

---

## Construction And Validation Of Mathematics Aptitude Test For Junior Secondary School Students In Rivers State Using Programming Technology

Dr. Ibiene T. Longjohn<sup>1</sup>, Dr. Allwell Sunny Njigwum<sup>2</sup>, Osila Kenneth<sup>3</sup>,  
Anunaobi James Chibuike<sup>4</sup>, Amadi Chizoba Perpetua<sup>5</sup>

Department of Educational Psychology, Guidance and Counselling, Faculty of Education, Ignatius Ajuru University of Education, Rumuolumeni, Rivers State, Nigeria.

---

### Abstract

The study investigated the construction and validation of a Mathematics Aptitude Test (MAT) for Junior Secondary school students in Rivers State using R programming technology. The study employed an instrumentation design using Classical test theory approach to construct and validate a Mathematics Aptitude Test (MAT) for Junior secondary school students. A sample of 420 students were randomly selected through stratified random sampling technique. Three research questions guided the study. The Mathematics Aptitude Test (MAT) covered ten topics, which were used for the generation of 150 items. The content validity of the MAT was established using a test blueprint. The data collected were analysed using R software to find out the reliability, item difficulty and item discrimination under the CTT framework. The test reliability was established through Cronbach Alpha, Split-half and Kuder Richardson-20 statistics using R-software and SPSS, which produced reliability coefficients of 0.94, 0.96 and 0.97 respectively. The result indicated that 82 items (54.6%) were considered good items, while 68 items (45.4%) were marked for rejection. From the selected items, the difficulty indices ranged from 0.40 to 0.60 while discrimination indices ranged from 0.22 to 0.64. Since the mathematics aptitude test (MAT) was highly valid and reliable, it was recommended among others that the instrument be used as an assessment tool for mathematics during promotional examinations from JSS 1 to JSS 2 for junior secondary school students in Rivers State.

**Keywords:** Aptitude test, Mathematics, Test construction, Classical Test Theory (CTT), R-software.

---

### Introduction

The introduction of education from the Western countries to Nigeria mostly consisted of oral and informal methods. Nevertheless, the evaluation of traditional education mostly relied on oral methods. During this time span, formal or written exams were not used as a method of assessment (Ohuche, et. al. in Patrick, et. al. 2018). The introduction of western education in Nigeria led to the establishment of a standardized system of testing and assessment in the school system. The exams were divided into

two categories: internal and external examinations. Internal examinations refer to test items that are created and administered within the school, such as end-of-term examinations and promotion examinations. On the other hand, external examinations are administered by public examination bodies and serve the purpose of certifying students, determining their placement, and providing job opportunities. Some of the examination organizations mentioned include the West African Examination Council (WAEC), State's Ministries of Education (SMOE), Joint Admission and Matriculations Board (JAMB), and the National Examination Council of Nigeria (NECO). Egbule (2002) observed that the word "test" refers to a collection of activities or a presentation of a predetermined set of things that elicit answers from the individuals being examined.

Aptitude refers to an individual's inherent capacity for learning and their specific potential to excel in a certain vocation or activity. It encompasses their appropriateness, preparation, propensity, and natural or learned temperament. Aptitude tests are cognitive assessments that help predict pupils' future success in schools where they participate in activities (Aiken in Usha, et. al. 2022). Aptitude testing is an effective method for enhancing the cognitive abilities and memory capabilities of pupils. These tests are used to assess an individual's aptitude for doing a certain activity, without any prior knowledge or instruction. Typically, those with a passion for research choose to design experiments in fields where they feel more at ease. In Usha, et. al. (2022), Freeman has provided a definition of aptitude as a combination of traits that specifically pertain to an individual's ability to acquire various abilities.

An aptitude test is a kind of assessment that aims to evaluate an individual's inherent qualities or attributes in relation to their capacity to acquire a skill or talent that is highly regarded in society, under favorable conditions (Schneider, 2013). The latent potential of aptitude, an inherent quality of an individual, cannot be readily seen. This trait may only be deduced by assessing observable qualities that are either commonly associated with the testee's ability or are indicative of the future progression of the ability. Aptitude is often evaluated by examining attributes that are considered significant precursors of success. Those with speech comprehension has a superior capacity for reading comprehension compared to those without speech comprehension. These antecedents may themselves constitute a kind of achievement. It is crucial to highlight that aptitudes are inherent traits, meaning they are present at the time they are assessed. Therefore, the concept of aptitude does not imply whether they are acquired or inherited, or if they result from a combination of environmental and genetic factors (Miller, et. al. 2009).

According to Ukuije, et. al. (2012), an aptitude test is able to forecast an individual's future capabilities. According to Onunkwo in Inko-Tariah, et. al. (2017), they solely assess a single, particular area of skill. Cognitive measurements are used to predict individuals' performance in various domains such as mechanical, clerical, algebraic, foreign languages, musical, and welding. Some examples of standardized aptitude exams include the Bloomberg Aptitude Test (BAT) and the Armed Service Vocational Aptitude Battery (ASVAB). An aptitude test evaluates cognitive capacities that are often not explicitly taught inside educational institutions (Fraenkel, et. al. 2003). Aptitude tests are designed to assess an individual's capacity to accomplish. In essence, they evaluate the innate talents or abilities of a person. An aptitude test assesses an individual's potential performance and has a significant

predictive component. An aptitude test aims to assess the examinees' potential for future success in a certain activity. Aptitude tests only evaluate previous knowledge and skills. They do not immediately evaluate inherent aptitude or capability. Moreover, due to the challenge of acquiring a future performance sample, these tests are unable to directly evaluate future capability.

Nitko (2004) observed that an aptitude test is only one indicator of aptitude. An individual's aptitude for X refers to their current ability and potential for future success in X, assuming that the circumstances of the past and present persist. Additional indicators of aptitude may encompass scores on achievement tests, data on past performance in activities analogous to those in which one seeks to forecast success, and information obtained from procedures for evaluating personality, interests, attitudes, physical abilities, and psychological condition (Nitko, 2004). There are two main types of aptitude testing: specialized exams and multi-aptitude batteries (Salkind, et. al., 2007). They emphasized that job-specific aptitude tests are often designed to assess whether candidates are most appropriate for a certain job or function within a particular profession. The multi-aptitude batteries are often used in educational settings and large-scale job testing scenarios to determine selection, placement, or categorization choices (Cohen, 2018).

In the present day, there is a growing need for pupils to develop and enhance their cognitive abilities via creativity and invention. Furthermore, there is a rapid and increasing progress in the field of science and technology. Mathematics, being an abstract discipline, has extensive applications in other domains. It exhibits correlation with other topics and has relevance to daily life. This would have enticed the pupils to engage with maths. Success depends on a multitude of circumstances, with the presence of fundamental skills being crucial. The pressure exerted by parents compels them to make misguided choices, ultimately resulting in the deterioration of their physical and emotional well-being. Assessing or approximating a student's potential or skill prior to their selection or enrollment in a particular area might help save a significant amount of wasted effort. The assessment of aptitude is crucial in this situation. Presumably, this research aims to empower pupils to identify their aptitudes and deficiencies in order to achieve success in life. The aptitude assessments determine the future trajectory of the pupils (Chatterjee, 2007).

Each student's aptitude is distinct since it reflects the skill level of an individual (Ramsay, 2008; Raza, 2011). Kubiszyn, et. al. (2003) have proposed a novel term for aptitude, defining it as one's inherent potential or capacity. In Usha, et. al. (2022), the term "aptitude test" is defined as a tool used by teachers to evaluate students' talents in a practical manner and identify those who are not performing up to their potential. The specific aptitude required to improve learning, such as for a profession, readiness, suitability, or inherent or acquired inclination or capability for a specific task or assignment (Raza, 2011; Reeves, 2002), and the level of readiness to learn and perform effectively in a specific situation or within a defined area (Raza, 2011). An aptitude exam assesses the comprehensive performance over a broad range of mental capacities of the pupils. It consists of verbal and numerical skills that aim to assess essential mathematical abilities.

According to Longjohn, et. al. (2021), a test may be analyzed from several viewpoints, and the questions inside the exam can be assessed based on multiple theoretical frameworks. Two theories that fall within this category are the Classical

Test Theory (CTT) and the Item Response Theory (IRT). These ideas, known as the two primary frameworks in educational measurement, are used to construct, assess, ascertain the dependability and validity of tests, and enhance the quality of test items. These frameworks rely on distinct assumptions and use diverse statistical methodologies. Initially, CTT served as the predominant framework for creating and evaluating standardized exams. Subsequently, IRT was created to enhance the function of CTT. However, the research will primarily focus on the use of CTT procedures. CTT is predicated on the premise that a test taker has both an observed score and a real score. The seen score of a test-taker is often regarded as the sum of an estimation of the real scores of that test-taker, together with some unobservable error that may be positive or negative.

The real score represents the test-taker's actual knowledge, however it is inevitably influenced by several sources of inaccuracies. CTT employs assessments of item features, item difficulty, and item discrimination, which are influenced by the distribution of examinee competence in a given sample. While traditional test theory may be applied to many test building circumstances, the assumptions it is founded on seem to introduce vulnerabilities in the model. The statistical indices based on CTT are easily calculable, manipulable, and comprehensible to individuals without specialized knowledge. However, these indices exhibit variability across different samples. CTT has theoretical assumptions that are rather weak, hence facilitating its use in various testing scenarios (Hambleton, in Awopeju, et. al. 2016). Although CTT has shown great use in test construction, it is important to note that the two fundamental statistics it relies on, item difficulty and item discrimination, are contingent upon the specific sample being analyzed.

R is a programming language and a freely available software environment for doing statistical computation and generating visual representations of data (Weston, et. al. 2017). Researchers from many fields often use it to assess and present findings, while statistics and research techniques instructors frequently employ it. Although it is free, this option requires the use of computer code rather than drop-down menus or buttons to accomplish the task, which may be appealing to some. Although programming languages may seem daunting, we use R for both research and teaching, since we firmly think that the advantages surpass the initial time and effort required to begin. An exceptional attribute of R is its open-source nature, which grants unrestricted access to the underlying code used to execute the program, allowing users to freely include their own code. This implies that R will consistently possess the capability to execute the most recent statistical analyses promptly upon their conception.

Additionally, the R community is renowned for its dynamic contributions in the form of packages and has fostered a network of programming and statistics enthusiasts (referred to as "use Rs") who are available to provide assistance. Furthermore, individuals have the ability to compose their own R code, therefore expanding the extensive repertoire of tools available in R. Programmers submit their work to R by packaging it into software modules called packages. Certain programs are designed to focus on certain types of analysis, whilst other packages have a more comprehensive range of capabilities. As an example, the 'psych' package developed by William R. Revelle is capable of performing a wide range of tasks, including descriptive statistics, item-response theory, and mediation analysis (Weston, et. al. in Longjohn, et. el. 2021). As of the beginning of 2017, there are slightly less than

10,000 bundles available. Whenever a novel statistical methodology is devised, it is certain that someone will promptly design a new software package or include more features into an already existing program.

Test building refers to the methodical process of creating test items or preparing a test by gathering and organizing a sequence of questions that make up the assignment for students. Test validation, as described by Gronlund in Patrick, et. al. (2018), is a statistical approach used to standardize test items. Its purpose is to eliminate any biases and ensure that the test is valid, reliable, objective, and useful. Patrick and his colleagues (2018) proposed a framework for creating a test that is efficient and well-structured. He observed that the most probable way to produce a suitable sample is by following a methodical strategy for test preparation. He listed the following procedures as advantageous for this objective, The process involves formulating broad instructional goals, creating a detailed summary of the material to be taught, developing a table that specifies the desired outcomes, and designing test questions that assess the achievement of these objectives as outlined in the table.

Furthermore, according to Okobia in Patrick, et. al. (2018), certain protocols are adhered to while developing and verifying a test. Consequently, the majority of measurement and evaluation specialists have reached a consensus that test preparation is a methodical and organized procedure. Nevertheless, the building and validation procedures used by authors may differ based on the characteristics of the test. In Patrick et al.'s (2018) study, Brown states that Educational Testing Services (ETS) proposed a set of standard procedures for constructing and validating tests. These procedures include planning the test, writing the items, pretesting the items, preparing the form, collecting reliability and validity evidence, and developing normative and interpretative materials.

According to Osadebe (2001), principles of test building serve as the recommendations for creating a test. Test construction is a specialized field that pertains to the domain of differential psychologists and professionals in test and measurement. The test is a genuine and dependable tool. Teachers should adhere to a set of rules while developing test items, including planning, item authoring, item analysis via trial testing, composing the items, ensuring reliability, printing and administration, marking and scoring, and creating a manual.

### **Statement of the Problem**

An issue of significant concern in the Nigerian educational industry is the lack of high-quality evaluation tools. Inadequate assessment is a significant factor contributing to the persistent issue of high rates of test misconduct and fails, particularly in scientific areas such as mathematics, as reported annually by examination organizations. Many students often fear external tests because they perceive them to be challenging, which motivates them to explore all available methods to succeed in these exams. One of the elements that might contribute to examination malpractice and failure is the insufficient preparation by classroom instructors, who do not provide pupils with high-quality tests similar to those utilized by external examining organizations.

Furthermore, the issue of insufficient quality assessment tools is exacerbated in educational institutions, where instructors are instructed to administer a reduced number of questions for end-of-term exams in order to minimize the expenses

associated with question papers. The issue might escalate to the extent of having a mere five to ten objective exam elements and two essay questions, which often lack subject validity. The inability of classroom instructors to create low-quality test items is partially attributed to their insufficient expertise in test creation and validation. According to Awai (2017), many competent instructors struggle with test creation. They may believe they are doing well and evaluating well, but without conducting item analysis, they cannot really ascertain their performance.

It is important to emphasize that in order for education to be really beneficial, it must be based on high-quality evaluation. In order for the Nigerian economy to effectively compete on a global scale in the fields of science and technology, it is imperative to prioritize the advancement of mathematics education via the promotion of high-quality learning and evaluation. Therefore, to fill in the gap of providing appropriate training for teachers in test construction, the researchers have carried out the construction and validation of mathematics aptitude test for junior secondary school students in Rivers State using R programming technology.

### **Aim and Objectives of the Study**

The main purpose of the study was to construct and validate a mathematics aptitude test (MAT). Specifically, the study determined the following:

1. Reliability of the constructed mathematics aptitude test (MAT) items.
2. Item parameters of the mathematics aptitude test (MAT) using CTT.
3. The number of mathematics aptitude test (MAT) that survived under CTT framework.

### **Research Questions**

The following research questions were used for the purpose of this study.

1. What is the reliability of the constructed Mathematics Aptitude Test (MAT) items?
2. What are the item parameters of the mathematics aptitude test (MAT) using CTT?
3. How many of the mathematics aptitude test (MAT) survived under CTT framework?

### **Methodology**

This study focuses on development of a multiple-choice mathematics aptitude test (MAT). Instrumentation research design was adopted. Instrumentation research design is seen as a study which aims at introducing new or modified instrument for educational practices. The population for this study consists of all Junior Secondary School One (JSS I) students in Rivers State. Six (6) junior secondary schools were sampled. The stratified sampling technique was used to select the six (6) junior secondary schools. The sample size for the study was 420 respondents. The mathematics aptitude test (MAT) developed by the researchers contained 150 items. The test content consists of eight (8) topics which covered a period of 10 weeks. The test content was based on a well-designed test blueprint covering the six levels of Bloom's taxonomy of knowledge, comprehension, application, analysis, synthesis and evaluation (KCAASE) of cognitive domain of learning. It consists of whole numbers with twenty (20) items, four basic operations with fifteen (15) items, factors and multiples with fifteen (15) items, fractions with forty (40) items, decimal fractions with twenty (20) items, order of operations with five (5) items, binary number system with ten (10) items and estimation with twenty (25) items. This shows how the 150

test items in the MAT were distributed among the content areas which covered the six domains of Bloom's taxonomy of learning.

The data for this research was obtained directly with the assistance of subject instructors, who were responsible for administering the exam as an internal assessment. Of the 450 exams administered, 420 test papers (equivalent to 94.5%) were collected and considered complete for analysis. The content validity was determined by using a table of specification. Okorodudu (2012) and Kpolovie (2010) concur that the most accurate method to determine the validity of an accomplishment exam is via content validity, specifically by using a test blueprint. In addition, the test's face and content validities were further confirmed by topic experts in Mathematics. The specialists were given the test form and test blueprint to review and evaluate the instrument. All recommendations and adjustments were implemented prior to conducting trial testing.

The test's reliability was determined using statistical methods such as Cronbach Alpha, Split-half, and Kuder Richardson-20. These methods were implemented using R-software and SPSS, resulting in reliability coefficients of 0.937, 0.956, and 0.974, respectively. The data analysis strategy included utilizing SPSS for data coding and cleaning, and many R tools were used for item analysis using CTT methodologies.

### **Steps taken to validate the mathematics aptitude test (MAT)**

The construction and validation of the mathematics aptitude test followed the recommended steps for an achievement test, as suggested by Asuru (2015) and Orluwene (2012) in Longjohn et al. (2021). These steps include test planning, stating objectives, outlining content, developing a test blueprint, determining the format and length of test items, writing items, ensuring face and content validity, conducting trial testing, analyzing items, computing discriminative and difficulty indices, selecting items, revising, and estimating test reliability. The aforementioned processes were used to direct the development and verification of the 150-item multiple choice aptitude questions. Some phases will be explicitly emphasized here, while others will be included into the methods and outcomes of the research.

**Test Planning:** According to Asuru (2015), test planning encompasses all the preliminary procedures involved in creating a test. These processes include, among other things;

**Stating of the Objectives:** Given that the exam is designed to assess educational or behavioral purposes in the cognitive domain, Bloom's taxonomies of educational aims were used. The elements included are: knowledge, understanding, application, analysis, synthesis, and assessment. Nevertheless, due to the multiple-choice nature of the exam and the fact that it is designed for Junior Secondary School Students, the test will primarily assess lower-level educational goals, including knowledge, understanding, and application (Asuru, 2015).

**Outlining of the Content:** The material for this exam is derived from the mathematics curriculum of JSS (I). The topic encompasses seven overarching concepts derived from the new curriculum for junior secondary school pupils, as presented in Book 1 of the general mathematics course. The topics covered include whole numbers, fundamental operations, factors and multiples, fractions, order of operation, binary system, and estimates. See table 1 below;

**Table 1:** Content area covered during the construction of the mathematics aptitude test (MAT)

S/No.	Topic(s)	Content(s)
1	Whole Numbers	(i) Count and write in million, billion and trillion. (ii) Count and write small numbers. (iii) Roman system of counting. (iv) Place value of numbers.
2	Exercise in Four Basic Operations	(i) Addition and subtraction of numbers. (ii) Division and multiplication of numbers.
3	Factors and Multiples	(i) Definition of prime numbers (ii) Identification of prime numbers (iii) Finding factors of a given whole number (iv) Expression of numbers as a product of its factors in index form. (v) Finding the highest common factor (HCF) of two or more whole numbers. (vi) Writing of two or more multiples of a given number (vii) Finding the lowest common multiple (LCM) of two or more whole numbers.
4	Fractions I	(i) Expressing Quantities in terms of common fractions and mixed numbers. (ii) Changing mixed numbers to improper fractions and vice versa. (iii) Obtaining equivalents of a given fraction. (iv) Reducing fractions to the lowest terms. (v) Arranging fractions in ascending or descending order. (vi) Add, subtract, multiply and divide fractions. (vii) Expressing a quantity as a fraction or percentage of another quantity. (viii) Converting fractions to percentages and vice versa. (ix) Solving word problems involving fractions.
5	Fractions II	(i) Conversion of fractions to decimals and vice versa. (ii) Conversion from percentages and vice versa.
6	Decimal Fractions	(i) Meaning of decimal fractions. (ii) Addition and subtraction of fractions. (iii) Division and multiplication of fractions. (iv) Solve word problems involving fractions.
7	Order of Operations	(i) Add fractions (ii) Subtract fractions
8	Binary Number System	(i) Convert numbers from base 10 to base 2. (ii) Convert numbers from base 2 to base 10. (iii) Add and subtract numbers in base 2.
9	Estimation	(i) Estimation of quantities. (ii) Rounding off numbers to various degrees of accuracy.
10	Estimation Contd.	(i) Use of rounded numbers to approximate the answers to addition, subtraction, division and multiplication problems. (ii)

**Development of a test blueprint:** The test blueprint is a grid that includes course material in one way and the behavioral goals or learning outcomes in the other direction, specifically related to the cognitive domain (Ogunleye, 2000, Orluwene, 2012). The determination of the weighting for each cell may be done subjectively, relying on personal judgment, or objectively, by considering the amount of time dedicated to teaching each topic, as stated by Asuru (2015). Downing (2006) observed that a table of specification is a process that lists the exact information and cognitive activities that will be used to evaluate examinees. Students are evaluated in the test according to their acquired knowledge, which is anticipated to include all subject areas. Linn, et. al. (2002) described a table of specification as the comprehensive representation of the content of a course or program, including both the subject matter material and the instructional goals. This refers to the anticipated level of achievement that children may exhibit. The following table, Table 2, displays the specifications for the 150-item mathematical aptitude exam (MAT) designed for JSS I pupils.

**Table 2:** Test blueprint for 150-item mathematics aptitude test (MAT) constructed for JSS 1 mathematics 1<sup>st</sup> term 2022/2023 academic session

S/No	Contents	Wks	<u>Behavioural Objectives</u>						Total
			K	C	A	A	S	E	
1.	Whole Numbers	1	5	4	10	1	-	-	20
2.	Exercises in four basic operations	2	5	2	8	-	-	-	15
3.	Factors and Multiples	3	1	6	2	1	2	3	15
4.	Fractions 1	4	6	5	6	3	2	3	25
5.	Fractions 2	5	2	2	2	2	3	4	15
6.	Decimal fractions	6	6	3	2	2	3	4	20
7.	Order of Operations	7	-	-	2	3	-	-	5
8.	Binary Number System	8	4	2	2	2	-	-	10
9.	Estimation	9	2	2	2	3	5	1	15
10.	Estimation	10	3	-	5	-	-	2	10
	<b>Total</b>					1	1		<b>150</b>
			34	26	41	7	5	17	

**Determination of the format and length of test items**

The exam that was created consists of 150 items and follows a four-option multiple choice objective structure. A total of 420 pupils in JSS I (UBE Section) took the exam, which lasted for 2 hours and 30 minutes.

**Item Writing:** The formulation of the test items was determined by factors such as the test design, the cognitive abilities of the students, the test's objectives, and the characteristics of the examinees.

### Characteristics of the Testees

- Ownership of School: Public
- Subject: Mathematics
- Class: JSSI
- No. of Testees: 420 students
- Average Age: 11-13 years
- School Type: Single and Co-educational
- Content: JSS I curriculum.

### Results

The results obtained in this study are presented and discussed below.

**Research Question 1:** What is the reliability of the mathematics aptitude test (MAT)? The test's reliability was assessed using three methods: Cronbach alpha, Split-half, and Kuder-Richardson 20. The reliability techniques used a single testing or administration approach, where the resulting scores were used to calculate the reliability coefficient. Upon calculation, the resulting coefficients were acquired as indicators of the test's internal consistency.

**Table 3:** Reliability coefficients using Cronbach alpha, split-half and Kuder-Richardson

Reliability Method	Coefficient
Cronbach alpha	0.937
Split-half	0.956
Kuder-Richardson-20	0.974

Given the sufficiently high coefficient of the MAT, it can be confidently said that the test was trustworthy, as its value above the threshold of 0.7, which is the standard for a dependable instrument.

**Research Question 2:** What are the item parameters of the mathematics aptitude test (MAT) using CTT?

In order to address this inquiry, the test items underwent psychometric analysis using the R programming language. Table 4 displays the CTT data, which include p representing the item difficulty indices and the item discrimination indices.

**Table 4:** Item parameter estimates and survived items under CTT

Items	Difficulty	Remarks	Discrimination	Remarks
ITEM1	0.557	Moderate	0.259	Good
ITEM2	0.545	Moderate	0.437	Excellent
ITEM3	0.298	Difficult	0.634	Excellent
ITEM4	0.774	Easy	-0.023	Poor
ITEM5	0.367	Difficult	0.591	Excellent
ITEM6	0.314	Difficult	0.744	Excellent
ITEM7	0.795	Easy	-0.070	Poor
ITEM8	0.319	Difficult	0.424	Excellent
ITEM9	0.260	Difficult	0.503	Excellent
ITEM10	0.598	Moderate	0.169	Poor
ITEM11	0.698	Moderate	0.093	Poor

ITEM12	0.348	Difficult	0.640	Excellent
ITEM13	0.457	Moderate	0.429	Excellent
ITEM14	0.386	Difficult	0.369	Good
ITEM15	0.681	Moderate	0.166	Poor
ITEM16	0.495	Moderate	0.545	Excellent
ITEM17	0.357	Difficult	0.537	Excellent
ITEM18	0.426	Moderate	0.516	Excellent
ITEM19	0.419	Moderate	0.600	Excellent
ITEM20	0.514	Moderate	0.611	Excellent
ITEM21	0.400	Moderate	0.638	Excellent
ITEM22	0.455	Moderate	0.258	Good
ITEM23	0.350	Difficult	0.470	Excellent
ITEM24	0.481	Moderate	0.477	Excellent
ITEM25	0.431	Moderate	0.566	Excellent
ITEM26	0.386	Difficult	0.570	Excellent
ITEM27	0.550	Moderate	0.131	Poor
ITEM28	0.307	Difficult	0.480	Excellent
ITEM29	0.583	Moderate	0.107	Poor
ITEM30	0.488	Moderate	0.576	Excellent
ITEM31	0.671	Moderate	-0.052	Poor
ITEM32	0.326	Difficult	0.584	Excellent
ITEM33	0.426	Moderate	0.588	Excellent
ITEM34	0.433	Moderate	0.566	Excellent
ITEM35	0.526	Moderate	0.409	Excellent
ITEM36	0.495	Moderate	0.345	Good
ITEM37	0.340	Difficult	0.591	Excellent
ITEM38	0.364	Difficult	0.470	Excellent
ITEM39	0.414	Moderate	0.425	Excellent
ITEM40	0.531	Moderate	0.311	Good
ITEM41	0.457	Moderate	0.332	Good
ITEM42	0.614	Moderate	0.256	Good
ITEM43	0.333	Difficult	0.573	Excellent
ITEM44	0.648	Moderate	0.136	Poor
ITEM45	0.550	Moderate	0.614	Excellent
ITEM46	0.717	Easy	0.156	Poor
ITEM47	0.429	Moderate	0.571	Excellent
ITEM48	0.383	Difficult	0.524	Excellent
ITEM49	0.388	Difficult	0.477	Excellent
ITEM50	0.529	Moderate	0.305	Good
ITEM51	0.464	Moderate	0.606	Excellent
ITEM52	0.410	Moderate	0.261	Good
ITEM53	0.674	Moderate	0.230	Good
ITEM54	0.693	Moderate	0.290	Good
ITEM55	0.419	Moderate	0.633	Excellent
ITEM56	0.398	Difficult	0.573	Excellent
ITEM57	0.443	Moderate	0.553	Excellent
ITEM58	0.543	Moderate	0.449	Excellent
ITEM59	0.533	Moderate	0.510	Excellent
ITEM60	0.393	Difficult	0.553	Excellent
ITEM61	0.448	Moderate	0.571	Excellent

ITEM62	0.467	Moderate	0.448	Excellent
ITEM63	0.436	Moderate	0.487	Excellent
ITEM64	0.495	Moderate	0.553	Excellent
ITEM65	0.081	Difficult	-0.125	Poor
ITEM66	0.407	Moderate	0.337	Good
ITEM67	0.510	Moderate	0.427	Excellent
ITEM68	0.355	Difficult	0.502	Excellent
ITEM69	0.326	Difficult	0.404	Excellent
ITEM70	0.455	Moderate	0.378	Good
ITEM71	0.467	Moderate	0.311	Good
ITEM72	0.407	Moderate	0.490	Excellent
ITEM73	0.460	Moderate	0.386	Good
ITEM74	0.538	Moderate	0.541	Excellent
ITEM75	0.360	Difficult	0.472	Excellent
ITEM76	0.452	Moderate	0.424	Excellent
ITEM77	0.579	Moderate	0.224	Good
ITEM78	0.390	Difficult	0.569	Excellent
ITEM79	0.524	Moderate	0.160	Poor
ITEM80	0.648	Moderate	0.281	Good
ITEM81	0.493	Moderate	0.490	Excellent
ITEM82	0.405	Moderate	0.568	Excellent
ITEM83	0.545	Moderate	0.335	Good
ITEM84	0.350	Difficult	0.309	Good
ITEM85	0.433	Moderate	0.399	Good
ITEM86	0.457	Moderate	0.605	Excellent
ITEM87	0.393	Difficult	0.582	Excellent
ITEM88	0.502	Moderate	0.279	Good
ITEM89	0.536	Moderate	0.423	Excellent
ITEM90	0.369	Difficult	0.652	Excellent
ITEM91	0.388	Difficult	0.636	Excellent
ITEM92	0.431	Moderate	0.528	Excellent
ITEM93	0.514	Moderate	0.488	Excellent
ITEM94	0.452	Moderate	0.297	Good
ITEM95	0.540	Moderate	0.227	Good
ITEM96	0.421	Moderate	0.560	Excellent
ITEM97	0.505	Moderate	0.433	Excellent
ITEM98	0.540	Moderate	0.407	Excellent
ITEM99	0.348	Difficult	0.545	Excellent
ITEM100	0.426	Moderate	0.394	Good
ITEM101	0.350	Difficult	0.534	Excellent
ITEM102	0.424	Moderate	0.603	Excellent
ITEM103	0.336	Difficult	0.476	Excellent
ITEM104	0.348	Difficult	0.487	Excellent
ITEM105	0.298	Difficult	0.453	Excellent
ITEM106	0.395	Difficult	0.378	Good
ITEM107	0.390	Difficult	0.350	Good
ITEM108	0.426	Moderate	0.462	Excellent
ITEM109	0.505	Moderate	0.453	Excellent
ITEM110	0.405	Moderate	0.477	Excellent
ITEM111	0.424	Moderate	0.356	Good

ITEM112	0.543	Moderate	0.370	Good
ITEM113	0.398	Difficult	0.454	Excellent
ITEM114	0.486	Moderate	0.499	Excellent
ITEM115	0.333	Difficult	0.558	Excellent
ITEM116	0.429	Moderate	0.425	Excellent
ITEM117	0.500	Moderate	0.535	Excellent
ITEM118	0.448	Moderate	0.452	Excellent
ITEM119	0.343	Difficult	0.627	Excellent
ITEM120	0.457	Moderate	0.470	Excellent
ITEM121	0.467	Moderate	0.482	Excellent
ITEM122	0.476	Moderate	0.494	Excellent
ITEM123	0.440	Moderate	0.562	Excellent
ITEM124	0.448	Moderate	0.532	Excellent
ITEM125	0.371	Difficult	0.554	Excellent
ITEM126	0.369	Difficult	0.495	Excellent
ITEM127	0.336	Difficult	0.570	Excellent
ITEM128	0.388	Difficult	0.228	Good
ITEM129	0.319	Difficult	0.443	Excellent
ITEM130	0.317	Difficult	0.448	Excellent
ITEM131	0.471	Moderate	0.485	Excellent
ITEM132	0.471	Moderate	0.565	Excellent
ITEM133	0.407	Moderate	0.623	Excellent
ITEM134	0.512	Moderate	0.411	Excellent
ITEM135	0.295	Difficult	0.491	Excellent
ITEM136	0.457	Moderate	0.540	Excellent
ITEM137	0.407	Moderate	0.563	Excellent
ITEM138	0.367	Difficult	0.518	Excellent
ITEM139	0.452	Moderate	0.535	Excellent
ITEM140	0.395	Difficult	0.507	Excellent
ITEM141	0.374	Difficult	0.558	Excellent
ITEM142	0.374	Difficult	0.518	Excellent
ITEM143	0.345	Difficult	0.510	Excellent
ITEM144	0.362	Difficult	0.627	Excellent
ITEM145	0.333	Difficult	0.464	Excellent
ITEM146	0.355	Difficult	0.487	Excellent
ITEM147	0.386	Difficult	0.608	Excellent
ITEM148	0.376	Difficult	0.517	Excellent
ITEM149	0.338	Difficult	0.380	Good
ITEM150	0.269	Difficult	0.412	Excellent

Table 4 revealed that the CTT framework provided estimates for all item parameters of the 150 items that underwent item analysis. The item difficulty indices revealed that there were three (3) things classified as easy (above 0.70), accounting for (2%) of the total. Additionally, there were 90 items classified as intermediate (0.4 - 0.60), representing (60%) of the total. Lastly, there were 57 items classified as tough (0.0 - 0.39), accounting for (38%) of the total. Table 4 shows that the discrimination index findings reveal that 12 items, accounting for 8% of the total, were regarded bad (below 0.19). Additionally, 30 items, representing 20% of the total, were deemed good (ranging from 0.2 to 0.39). The majority of the items, 108 in total, which accounts for 72% of the total, were rated excellent (0.4 and above).

**Research Question 3:** How many of the mathematics aptitude test (MAT) survived under CTT framework?

In order to address the research inquiry, an analysis was conducted on table 5. Items that meet the specified criteria, with item difficulty values between 0.40 and 0.70 and discrimination values more than or equal to 0.20, are considered excellent and chosen. The range of p-values between 0.4 and 0.7 was deemed good according to the research conducted by Obowu-Adutchay in Obagah, et. al. (2014) and Kolawole in Aborisade, et. al. (2021). Additionally, a discrimination value of 0.2 or higher was judged appropriate based on the findings of Metibemu in Asuru, et. al. (2023). Table 5 displays the items that will be removed as determined by the CTT analysis.

**Table 5:** Items deleted and accepted under CTT Framework

Item parameter	Number Deleted	Items Rejected	Items accepted on both difficulty and discrimination
Item Difficulty	60	3, 4, 5, 6, 7, 8, 9, 12, 14, 17, 23, 26, 28, 32, 37, 38, 43, 46, 48, 49, 56, 60, 65, 68, 69, 75, 78, 84, 87, 90, 91, 99, 101, 103, 104, 105, 106, 107, 113, 115, 119, 125, 126, 127, 128, 129, 130, 130, 138, 140, 141, 142, 143, 144, 145, 146, 147, 148, 148, 150.	1, 2, 13, 16, 18, 19, 20, 21, 22, 24, 25, 30, 33, 34, 35, 36, 39, 40, 41, 42, 45, 47, 50, 51, 52, 53, 54, 55, 57, 58, 59, 61, 62, 63, 64, 66, 67, 70, 71, 72, 73, 74, 76, 77, 80, 81, 82, 83, 85, 88, 89, 92, 93, 94, 95, 96, 97, 98, 100, 102, 108, 109, 111, 112, 114, 116, 117, 118, 120, 121, 122, 123, 124, 131, 132, 133, 134, 136, 137, 139.
Item Discrimination	11	4, 7, 10, 11, 15, 27, 29, 31, 44, 65, 69.	
<b>Total for both</b>	<b>68</b>		<b>82</b>

According to the criterion for item deletion, a total of 68 items from table 5 were identified as poor items and designated for removal from the item pool. Here, the selection of excellent things is dependent on the discriminative index and the complexity of the items. However, using the contemporary method for item analysis, the criterion for item selection inside the CTT framework was  $r_{pbs} \geq 0.20$  and  $0.40 \leq p \leq 0.70$ . P represents the level of difficulty of an item, whereas  $r_{pbs}$  represents the indices of item discrimination (point biserial correlation). Consequently, of the 150 things created for the test, 68 items were deemed unsuitable and so discarded, while 82 items were deemed suitable and included in the final exam form.

### Discussion

To answer the first research question, the reliability of the test was determined using three reliability methods, namely; Cronbach alpha method, split-half method and Kuder-Richardson-20. These strategies were suggested for evaluating the dependability of instruments that evaluate the cognitive domain (Asuru, 2015; Onunkwo, 2002; Bloom et al. in Longjohn et al., 2021). The test's internal consistency was measured using three reliability methodologies, resulting in the following coefficients: Cronbach alpha=0.937, split-half=0.956, and Kuder-Richardson-20=0.974. The coefficients above the 0.7 criterion, indicating that the test is trustworthy. Additionally, it is noteworthy that the test exhibited consistent

stability across several reliability methodologies, as shown by the near reliability coefficients.

To answer research question two and three, the item parameter was computed under the classical test theory (CTT) framework and good items were selected based on the difficulty and discriminative index of the items. The components of the test instrument (MAT) were chosen using item analysis. Their difficulty indices and discrimination indices were calculated. Regarding difficulty indices, scholars in the field of measurement and assessment, such as Nworgu (2003), have said that an optimal item should possess a facility index of 0.50. However, in practical scenarios, this index often varies between 0.30 and 0.70. All the components of the instrument fell between the range of 0.30 and 0.70, making them very suited, appropriate, and efficacious. This is analogous to the works of Akpoguma (2008) and Osiobe (2012). The discriminatory indices, which assess the ability of items to differentiate between high-achieving and low-achieving pupils, were also calculated. The discriminatory index of an item ranges from 0.00 to 1.00. Negative indices are considered atypical since they disproportionately punished high-achieving pupils compared to low-achieving kids, resulting in their rejection. Nworgu (2003) concurred that an optimal item should have discriminating indices of +1, however in practical terms it should fall within the range of +0.30 to 1.00. The researchers used a comprehensive set of discerning indices ranging from 0.20 to 1.00 to only include superior items into the instrument. This is analogous to the works of Akpoguma (2008) and Osiobe (2012).

Also, out of the 150 items subjected for analysis, the difficulty ranged from 0.081-0.795, while the discrimination index ranged from -0.125 to 0.744. For this test, the researchers chose 82 items that were acceptable and included in the final exam form. These items had a difficulty range of 0.40-0.693 and a discrimination index ranging from 0.224 to 0.638. The test items have moderate difficulty indices, indicating their strong performance within the CTT framework. Additionally, the items demonstrate high discriminatory power according to CTT.

### **Conclusion**

The findings revealed that out of the total number of goods, 82 (54.6%) were classified as excellent products, whilst 68 (45.4%) were designated for rejection. The mathematical aptitude test (MAT) demonstrated a high level of dependability, with a range of 0.94 to 0.97. The researchers have created the MAT instrument, which is a test with excellent psychometric features in terms of difficulty and discrimination indices. The results demonstrated that the MAT had both validity and reliability.

### **Recommendations**

The study recommended the following:

1. The mathematics aptitude test should be adopted for teaching and learning in measuring students' academic performance in mathematics during promotion examinations from JSS 1 to JSS 2 in junior secondary school in Rivers State.
2. The school boards of the Universal Basic Education (UBE) and the Senior Secondary should provide regular training workshops on test construction for teachers at all levels especially in Mathematics to aid proliferation of quality teacher-made tests.

## References

- Aborisade, O. J. & Ibitoba A. O. (2021). Construction and validation of differential aptitude test for secondary school students in the Southwest Nigeria. *African Journal of Behavioural and Scale Development Research (AJB-SDR)*, 3(1), 45-56.
- Akpoguma, J. (2008). *Construction and validation of mathematics achievement test for senior secondary school students in Delta State*. Unpublished M.Ed. Thesis, Delta State University, Abraka.
- Asuru, V. A. (2015). *Measurement and evaluation in education and psychology (2<sup>nd</sup> ed.)*. Pearl Publishers International Ltd.
- Asuru, V. A., Njigwum, A. S. & Letam, K. (2023). *Development and validation of Basic Science Achievement test using classical test approach: Equipping teachers for global competitiveness*. A paper presented at the 2022 Committee of Deans of Education in Nigerian Universities (CODENU) National Conference held at the Faculty of Education, Ignatius Ajuru University of Education, Port Harcourt, from 11th – 14th January, 2023.
- Awai, D. (2017). *Continuous assessment and teaching practicum in schools*. Spiel Media Concept. p.179.
- Awopeju. O. A. & Afolabi, E. R. (2016). Comparative analysis of classical test theory and item response theory-based item parameter estimates of senior school certificate mathematics examination. *European Scientific Journal*, 12(28), 263-270. <http://dx.doi.org/10.19044/esj.2016.v12n28p263>
- Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests and measurement (9th ed.)*. McGraw-Hill Education.
- Egbule, J. F. (2002). *Principles of psychological testing*. Whyte and Whyte Publishers.
- Fraenkel, J. R., & Wallen, N. E. (2003). *How to design and evaluate research in education*. McGraw Hill Company.
- Inko-Tariah, D. C. & Ogidi, R. C. (2017). *Fundamentals of psychological testing for psychologists, counsellors and educationists*. Rodi Printing and Publishing company.
- Kubiszyn, T. & Borich, G. (2003). *Educational testing and measurement: Classroom application and practice*. John Wiley & Sons.
- Longjohn, I. T., Njigwum, A. S., & Ibeh, C. P. (2021). Validation of basic science achievement test using R-software. *International Journal of Innovative Research and Development*, 10(9), 77-87.
- Miller D. M. Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching (10<sup>th</sup> ed.)*. Pearson Education Upper Saddle River

- Nitko, J. A. (2004). *Educational assessment of students*. Prentice Hall.
- Nworgu, B. G. (2003). *Educational measurement and evaluation, theory and practices*. University Trust Publishers.
- Obowu-Adutchay, V. (2014). Test development. In Obagah, M.O.N & Inko-Tariah, D. C. (Eds). *Educational measurement and evaluation*, (87-107), Rodi Printing and Publishing Company.
- Ogunleye, A. O. (2000). *An introduction to research methods in education and social sciences*. Sunshine International Publication (Nig.) Ltd.
- Onunkwo, G. I. N. (2002). *Fundamentals of educational measurement and evaluation*. Cape publisher International Limited.
- Onunkwo, G. I. N. (2005). *Continuous assessment for Nigerian scholars*. Vigo publishers International.
- Orluwene, G. W. (2012). *Fundamentals of testing and non-testing tools in educational psychology*. Harey Publications Coy.
- Osadebe, P.U. (2001). *Construction and standardization of an economics achievement test for senior secondary school students*. Unpublished Doctoral thesis, University of Port Harcourt.
- Osiobe, G. A. (2012). *Construction and standardization of geography objective test for senior secondary schools in Delta State*. Unpublished Ph.D thesis, Delta State University Abraka.
- Patrick, U. O. & Chukuka, P. N. (2018). Construction and validation of physics aptitude test as an assessment tool for senior secondary school students. *International Journal of Assessment Tools*, 5(3), 461-473.
- Ramsay, T. (2008). *Advantages of multiple-choice aptitude tests*. <http://www.ramsaycorp.com/categories/aptitude.asp>
- Raza, M. A., & Shah, F. A. (2007). Impact of favourite subject towards the scientific aptitude of the students at the elementary level. *Pakistan Journal of Social Sciences*, 31(1), 135-143.
- Reeves, J. F. (2002). *Aptitude assessment*. <Http://Www.Theworksuite.Com/Id15.Html>
- Salkind, N. J., & Rasmussen, K. (2007). *Encyclopaedia of measurement and statistics (Vol. 1)*. Sage Publications.
- Schneider W. J. (2013). Principles of assessment of aptitude and achievement. H. S. Donald, C. R. Reynolds, Schwan, V. (Eds). *The oxford handbook of child psychological assessment (286-330)*. Oxford University Press.
- Ukuije, R. P. I. & Oruluwene, G. W. (2012). *Peanuts educational statistics*. Chadik

printing press.

Usha, G, Vijaya, R. & Rajeshkumar, M. (2022). Construction and validation of mathematics aptitude test for secondary level learners. *Journal of Positive School Psychology*, 6(6), 2883-2887.