ISSN 2349-7831

International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 12, Issue 1, pp: (31-36), Month: January - March 2025, Available at: www.paperpublications.org

ITEMS DISTRIBUTION AND RELIABILITY COEFFICIENT OF MATHEMATICS ACHIEVEMENT TEST ADMINISTERED AMONG JUNIOR BASIC SECONDARY SCHOOL STUDENTS IN DELTA STATE

[1]Barnabas Othuke UMUOIGHORO, (Ph.D), [2]Andrew Moyioritse MEGBELE, (Ph.D), [3]OKAGBARE freedom

[1]Department of Educational Psychology, School of Education, Çollege of Education, Warri, Delta State, Nigeria.

[2]Department of Educational Psychology, School of Education, Çollege of Education, Warri, Delta State, Nigeria.

[3]Department of Guidance and Counselling, Faculty of Education, Delta State University, Abraka

DOI: https://doi.org/10.5281/zenodo.15100866

Published Date: 28-March-2025

Abstract: **The study investigated questions distribution and reliability coefficient of mathematics achievement test administered among junior Basic secondary school students in Delta State. Two research questions guided the study. Instrumentation research design was used for the study. The population of the study was 19, 693 Junior Basic secondary school two (JSII) students in Delta State. The sample size of this study was 900 students drawn from secondary schools. The sample for the study was obtained using the multi-stage sampling procedures. The research instrument was designed by the researcher. The items in the instrument were selected from the SS2 Delta State Ministry of Education curriculum. 150 test items were drawn based on table of specification that covered the contents of Numbers and Numeration, Basic Operations Algebraic Processes Measuration, Geometry and Everyday Statistics. The research question one was answered using frequency and percentage while research question 2 was answered using Cronbach's alpha reliability coefficient and Test Response Function (TRF). The findings of the study showed that items in the test are evenly and adequately distributed according to the curriculum of Delta State Ministry of Education and reliability coefficient of the mathematics achievement test is 0.96. In line with the findings and conclusions of the study, it was recommended among others that all items having bad difficulty and discrimination should be modified so that they will be useful in the assessment of Junior Basic secondary school students in Mathematics**

Keywords: **Items Distribution, Reliability Coefficient, Mathematics Achievement Test.**

1. INTRODUCTION

Mathematics is one of the compulsory subjects that students should get at credit grade level to secure admission into institution of higher learning and pursuant of other academic activities. During the period of acquiring mathematics skills at the secondary level, most teachers apply designed mathematics achievement tests for determining students' continuous assessment process, mock examination and weekly assignment (Akemieyefa, Okagbare & Odili, 2025). This means that various ethnic groups irrespective of religion in Nigeria used the developed achievement test. Achievement tests are interval in nature meaning that in educational measurement, there is no zero with regard to achievement tests. Achievement tests are of two broad categories which are the classroom test (popularly called teacher-made test) and standardized tests (Igabari & Okagbare, 2023). The distribution of mathematics achievement test items among students has been a focal point of educational research. Simply put, in order to meet the mathematics curriculum as a touchstone of intelligence for scientific

and technological innovations, questions/items must be well distributed. Umuoighoro (2022) opined that a bid to assessing learner's performance, certain action verbs should be used in test development for adequate representation of both lower and higher order levels of cognitive domain (Knowledge, Comprehension, Application Analysis, Synthesis and Evaluation) for appropriate skills acquisition in line with the 9-Year revised basic curriculum. The distribution of mathematics test items is a fundamental aspect of educational assessment, reflecting the emphasis placed on various mathematical content areas. This distribution ensures that assessments accurately measure students' proficiency across the curriculum and align with established learning standards. It was on this note that NAEP (2023) design strategic distribution of items across different mathematical content areas, such as number properties, algebra, geometry, and data analysis. This approach ensures a comprehensive assessment of students' mathematical understanding. NAEP's distribution of mathematics questions for 2019 and 2024 highlights the varying emphasis on these content areas at different grade levels, reflecting the relative importance assigned to each within the curriculum.

State assessments also adhere to specific guidelines regarding the distribution of mathematics questions. Mathematics Learning Standards outline the standards designations, detailing the allocation of questions across various mathematical domains (NYS, 2024). These designations are intended to guide educators in aligning instruction with the standards and preparing students for the assessments that provide a clear framework for the distribution of test items. The design and development of mathematics tests involve careful consideration of the distribution of questions to ensure they accurately assess the intended learning outcomes. For example, the Massachusetts Comprehensive Assessment System (2020) provides detailed information on the test design for mathematics, including the distribution of questions across different content areas and grade levels. This structured approach ensures that assessments are both fair and reflective of the curriculum. The African Mathematics Programme (AMP) was tasked with preparing and producing teaching textbooks for use in schools to ensure quality distribution of test questions. According to Umuoighoro (2022) stated that topics like algebra, geometry, and trigonometry were not taught, hence the curriculum was largely dependent on arithmetic processes which were not well distributed. Based on this, it is evident that certain changes to the Mathematics curriculum were needed to plea for curriculum reform across Africa was issued a few months before the country's independence.

On the other hand, a reliable test is a test whose measures are consistent and stable. This means that when one uses a reliable test to measure a similar behaviour, the same outcomes are expected (Igabari & Okagbare, 2023). An unreliable test is a test which today gives one result and at other times gives a totally different result, such a test cannot produce acceptable results. Achievement tests are expected to be reliable. In other words, Mathematics Achievement Tests, whether teacher-made or standardized, are expected to measure the knowledge of the students in the content area at all times, and for any group of students. Hence, they are supposed to be stable and consistent. Various statistical techniques are used in establishing the reliability coefficient of measurement instruments according to Umuoighoro (2022) includes test re-test, parallel form, split-half, Kuder Richardson formula 20 and 21, and Cronbach Coefficient Alpha.

Test re-Test: This is a method of establishing reliability; it is the extent to which a measuring instrument measures the stability of scores overtime. This method of estimating the reliability coefficient of a measuring instrument indicates whether students or testees would get essentially the same scores if they took the test at different times. The instrument is administered twice on the same respondents with at least two weeks' interval and then scored. The two sets of scores on the different administrations are correlated using Pearson Product Moment (PPM) correlation to obtain an index of stability (Kpolovie, 2014; Iweka, 2014; Osadebe, 2013; Orluwene, 2012; Ukwuije & Opara, 201, Igabari & Okagbare, 2023).

Parallel Form: This is another method of establishing reliability; it is the degree to which two measuring instruments with similar difficulty level measure equivalence. The respondents are given the two instruments at the same time. The scores from the two instruments are then correlated using PPM to obtain a coefficient of equivalence.

Split-Half: This is a method of estimating reliability coefficient by administering a single measuring instrument whose scoring is based on odd and even numbers; to obtain a measure of internal consistency. Coefficient of internal consistency is an estimate of reliability that describes the extent to which the items in a test are internally consistent or homogeneous with one another. The scores so obtained are correlated using PPM to obtain a coefficient called half reliability (rh) or half internal consistency. Since the full internal consistency is required, the Spearman Brown statistical formula is then applied.

Kuder Richardson formulae (KR-20 and KR-21): These statistics measure internal consistency of test items. The method is also a single administration type of technique for establishing reliability of test instruments. Both formulae are denoted by r. KR_{20} determines the reliability of test items that are dichotomous. That is, test items are scored 1 or 0 for right and wrong responses respectively, based on the proportion of correct or incorrect responses to each item in a test. Proportion of individuals passing an item is denoted by 'p' and that failing the item by 'q' (that is, q=1-p) (Osadebe, 2013). KR21 on the

ISSN 2349-7831

International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 12, Issue 1, pp: (31-36), Month: January - March 2025, Available at: www.paperpublications.org

other hand, is based on the mean and variance of the test scores. It assumes all the test items are of equal difficulty (or that the average difficulty level is 0.5). The difference between KR20 and KR21 is that, KR21 assumes all items to be of equal difficulty. That is, p is constant for all items (Ukwuije & Opara, 2012; Igabari & Okagbare, 2023). According to these researchers, KR21 is more useful to the classroom teacher. It does not require much computation compared with KR20 (Mehrens & Lehmann as cited in Ukwuije & Opara, 2012).

Cronbach Coefficient Alpha: This is also a single administration technique where multiple levels of response like the Likert scale type, essay test, personality tests, etc are involved. In other words, test or self-reporting inventories that the scoring of each item takes a range of values, the appropriate procedure for the determination of the internal consistency reliability is the Cronbach Coefficient Alpha (Kpolovie, 2014). This formula was developed by Cronbach (Ukwuije & Opara, 2012).

Research Questions

The following research questions guided the study:

1. How do the questions distribute across the topics in the mathematics achievement test curriculum administered among students in Delta State?
2. What is the reliability coefficient of the mathematics achievement test administered among students in Delta State?

2. METHODOLOGY

Instrumentation research design was used for the study. Ileh (2017) defined instrumentation design as the procedure that is used by measurement experts for developing tests or modifying an existing instrument for collecting data. The population of the study is 19,693 junior Basic secondary school two (JSII) students in Delta State. The sample size of this study was 900 students drawn from the secondary schools. The study adopted the multi-stage sampling procedures. Firstly, the State was stratefied into Delta Central, Delta North and Delta South senatorial districts. Then, Simple Random Sampling Technique (SRST) was used to select nine (9) Local Government Areas from the senatorial districts. At the second stage, Simple Random Sampling Techniqe was applied to select 10 junior basic secondary schools in each of the Senatorial Districts making a total of 90 jenior basic secondary schools. In the final stage, ten (10) students were drawn from each junior basic of Delta State making a total of 900 students. The research instrument was designed by the researcher. The items in the instrument were selected from the JS2 Delta State Ministry of Education curriculum. 150 test items were drawn based on table of specification that covered the contents of Numbers and Numeration, Basic Operations, Algebraic Processes, Measuration, Geometry and Everyday Statistics. The research question one was answered using frequency and percentage while research question 2 was answered using Cronbach's alpha reliability coefficient and Test Response Function (TRF).

3. PRESENTATION AND DISCUSSION OF RESULTS

Research Question One

How do the questions distribute across the topics in the mathematics achievement test curriculum administered among students in Delta State?

Table 1: Frequency and Percentage distribution of questions across topics in the mathematics achievement test curriculum administered among students in Delta State

Content Area	Levels of Cognitive Objectives						
	Knowledge 10%	Comprehension 12%	Application 16%	Analysis 20%	Synthesis 18%	Evaluation 24%	Total 100%
Numbers and Numeration (24%)	3	9	6	6	3	9	36
Basic Operations (20%)	1	2	6	6	6	9	30
Algebraic Processes (16%)	3	3	6	6	3	3	24
Measuration and Geometry (22%)	3	6	3	6	6	9	33
Everyday Statistics (18%)	1	2	3	6	9	6	27
Total 100%	11	22	24	30	27	36	150

Table 1 shows the frequency and percentage distribution of questions in the Mathematics Achievement Test across the topics. The result showed that the items in the test are evenly and adequately distributed according to the curriculum of Delta State Ministry of Education.

Research Question Two

What is the reliability coefficient of the mathematics achievement test administered among students in Delta State?

Table 2: Summary Statistics for the Total Scores indicating person separation reliability of the MAT Administered among Students in Delta State

Test	Items	Alpha	Mean	SD	Skew	Min	Median	Max	IQR
Full Test	150	0.96	105.54	24.71	-0.67	29.00	106.00	148.00	32.00

Table 2 shows the Alpha value of the full test. The Alpha value is 0.96 which tends towards 1 and it indicates a strong reliability. Therefore, the reliability of the MAT scale within the framework of IRT as indicated by the Alpha value is 0.96. Figure 1 display a graph of the Test Characteristics Curve (TCC) for all calibrated items. The TCC predicts the proportion or number of items that an examinee would answer correctly as a function of theta. The left Y-axis is in proportion-correct units while the right X-axis is in number-correct units. In this case TCC will predict 96% or its equivalent of the score of each examinee on the MAT.

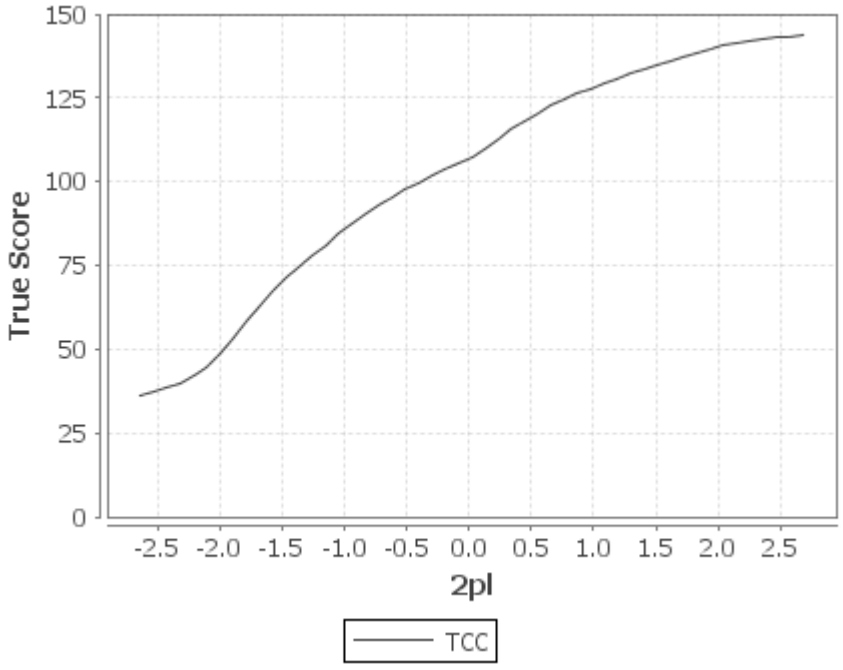

Fig 1: Test Characteristics Curve for the MAT

4. DISCUSSION OF FINDINGS

Questions Distribution across Topics in Mathematics Achievement Test

The first finding revealed that the items in the test are evenly and adequately distributed according to the curriculum of Delta State Ministry of Education. This was as a result of the high degrees of agreement among experts on the percentage weight assigned to the objectives in the cognitive domain and on content areas of School Mathematics. The result indicates that all the objectives and content areas were well covered in the test blueprint. This is an evidence that the test has adequate content validity and can be used to assess upper basic education students in Delta State. The test can be used during mock exam as well as in the classroom for the assessment of students' achievement in Mathematics. The above finding agrees with Ene (2017), who developed and calibrated Basic Science Achievement Test Using the Two-Parameter Logistic Model of Item Response Theory (IRT), and found that the validity and reliability of the instrument were established according to

the two-parameter IRT model. The finding is also consistent with the finding of Taiwo (2020), who developed and investigated the item discrimination /difficulty of each item in chemistry achievement test. His finding showed that the developed test is valid and fit for use.

Reliability Coefficient of the Mathematics Achievement Test

The second finding revealed that the Mathematics Achievement Test is highly reliable. Reliability in item response theory is usually inferred from the Test Characteristics Curve (TCC) since the concept of TCC is similar to the concept of reliability in Classical Test Theory. The TCC is pictorially displayed in Figure 1. Reliability in this case is conceived as the person separation reliability or item separation reliability. The person separation reliability is similar to Cronbach's α. This is the degree to which the Mathematics Achievement Test differentiates persons in the test's outcome. The range is 0 – 1. Item separation reliability, on the other hand, is the degree to which item difficulties are differentiated. Again, the range is 0 – 1. The above finding agrees with Kpolovie and Emekene (2016), who in their study of item response theory validation of advanced progressive matrices in Nigeria, used the same method and obtained a reliability index of 0.94 for the test. The finding also agrees with Ani (2014), who applied item response theory in the development and validation of multiple-choice test in Economics. The result of her study showed that 49 items of the multiple choice questions in Economics were reliable based on three parameter model (3PL) logistic model. The finding is further consistent with the finding of Eleje and Esomonu (2018). The authors developed and validated a test to measure achievement in quantitative economics among secondary school students. Analysis were attained through SPSS and Bilog MG using Item Response Theory (IRT) three-parameter logistic model (3PL) to establish item difficulty, item discrimination, and the guessing value. The empirical reliability of the test was 0.86. The test was found to be of good quality, valid and highly reliable. The result of the second finding also seems to be in line with the study of Ezechukwu, Oguguo, Ene and Ugorji (2020), who determined the psychometric properties of the Economics Achievement Test (EAT) using Item Response Theory (IRT). Two popular IRT models namely, one-parameter logistics (1PL) and two-parameter logistics (2PL) models were utilized just like in the current study. Reliability and validity for each item and for the whole test were established according to the one-parameter and two-parameter logistic models. The finding revealed that the instrument was highly reliable and fit for use.

5. CONCLUSION

Based on the findings, it can be concluded that the items in mathematics achievement test are well distributed. The study also concluded that mathematics achievement test is reliable and can be used to assess the Mathematics achievement of upper basic school students in Delta State.

6. RECOMMENDATIONS

Based on the findings from this study, the following recommendations were made:

1. All items having bad difficulty and discrimination should be modified so that they will be useful in the assessment of school students in Mathematics
2. Psychometricians should add this reliable test for item bank development used for examinations especially in computer-based test.

REFERENCES

[1] Akemieyefa, S., Okagbare, F. & Odili, J.N. (2025). Difficulty and Discriminatory Indexed of Developed Mathematics Achievement Test in Bayelsa State.

[2] Ani, E. N. (2014). *Application of item response theory in the development and validation of multiple choice test in economics.* (Master's thesis). University of Nigeria, Nsukka.

[3] Eleje, L. I., & Esomonu, N. P. M. (2018). Test of Achievement in Quantitative Economics for Secondary Schools: Construction and Validation Using Item Response Theory. *Asian Journal of Education and Training,* 4(1), 18-28.

[4] Ene, C. U. (2017). Development and calibration of a basic science achievement test using the two-parameter logistic model of Item Response Theory (IRT)

[5] Ezechukwu, R. I., Oguguo, B. C. E., Ene, C. U., & Ugorji, C. O. (2020). Psychometric Analysis of Economics Achievement Test Using Item Response Theory. *World Journal of Education,* 10(2), 59-68.

[6] Igabari, Q & Okagbare, F. (2023). *Assessment of test development skills of Secondary School Physics Teachers in Delta Central Senatorial District*. Innovations, Number 73

[7] Iweka, F. (2014). *Comprehensive guide to test construction and administration.* Omoku: Chifas Nigeria.

[8] Kpolovie, P. J. & Emekene, C.O. (2016) Psychometric advent of advanced progressive matrices- smart version (APM-SV) for use in Nigeria. European journal of statistics and probability, 4(3), 20-30.

[9] Kpolovie, P. J. & Emekene, C.O. (2016) Psychometric advent of advanced progressive matrices- smart version (APM-SV) for use in Nigeria. European journal of statistics and probability, 4(3), 20-30.

[10] Kpolovie, P. J. (2014). *Test, measurement and evaluation in education* (2nd edition). New Owerri: Spring Field Publishers Ltd.

[11] NAEP (2023). NYS Next Generation Mathematics Learning Standards for Grades 3-8: Post-Test Standards Designations Applicable to Test Administrations.

[12] Orluwene, G. W. (2012). *Introduction to test theory and development process.* Port Harcourt: Chris-Ron Integrated Services.

[13] Osadebe, P. U. (2013). Validity and reliability of evaluation techniques for quality education. *DELSU Journal of Educational Research and Development,* 12 (1), 56-63.

[14] Osadebe, P. U. (2016). *Practical guide to item generation.* A seminar presented at Delta State University, Abraka

[15] Taiwo, A. K. (2020). A Three Model Parameter Of Chemistry Achievement Test Among Secondary School Students In Ibadan Metropolis, Nigeria. *International Journal of Scientific and Research Publications,* 10(6), 751-760.

[16] Ukwuije, R. P. I., & Opara, I. (2012). *Test and measurement for teachers.* Alakahia-Akpor: Chadik Printing Press.

[17] Umuoighoro, B. O. (2022). Development And Validation Of Mathematics Achievement Test For Junior Basic Schools In Delta State. Unpublished Ph.D Thesis of Delta State University, Abraka.