of National Examinations Council (NECO) Geography individually 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.

**Research Question 2:** What IRT model best fits the data of 2022 NECO mathematics multiple-choice test items?

**Table 2: IRT Model Best fit of 2022 NECO Mathematics Multiple-Choice Test Items**

Model	AIC	-2Loglik
1PL	131533.5	131321.4
2PL	131534.5	131324.4
3PL	130648.8	130318.8
4PL	130630.4	130180.0

The analysis of model data fit was addressed by subjecting the responses to full information item factor analysis, and compared using Akaike Information Criterion (AIC), and Likelihood Ratio Test (LogLik). The essence of this analysis is to establish the best fit model that provided the information for the calibration of item parameters embedded in the test data, and this, according to Thorpe and Favia (2012), is to ensure the validity of the test items. The result in Table 2 indicated that 4PLM has the smallest information criteria, and hence, the best model fit for the data. This is consistent with Thorpe and Favia (2012) that the smallest -2loglik value is the best fit.

### **Conclusion and Recommendation**

From the result of this study, it was concluded that the 2022 NECO Mathematics multiple-choice test items were multidimensional, and hence violated the unidimensionality assumption of the IRT. The 4-parameter logistic IRT model was also found out to be the best fit for the 2022 NECO Mathematics multiple-choice test items. It was recommended that examination bodies, and researchers involved in the use of IRT should ascertain the conformity of the test items to the IRT assumptions in order

to avoid bias in item parameter estimation, and hence, reduced validity and reliability of the test items.

## **References**

- Adedoyin, C. (2010). Investigating the invariance of persons parameter estimates based on classical test and item response theories. *An International Journal on Education Science*, 2(2), 107-113.
- Anyanwale, M. A., Isaac-Oloniyo and 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.
- Demirtaslı, N.C. (2002). A study of Raven Standard Progressive Matrices test's item measures under classical and item response models: An empirical comparison. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 35(2), 71-79.
- Federal Republic of Nigeria (FRN) (2013). *National policy on education* (6th ed.). Lagos: NERDC press.
- Hambleton, R. K., Robin, F., and Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H. Tinsley and S. Brown (Eds.) *Handbook of applied multivariate statistics and modelling*. San Diego, CA: Academic Press.
- Kyung, T. H. (2013). *Windows software that generates IRT parameters and item responses: Research and evaluation program methods (REMP)*. University of Massachusetts Amherst. Retrieved from <https://www.umass.edu/remf/software/simcata/wingen/homeF.html>
- Nworgu, B.G. (2015). *Education research: Basic issues and methodology*. Nsukka: University Trust Publishers.
- Obinne, A.E. (2013). Test item validity: Item response theory (IRT) perspective for Nigeria. *Research Journal in Organizational Psychology and Educational Studies* 2(1). Retrieved March, 03, 2019, from [www.emergingresource.org](http://www.emergingresource.org)

- Oguoma, C. C., Metibemu, M.A. and 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.
- Ojerinde, D.P. (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 23<sup>rd</sup> May.
- Okoye, R.O. (2015). *Educational and psychological measurement and evaluation*. Awka: Erudition Publishers.
- Okwilagwe, M. A. and Ogunrinde, E.A. (2017). Assessment of unidimensionality and local independence of WAEC and NECO 2013 Geography Achievement Tests. *African J. Theory Pract. Educ. Assess.*, 5, 31-44.
- Omorogiuwa, O.K. (2010). *An introduction to educational measurement and evaluation*. Benin: Perfect Touch Prints.
- Thorpe, G. L., and Favia, A. (2012). *Data analysis using item response theory methodology: An introduction to selected programs and applications*. Retrieved from [http://digitalcommons.library.umaine.edu/psy\\_facpub/20](http://digitalcommons.library.umaine.edu/psy_facpub/20)
- Ukwuije, R.P.I. (2012). *Educational assessment: A sine qua non for quality education*. 83<sup>rd</sup> edition of the inaugural lecture series at University of Port Harcourt.