WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Development and Validation of Mathematics Achievement Test for Upper Basic School Students in Delta State

[1]Barnabas Othuke UMUOIGHORO, Ph. D; [2] Andrew Moyioritse MEGBELE, Ph.D

[1]Department of Educational Psychology, School of Education, College of Education, Warri, Delta State, Nigeria.
Email: bbfego@gmail.com
[2]Department of Educational Psychology, School of Education, College of Education, Warri, Delta State, Nigeria.
Email: megbeleandrew@gmail.com

ABSTRACT

This study developed and validated Mathematics Achievement Test (MAT) for the assessment of upper basic school students in Delta State. Two research questions guided the study. The instrumentation research design was adopted for the study. The population comprised 72,854 JSS 3 students and a sample size of 1,200 was selected using simple random and cluster sampling techniques. A 150-multiple choice item Mathematics Achievement Test (MAT) was developed by the researchers, which was used as instrument for the study. The data were analysed using Chi-square (X^2), histogram, frequency and percentage. According to the study's findings, 86 of the 150 items had a good fit with the overall model, while 64 did not; the test was distributed well in accordance with the goals of inquiry, intellectual, manipulative, and societal value skills; the test's items were distributed evenly and adequately in accordance with the Upper Basic Education curriculum; and the Mathematics Achievement Test is extremely reliable, as evidenced by its alpha value of 0.96. Additionally, the study found that all 150 items measure the same construct, suggesting unidimensionality; the majority of the test's items are easy and appropriate for the students for whom it was designed, as indicated by the b-parameter model; and the majority of the test's items effectively distinguish between high and low achievers, as indicated by the a-parameter model. Based on the results, it was determined that the Mathematics Achievement Test is a valid and reliable tool for evaluating pupils' mathematical proficiency in Delta State's Upper Basic Schools. Among other things, it was suggested that math instructors utilise the created Mathematics Achievement Test to evaluate Upper Basic School pupils, particularly during practice exams, in order to get them ready for external exams.

Keywords: Development, Validation, Mathematics Achievement Test (MAT), Upper Basic School, Parameter Models

INTRODUCTION

A number of shortcomings in the existing mathematics achievement test have been noted by the academics, including its inconsistency with the state's updated 9-Year Basic Education Curriculum. Additional factors include the use of unsuitable test theories as a guide for developing tests, linguistic ambiguity, violations of unidimensionality, and the introduction of extraneous variables. According to the experts, Delta State's mathematics performance exam from 2019 to 2022 does not align with the core principles of the updated 9-Year Basic Education Curriculum. The curriculum in schools is a living, breathing document that is always evolving to meet the demands, difficulties, and goals of society. Igbokwe (2015) claims that the Nigerian government updated the 9-Year Basic Education Curriculum in 2014 to conform to international best practices, taking into account modern national and international issues as well as input on the curriculum's implementation.

As a result, the updated curriculum has a number of modifications that the previous curriculum did not have. For instance, according to the updated 9-Year Basic Education Curriculum, graduates of 9 years of basic education should possess the necessary levels of literacy, numeracy, manipulative, communication, and life skills in addition to the civic, moral, and ethical values that serve as a foundation for scientific and reflective thinking and as a strong basis for lifelong learning. Researchers found that the present math success exam places a lot of focus on lower level thinking, which will make it difficult for students to meet the requirements of the updated Basic Education Curriculum, which places more emphasis on higher order thinking. The updated curriculum therefore emphasises the development of talents, including work, entrepreneurial, and creative thinking abilities. The curriculum stipulates that pupils must be trained in critical thinking, which will help them meet the educational goals. This may be achieved by helping kids develop the cognitive, investigative, manipulative, and social value abilities that are part of the updated curriculum. Therefore, it is necessary to include the modifications that are part of the updated curriculum into the math achievement exam.

The two main test theories that are often used to influence the creation of test items are item response theory (IRT) and classical test theory (CTT). According to Odili (2016), CTT was the more established of the two and dominated testing and measuring practices until the middle of the 20th century. The theory states that the true score component (T) and the error score component (E) are two orthogonal values that may be separated from the testee's raw score (i.e., the count of the number of right replies) or observed score (X). X = T + E is the mathematical equation that represents it.

The error score component reflects contributions from variables unrelated to the variable or attribute being measured, while the true score indicates the amount of the variable or attribute being measured. Language ambiguity, invigilator announcements, test taker nervousness, and distraction from random elements are some of the causes of mistake variation. Random mistakes in testing are ones that can be removed by sampling. Test takers should be given a lengthy test duration so that the mistakes may balance out, according to proponents of classical test theory, in order to eradicate such inaccuracies.

CTT is criticised due to some shortcomings, despite its relevance and robustness to testing practice. Odili (2016), for example, criticised the use of CTT to determine test takers' performance by comparing their scores to those of other test takers. The test taker who achieves a percentile score of 60 and a percentile rank of 99 in his or her group is considered to be at the top of the class, he said. The student takes advantage of the chance to be at the top of the class when they happen to be in another group with a score of 60 at the percentile rank of 75.

The aforementioned debate has shown that the questions on the mathematics success exams now administered by Delta State instructors do not correspond with the most recent revisions to the basic education curriculum. The quality of a test given by a teacher has a direct impact on its capacity to provide the kind of data required on students' performance. An accurate and consistent assessment of students' comprehension of particular material presented in class is made possible by a well-written exam. Teachers may evaluate the efficacy of their teaching to some degree based on the outcomes of these tests. In contrast, badly crafted test questions may result in imprecise assessments of learning and provide misleading data about students' performance and the efficacy of teaching (Agu, Onyekuba & Anyichie, 2013).

It is not possible to fully attribute the aforementioned issues with the present maths performance assessments on the instructors. This is due to the fact that subject instructors often lack the necessary skills to create effective tests. Some instructors disregard assessment as a distinct activity from their students' teaching and learning processes, neglecting to use the objectives and advantages of assessment in the classroom (Opara and Magnus-Arewa, 2017). According to some educators, the main objective of evaluation is to grade pupils. It seems that most instructors, especially those in Delta State, lack the skills essential to create the tools that are needed for the continuous assessment to function properly. As of right now, mathematics does not have these tools. Students' motivation and performance in mathematics are negatively impacted when poorly designed Mathematics Achievement Tests (MATs) are used. As Opara and Magnus-Arewa (2017) noted, poorly constructed tests may result in inaccurate evaluations and judgements of students' performance on a topic, as well as cause students to lose interest in the subject.

This means that teachers and school administrators will not be able to satisfy the requirements of each individual student and provide educational opportunities if children's accomplishment levels are not accurately assessed and understood. Experts are thus required to create a sufficient number of valid and reliable assessments with sufficient psychometric qualities for use in Delta State junior basic schools. This highlights the need of this work, which aims to first create an accomplishment test in mathematics for instructors to use in mock exams and continuous assessment, and then verify the tool using item response theory.

Statement of the Problem

One of the problems that have bedevilled the educational system of Nigeria and Delta State in particular, is the proliferation of poor achievement test in Mathematics, often used for the assessment of students. For instance, the current Mathematics achievement tests seem to be at variance with the philosophy of the revised curriculum for Basic education which emphasise the development of critical thinking skills among the students. There is therefore the need to include the changes inherent in the revised education curriculum in assessment of learning objectives.

In addition to the above, most of the Mathematics achievement tests currently in use are not in line with current Basic Education Curriculum, which emphasise higher order thinking. Tests that are locally constructed by teachers have been observed to lack the ability to appropriately measure the achievement of students in Mathematics, especially in the higher order thinking skills, to the extent that failure becomes imminent in external examinations often conducted by examination bodies. This is because often times, junior basic school students are prepared for external examinations with the use of poorly constructed test items. This is despite the fact that external examinations are conducted with the use of standardized achievement tests.

For these issues to be addressed in the Mathematics achievement test, there is need for the development of a Mathematics achievement test by a trained psychometrician, with serious emphasis on the possession of the various psychometric properties, the use of appropriate test theories, and ensuring unidimensionality assumption is strictly observed to avoid ambiguity of language or extraneous variables in line with the new curriculum. It is for this reason that this study is aimed to develop a Mathematics achievement test for the assessment of junior basic school students. The problem of this study therefore, is, to what extent is the developed Mathematics achievement test for junior basic schools valid and reliable?

Research Questions

The following research questions guided the study:

1. What level of theta (θ) does the item measure?
2. How do the questions distribute according to the objectives of enquiry, intellectual, manipulative and societal values?

Purpose of the Study

The purpose of this study was to develop and validate a Mathematics Achievement Test for the assessment of junior basic school students. The study specifically:

1. Examined levels of theta (θ) that items measure;
2. Ascertained the distribution of questions according to the objectives of enquiry, intellectual manipulative and societal values;.

Research Design

The instrumentation research design was used in this study. The creation of an instrument for evaluating human behaviour in the three areas of behavior—cognitive, emotional, and psychomotor—is known as an instrumentation research design.

Population of the Study

The population of this study comprised all Upper Basic Education Class III in Delta State. The choice of Upper Basic Education Class III students was because this stage of basic education marks the end of the basic education programme, and it is hoped that at this stage, the students had completed the curriculum of the programme, and they are often assessed based on the objectives of the curriculum. There are 452 Upper Basic Schools and 72,854 students in Delta State in the 2019/2020 academic session.

Sample and Sampling Techniques

The sample size for the study comprised 1,200 Upper Basic 3 students, who were selected from schools across the twenty-five Local Government Areas of Delta State. The choice of 1,200 students was based on the objective of the study and the sample size required to achieve the objectives. This was done through simple random and cluster sampling techniques

Research Instrument

The researchers developed a Mathematics Achievement Test (MAT). The test comprised 150 multiple choice items which were derived from Upper Basic 3 Mathematics Syllabus. The syllabus was obtained from the Ministry of Basic and Secondary Education, Asaba, Delta State. The items in the test comprised 5 options; one correct answer (key) and four incorrect options (distracters).The test was distributed across the objectives of enquiry, intellectual, manipulative and societal values, as stated in the revised curriculum for Upper Basic Schools in Nigeria. They also reflected both the lower and higher order levels of cognitive domain. These included knowledge, comprehension, application, analysis, synthesis and evaluation.

Procedure for Test Development

In developing the test, the following steps as adapted from Osadebe (2016), guided the researchers:

- Planning the test
- Constructing the test
- Initial validation of the test
- The test try-out
- Item Analysis
- Selection of good items
- Reliability

Planning the Test

This stage involved series of activities which the test developers embarked on. These activities include:

- Identifying the behavioural objectives to be measured;
- Identifying the content areas to be covered by the test;
- Deciding on the test format; and
- Designing a table of specification

Identifying the Behavioural Objectives to be Measured: When developing an achievement test, it is very necessary to be guided by the instructional and behavioural objectives that the test will measure. The Mathematics Achievement Test (MAT) that was developed in this study measured the cognitive domain of learning. It included the following six levels of cognitive domain:

1. Knowledge
2. Comprehension
3. Application
4. Analysis
5. Synthesis
6. Evaluation

Identifying the Content Areas to be Covered by the Test: This is the stage where one determines the contents of the test. Each subject or course has its content areas to be taught or examined. In the Mathematics Achievement Test (MAT), the content areas include Numbers and Numeration, Basic Operations, Algebraic Processes, Mensuration and Geometry and Everyday Statistics.

Deciding on the Test Format: In achievement test, there are basically two formats; essay and objective tests. In constructing an achievement test, Opara (2016) suggested that the test developer should make a major decision by determining the kind of items to be included in the test, whether the items should be in an objective format, essay format or both. The number of items to be included in the test should be enough to cover the content adequately and should be based on the levels of the cognitive domain to be measured. When developing the Mathematics Achievement Test (MAT), the researchers used objective test format.

Designing a Table of Specification: This is a very important step in the development of achievement test. According to Osadebe (2016), it is a two-way pattern drawn vertically and horizontally containing the content areas of a subject curriculum and the learning outcomes or behavioural objectives to be measured. The table of specification was used in ascertaining the content validity of Mathematics Achievement Test. It contained a vertical and horizontal parts. The vertical part contained the content (topics or sub-topics) which are to be tested while the horizontal part contained the behavioural process (or objectives) that was to be tested in the content area.

Constructing the Test: In this step, the researchers wrote the test items in the Mathematics Achievement Test (MAT). The researchers constructed a total of 150 multiple choice format objective test items that is in line with the upper basic School Mathematics syllabus.The items were written in line with the specifications before item analysis. In writing the test items, there are certain steps that were suggested by Onunkwo as cited in Opara (2016). The researchers followed the steps as follows:

1. Analysis of the content of most commonly used textbooks.
2. Analysis of the best available courses of study.
3. Analysis of examination questions.
4. Consensus of opinion of competent judges, teachers and others who have studied the material with which the test is concerned.
5. A combination of some or all the above methods.

Instructions were provided to guide the examinees. These instructions included the number of items or questions to be answered and the maximum amount of time allowed for the test should be stated.

Initial Validation of the Test: In its initial stage, the 150 items constructed were face validated by experts in the the field of measurement and evaluation as well as an experienced Mathematics teacher. Their expert observations, comments and suggestions were used in the modifications of the test.

The Test Try-Out: The test was administered on the 1,200 JSS 3 students. The test try-out yielded the data that were used for item analysis.

Item Analysis: The students' responses to the 150 items of the Mathematics Achievement test were subjected to item analysis using Item Response Theory (IRT) to ensure the quality of the items.

Selection of Good Items: Items that passed the quality check of the item analysis were retained in the final test.

Reliability: At this stage, the researchers analysed the items to determine the reliability index of the instrument. This produced the measure of internal consistency of each item in the MAT.

Method of Data Collection

The Mathematics Achievement Test was administered to the students directly by the researchers, with the help of five research assistants. The researchers visited the schools personally prior to the testing date to make his intention known to the principal or head of the school and to obtain permission.

Method of Data Analysis

After the administration of the test, the researchers collated the scores and entered them into a computer for item analysis. Chi-square (X^2)goodness of fit and histogram were used to answer research question 1; frequency and percentage were used to answer research questions 2.

The 150 items of the Mathematics Achievement Test were subjected to IRT psychometric analysis using Statistical Package for Social Sciences (SPSS) version 26 to determine the level of theta each each item measured.

Results and Discussion

In this chapter, the researchers presents the data obtained in the field and discussion was carried out on the findings.

Presentation of Data

Research Question 1: What level of theta (θ) does the item measure?

Table 1: Overall Model Fit of the MAT

S/N	Item	*S-X²*	*df*	P-Value	Remark
1.	MAT48	38.106	81.000	1.000	Good Fit
2.	MAT8	52.904	82.000	0.995	Good Fit
3.	MAT118	59.866	77.000	0.926	Good Fit
4.	MAT150	65.378	83.000	0.923	Good Fit
5.	MAT45	63.219	80.000	0.916	Good Fit
6.	MAT96	67.314	81.000	0.862	Good Fit
7.	MAT85	66.745	80.000	0.855	Good Fit
8.	MAT60	66.378	79.000	0.844	Good Fit
9.	MAT102	68.740	80.000	0.811	Good Fit
10.	MAT121	68.176	79.000	0.802	Good Fit
11.	MAT97	71.867	82.000	0.781	Good Fit
12.	MAT111	70.040	80.000	0.779	Good Fit
13.	MAT14	71.867	80.000	0.730	Good Fit
14.	MAT44	73.883	82.000	0.727	Good Fit
15.	MAT115	72.542	80.000	0.711	Good Fit

16.	MAT126	73.857	81.000	0.701	Good Fit
17.	MAT100	72.221	77.000	0.633	Good Fit
18.	MAT79	73.262	78.000	0.631	Good Fit
19.	MAT127	75.083	79.000	0.604	Good Fit
20.	MAT82	75.280	79.000	0.598	Good Fit
21.	MAT27	75.931	79.000	0.577	Good Fit
22.	MAT145	75.265	78.000	0.567	Good Fit
23.	MAT30	77.319	80.000	0.564	Good Fit
24.	MAT91	79.452	82.000	0.559	Good Fit
25.	MAT110	75.739	77.000	0.519	Good Fit
26.	MAT36	80.466	81.000	0.496	Good Fit
27.	MAT99	80.661	81.000	0.490	Good Fit
28.	MAT135	79.335	79.000	0.468	Good Fit
29.	MAT50	82.113	81.000	0.445	Good Fit
30.	MAT17	80.096	79.000	0.444	Good Fit
31.	MAT18	80.823	79.000	0.422	Good Fit
32.	MAT112	79.170	77.000	0.410	Good Fit
33.	MAT61	85.501	83.000	0.404	Good Fit
34.	MAT98	82.496	80.000	0.402	Good Fit
35.	MAT74	83.661	81.000	0.398	Good Fit
36.	MAT142	80.810	78.000	0.391	Good Fit
37.	MAT29	82.201	79.000	0.381	Good Fit
38.	MAT77	84.050	80.000	0.357	Good Fit
39.	MAT69	84.319	80.000	0.349	Good Fit
40.	MAT7	85.004	80.000	0.330	Good Fit
41.	MAT13	89.502	84.000	0.320	Good Fit
42.	MAT31	86.928	81.000	0.306	Good Fit
43.	MAT113	85.762	79.000	0.282	Good Fit
44.	MAT20	87.325	80.000	0.269	Good Fit
45.	MAT9	86.609	79.000	0.261	Good Fit
46.	MAT51	87.231	79.000	0.246	Good Fit
47.	MAT64	92.748	84.000	0.241	Good Fit
48.	MAT147	86.750	78.000	0.233	Good Fit
49.	MAT49	91.115	82.000	0.230	Good Fit
50.	MAT6	90.136	81.000	0.228	Good Fit

51.	MAT122	88.244	79.000	0.223	Good Fit
52.	MAT10	90.691	81.000	0.216	Good Fit
53.	MAT75	89.681	80.000	0.215	Good Fit
54.	MAT107	91.128	81.000	0.207	Good Fit
55.	MAT109	90.105	80.000	0.206	Good Fit
56.	MAT104	90.365	80.000	0.201	Good Fit
57.	MAT42	88.283	78.000	0.200	Good Fit
58.	MAT93	90.698	80.000	0.194	Good Fit
59.	MAT92	90.962	80.000	0.189	Good Fit
60.	MAT22	90.533	79.000	0.177	Good Fit
61.	MAT94	89.562	78.000	0.175	Good Fit
62.	MAT106	91.838	80.000	0.172	Good Fit
63.	MAT144	91.913	80.000	0.171	Good Fit
64.	MAT33	90.505	78.000	0.158	Good Fit
65.	MAT28	93.283	80.000	0.147	Good Fit
66.	MAT86	91.453	78.000	0.142	Good Fit
67.	MAT65	97.331	83.000	0.135	Good Fit
68.	MAT57	92.224	78.000	0.130	Good Fit
69.	MAT143	91.786	77.000	0.120	Good Fit
70.	MAT26	97.333	82.000	0.119	Good Fit
71.	MAT117	94.072	79.000	0.119	Good Fit
72.	MAT34	98.569	82.000	0.103	Good Fit
73.	MAT46	96.832	80.000	0.097	Good Fit
74.	MAT132	98.054	81.000	0.096	Good Fit
75.	MAT41	97.294	80.000	0.092	Good Fit
76.	MAT136	97.201	79.000	0.081	Good Fit
77.	MAT43	102.934	84.000	0.079	Good Fit
78.	MAT83	98.655	80.000	0.077	Good Fit
79.	MAT120	99.074	80.000	0.073	Good Fit
80.	MAT19	96.936	78.000	0.072	Good Fit
81.	MAT108	98.068	79.000	0.072	Good Fit
82.	MAT68	99.872	80.000	0.066	Good Fit
83.	MAT90	102.187	82.000	0.065	Good Fit
84.	MAT73	100.835	80.000	0.058	Good Fit
85.	MAT131	101.013	80.000	0.056	Good Fit

86.	MAT5	101.590	80.000	0.052	Good Fit
87.	MAT101	99.888	78.000	0.048	Not Good Fit
88.	MAT105	102.210	80.000	0.048	Not Good Fit
89.	MAT66	102.952	80.000	0.043	Not Good Fit
90.	MAT32	108.037	84.000	0.040	Not Good Fit
91.	MAT129	100.006	77.000	0.040	Not Good Fit
92.	MAT149	108.173	83.000	0.033	Not Good Fit
93.	MAT23	103.803	79.000	0.032	Not Good Fit
94.	MAT15	106.607	79.000	0.021	Not Good Fit
95.	MAT81	106.494	79.000	0.021	Not Good Fit
96.	MAT133	108.513	80.000	0.019	Not Good Fit
97.	MAT38	107.842	79.000	0.017	Not Good Fit
98.	MAT123	105.520	77.000	0.017	Not Good Fit
99.	MAT146	110.315	80.000	0.014	Not Good Fit
100.	MAT2	110.718	80.000	0.013	Not Good Fit
101.	MAT139	109.818	79.000	0.013	Not Good Fit
102.	MAT78	110.341	79.000	0.011	Not Good Fit
103.	MAT67	114.454	82.000	0.010	Not Good Fit
104.	MAT138	110.029	78.000	0.010	Not Good Fit
105.	MAT37	114.117	81.000	0.009	Not Good Fit
106.	MAT54	113.408	80.000	0.008	Not Good Fit
107.	MAT80	114.243	80.000	0.007	Not Good Fit
108.	MAT124	114.719	80.000	0.007	Not Good Fit
109.	MAT3	116.775	81.000	0.006	Not Good Fit
110.	MAT35	115.792	80.000	0.006	Not Good Fit
111.	MAT63	119.481	83.000	0.005	Not Good Fit
112.	MAT114	116.815	80.000	0.005	Not Good Fit
113.	MAT47	120.113	80.000	0.003	Not Good Fit
114.	MAT11	122.053	81.000	0.002	Not Good Fit
115.	MAT59	120.852	79.000	0.002	Not Good Fit
116.	MAT84	123.717	81.000	0.002	Not Good Fit
117.	MAT134	121.244	80.000	0.002	Not Good Fit
118.	MAT12	127.586	81.000	0.001	Not Good Fit
119.	MAT71	123.980	77.000	0.001	Not Good Fit
120.	MAT1	189.918	82.000	0.000	Not Good Fit

121.	MAT4	146.132	80.000	0.000	Not Good Fit
122.	MAT16	155.744	77.000	0.000	Not Good Fit
123.	MAT21	185.543	81.000	0.000	Not Good Fit
124.	MAT24	168.437	81.000	0.000	Not Good Fit
125.	MAT25	142.887	84.000	0.000	Not Good Fit
126.	MAT39	251.806	77.000	0.000	Not Good Fit
127.	MAT40	136.982	80.000	0.000	Not Good Fit
128.	MAT52	155.882	80.000	0.000	Not Good Fit
129.	MAT53	473.342	80.000	0.000	Not Good Fit
130.	MAT55	151.090	81.000	0.000	Not Good Fit
131.	MAT56	143.301	77.000	0.000	Not Good Fit
132.	MAT58	189.774	80.000	0.000	Not Good Fit
133.	MAT62	136.428	81.000	0.000	Not Good Fit
134.	MAT70	180.615	81.000	0.000	Not Good Fit
135.	MAT72	153.765	79.000	0.000	Not Good Fit
136.	MAT76	131.370	79.000	0.000	Not Good Fit
137.	MAT87	146.714	81.000	0.000	Not Good Fit
138.	MAT88	150.432	79.000	0.000	Not Good Fit
139.	MAT89	130.776	78.000	0.000	Not Good Fit
140.	MAT95	133.099	80.000	0.000	Not Good Fit
141.	MAT103	245.298	81.000	0.000	Not Good Fit
142.	MAT116	137.675	79.000	0.000	Not Good Fit
143.	MAT119	287.162	79.000	0.000	Not Good Fit
144.	MAT125	131.999	82.000	0.000	Not Good Fit
145.	MAT128	154.885	80.000	0.000	Not Good Fit
146.	MAT130	230.127	81.000	0.000	Not Good Fit
147.	MAT137	198.746	80.000	0.000	Not Good Fit
148.	MAT140	175.304	80.000	0.000	Not Good Fit
149.	MAT141	131.693	78.000	0.000	Not Good Fit
150.	MAT148	186.098	84.000	0.000	Not Good Fit

Table 1 shows the overall model fit of the Mathematics Achievement Test, assessed using chi-square goodness of fit, which was used to estimate the theta level measured by the items of the test. The table shows that the p-value ranged from 0.000 to 1.000. Items with p-value greater than or equal to (≥) 0.05 are regarded as having a good fit in the overall model while items with p-value less than 0.05 are regarded as having no good fit in the overall model. Based on this criterion, out of a total of 150 items, 86 have a good fit in the overall model while 64 did not have a good fit. A total of 86 final items were therefore selected for final administration. To better appreciate the overall model fit of the test, the distribution of the theta estimates for all calibrated items is shown in Figure 1.

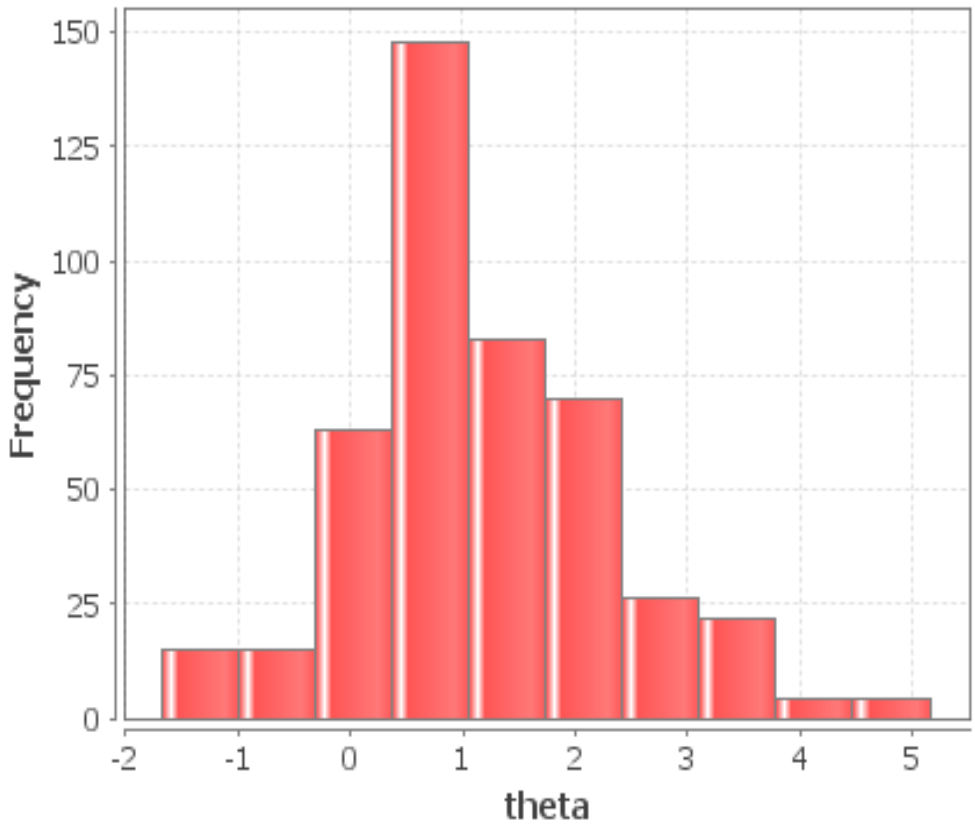

Figure1: Theta Estimates for All Calibrated Items

The histogram shown in figure 3 is a graphical representation of how much information the test is providing at each level of theta. In this case, histogram provides satisfactory information over the ability trait range since it takes the shape of a normal distribution curve.

Research Question 2: How do the questions distribute according to the objectives of enquiry, intellectual, manipulative and societal values?

Table 2: Percentage distribution of questions according to the objectives of enquiry, intellectual, manipulative and societal values

	Enquiry Skill (16%)	**Intellectual Skill (44%)**	**Manipulative Skill (22%)**	**Societal Value Skill (18%)**	**Total 100%**
Numbers and Numeration	0	36	0	0	36
Basic Operations	0	30	0	0	30
Algebraic Processes	24	0	0	0	24
Measuration and Geometry	0	0	33	0	33
Everyday Statistics	0	0	0	27	27
Total	**24**	**66**	**33**	**27**	**150**

Table 2 shows the frequency and percentage distribution of questions in the Mathematics Achievement Test according to the objectives of enquiry, intellectual, manipulative and societal values. The result shows that enquiry skills had a total of 24 items, representing 16%, intellectual skills had a total of 66 items, representing 44%, manipulative skill had a total of 33 items, representing 22% while societal values had a total of 27 items, representing 18% of the 150 items in the Mathematics Achievement Test (MAT).

Discussion of Findings

Level of Theta (θ) Measured by the Mathematics Achievement Test

. The first finding revealed that out of a total of 150 items, 86 have a good fit in the overall model while 64 did not have a good fit. In estimating the ability of the examinee, the Maximum Likelihood Estimation (MLE) was used. The Maximum Likelihood Estimation is a common method for model parameter estimation, sufficiently effective with large sample and valid model application (Longford, 2008). The “likelihood” means probability or possibility, while "maximum" means the highest extent. Therefore, maximum likelihood” is the occurrence with the highest possibility. The highest opportunity will depend on the probability of the correct answers and incorrect answers by the participants, and also on the logistic parameter employed, thus, the determination of maximum ability value is carried out through iteration calculation (Baker, 2001).

The above finding implies that the test generally is within the ability of the testees, which explains why the overall score of the students in the test is high. The testees have a high probability of correctly responding to majority of the items in the test. This finding agrees with the basic premise of IRT, as stated by Thompson (2009) that the probability of a correct or keyed response is a function of an underlying trait or ability, denoted by the Greek letter theta (θ) with a scale typically depicted as ranging from -3 to 3, with 0.0 representing average ability.

The use of Maximum Likelihood Estimation for the development of assessment tools has been recommended and used by several researchers (Odili, 2016; Osadebe, 2016). Egunsola, Denga and Pev (2014), in their study of development and standardization of agricultural science achievement test for senior secondary school students in Taraba State Nigeria, analyzed their test using the Maximum Likelihood estimation technique, and found that the test items had high validity based on one – parameter model of item response theory. The finding is also in line with the study of Oku and Iweka (2018), where item analysis was performed on each item using Maximum Likelihood Estimation Method and found that 99 items of the test fitted the One-Parameter Model (1-PLM). The finding further supports the study of Ani (2014), which maximum likelihood estimation technique of BILOG-MG computer programming to analyse the data generated and found that all 50 test items of Economics survived item calibration. The first finding further agrees with Ene (2017), who used maximum likelihood parameterization procedure to develop and calibrate a Science Achievement Test Using the Two-Parameter Logistic Model of Item Response Theory (IRT), and found that among others that all the item parameter estimates and person parameter estimates were within the acceptable range.

Distribution of Questions According to the Objectives of Enquiry, Intellectual, Manipulative and Societal Values

The second finding revealed that enquiry skills had a total of 24 items, representing 16%, intellectual skills had a total of 66 items, representing 44%, manipulative skill had a total of 33 items, representing 22% while societal values had a total of 27 items, representing 18% of the 150 items in the Mathematics Achievement Test (MAT). What this finding suggests is that the test can measure the objectives of the revised basic education curriculum.

The philosophy of the revised 9-Year Basic Education Curriculum is that every learner who has gone through the 9 years of basic education should have acquired appropriate levels of literacy, numeracy, manipulative, communicative and life skills; as well as the ethical, moral and civic values needed for laying a solid foundation for a life-long learning; as a basis for scientific and reflective thinking. The curriculum requires that students should be developed with critical thinking ability, which will aid in the achievement of the objectives of education. In order to do this, students need to be developed in the area of intellect, enquiry, manipulative and societal value skills inherent in the revised curriculum.

These skills are needed for daily living and the achievement of sustainable development. Before now, tests that measure intellectual, enquiry, manipulative and societal value skills are not available. This Mathematics Achievement Test developed in this study has therefore, filled the gap.

Conclusion

The study revealed that the test generally is within the ability of the testees, which explains why the overall score of the students in the test is high. The testees have a high probability of correctly responding to majority of the items in the test. The study also concluded that the test can measure the objectives of the revised basic education curriculum as the test is in line with the ability level of the students in the appropriate level of schooling.

Recommendations

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

- The developed Mathematics Achievement Test should be used by Mathematics teachers for the assessment of Upper Basic School students, especially during mock examination, in preparation for external examinations;
- The test should be added to the already existing item bank domiciled in the Ministry of Basic and Secondary Education, since the psychometric properties of the test has been shown to be sound.

REFERENCES

Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013.). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research and Reviews*, 8(8), 431-439.

Ani, E. N. (2014). Application of item response theory in the development and validation of multiple choice test in economics. Unpublished Masters dissertation, University of Nigeria, Nsukka

Baker FB (2001). Estimating an Examinee's Ability. In: Boston C, Rudner L, editors. The Basics of Item Response Theory: ERIC Clearinghouse on Assessment and Evaluation. pp. 85-102.

Egunsola, A., Denga, L., & Pev, I. (2014). Development and standardization of agricultural science achievement test for senior secondary school students in Taraba State Nigeria. *Journal of Education and Leadership Development*, 6(2), 72-85.

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.

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

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.

Field, A., (2013), Discovering Statistics Using IBM SPSS Statistics, London, Sage Publications Ltd.

Georgiev, N., (2008) Item analysis of C, D and E series form Raven's Standard Progressive Matrices with Item Response Theory two-parameter logistic model Europe's Journal of Psychology (2008) http://www.ejop.org/archives/2008/08/item_analysis_o.html

Igbokwe, C. O. (2015). Recent curriculum reforms at the basic education level in Nigeria aimed at catching them young to create change. *American Journal of Educational Research*, 3(1), 31-37.

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

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.

Longford, N. (2008). Maximum Likelihood Estimation. In: Casella G, Fienberg S, Olkin I, editors. Studying Human Populations: An Advanced Course in Statistics. New York, NY: Springer New York. pp. 37-66.

Odili, J. N. (2004). *Effect of language manipulation on differential item functioning of Biology multiple choice test.* Unpublished PhD. Thesis, University of Nigeria, Nsukka.

Odili, J. N. (2012). Effect of Simplifying English Language of biology test items on test fairness. *Journal of Educational Assessment in African*, 7.

Odili, J. N. (2016). Language issues in testing in a plural society. *Academic Scholarship Journal*, 13(1), 1-8.

Odili, J. N. (2019). *Refocusing the link between quality education and sustainable national development.* A paper presented at the 5th Annual National Conference of School of Education, College of Education, Agbor, Delta State, 3rd July 2019.

Ojerinde, D., & Ifewulu, B. C. (2012). *Item unidimensionality using 2010 unified tertiary matriculation examination Mathematics pre-test.* A paper presented at the 2012 international conference of IAEA, Kazastan. Retrieved from https://ui.edu.ng/sites/.../PROF%20OJERINDE'S%20LECTURE%20(Autosaved).pdf

Onunkwo, G. I. N. (2002). *Fundamentals of education measurement and evaluation.* Owerri: Cape Publishers Int'l Ltd.

Opara, I. M., & Magnus-Arewa, E. A. (2017). Development and validation of Mathematics achievement test for primary school pupils. *British Journal of Education*, 5(7), 47-57.

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

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: The Danish Institute for Educational Research.

Raven, J, Raven, J.C. & Court, J. H. (1997). Mill Hill Vocabulary Scale: 1998 Edition. Oxford: Oxford Psychologists Press.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. Cognitive Psychology, 41(1), 1-48. Retrieved June 15, 2012, from http://www.wpspublish.com

Thompson, B. (2004). Five methodology errors in educational research: The pantheon of statistical significance and other faux pas. In B. Thompson (Ed.), Advances in social science methodology (pp. 23-86). Stamford, CT: JAI Press.

Thompson, N. A. (2009). Ability Estimation with Item Response Theory. White Paper, Assessment Systems Corporation.

Thorpe, G. L., & Favia, A. (2012). *Data analysis using item response theory methodology: An introduction to selected programmes and applications.* Retrieved from http://digitalcommons.lidrary.umaine.edu/psy_facpub/20

Van der Ven A.H.G.S. & Ellis, J.L., (2000). A Rasch analysis of Raven's Standard Progressive Matrices Personality and Individual Differences, 29 (2000), pp. 45–64.

WPS (2015). Western Psychological Services California. Retrieved from www.wpspublish.com