Estimating the psychometric properties of 50 multiple choice mathematics item for senior secondary II students in Ikwerre Local Government area of Rivers State using classical test theory.

[1]AMADI, Boma Daba [2]OGUNKA, Richard Iheanyichukwu [3]LONGJOHN Ibiene Tandi

Abstract

Test development for assessment is an art that needs accurate estimation of item difficulty, discrimination and chance factor. This study was on estimating psychometric properties of 50 items multiple choice test item in mathematics. Three research questions were stated to guide the findings of the study. The population comprised all senior secondary II students in Ikwerre Local Government Area. Simple random sampling technique was used to select the sample of the study. A total of 900 SS II students were used for the study. A self-constructed multiple choice test in mathematics made up of 50 items was constructed from SSCE Mathematics syllabus, using a table of specification. The reliability of the instrument was determined using Kuder – Richardson formula 20 (KR20) and the reliability coefficient was 0.79. The test was administered to the selected sample and the psychometric properties estimated using Classical Test Theory. Results showed that thirty-five items were to be retained while fifteen items were discarded based on estimate of difficulty index. Two items have high positive discrimination, seventeen items have moderate positive discrimination, while sixteen items are on borderline positive discrimination and fourteen items had zero discrimination and one item has a negative discrimination. The distracter index is forty-two items on negative distracter, thirty items on positive distracter and eight items on zero distracter. Based on these recommendations were made including that test developers should be trained and know how to apply the formula effectively

Keywords: *Difficulty, Discrimination, Distracter and Multiple Choice Questions.*

Introduction

Mathematics is a tool that everybody uses daily to solve practical and abstract problems. It deals with logical reasoning and quantitative calculations. At a basic level, everybody needs to count, multiply, subtract and divide. Mathematics makes it possible for people to count days, weeks, months and years. Every area of mathematics has its own unique applications to the different career options. Thus, Mathematics is one of the compulsory subjects for primary and secondary school learners (Federal Republic of Nigeria. National Policy on Education 2004). Therefore, the minimum requirement to gain admission into most Nigerian universities include at least a credit in each of Mathematics and English language and three other relevant subjects. Over the years, students' Mathematics performance in external examinations has been low. For example, from 2006 to 2015, the percentage of students who passed Mathematics at credit level and above (Al to C6) in May/June West African Secondary School Certificate Examination was as low as *40.3 5%* in 2011 and as high as *63.15%* in 2014. Similarly, a review of the National Examinations Council Senior School Certificate Examination results from 2007 to 2016 reveals that the percentage of students who had credit and above in Mathematics, during the period, were as low as 25.62% in 2010 and as high as 80.76% in 2015. Consequently, many candidates were denied admission into tertiary institutions because of their low Mathematics performance.

Nigerian Educational Research and Development Council (NERDC) (2012) rightly observed, "Poor performance and failure in Mathematics are real in our secondary schools when globally considered". This situation is very disturbing when viewed against the nation's aspirations for scientific and technological advancement. This perennial decline has remained a source of concern to science educators, Mathematicians and Mathematics educators (Nnaka & Anaekwe, 2004). According to Afolabi (2010), the quality of a nation's education determines the quality of

the product of its education system and by extension, the quality and quantity, pace and level of its development. This is probably why every nation tends to invest more in getting her populace educated.

Efforts have been made towards improving performance in Mathematics within the country. The Mathematical Association of Nigeria (MAN) and Science Teachers Association of Nigeria (STAN) collaborated to produce Mathematics textbooks accompanied with students' workbooks to aid teaching and learning of Mathematics in both junior and senior secondary schools. The National Mathematics Centre (NMC) in its attempt to revamp Mathematics teaching and learning in Nigerian secondary schools, initiated the Mathematics Improvement Programme (MIP) project. The MIP project was designed to introduce new teaching methods to enhance students' performance in Mathematics.

Despite all these efforts and the importance accorded to Mathematics as one of the core science subjects and as a compulsory subject to be passed at credit level before admission into tertiary institutions, some students still consider Mathematics as a "monster" while others see teachers of mathematics as "monsters" Issues like the teaching method used by mathematics teachers, much talk about mathematics being the most difficult subject etc, sometimes give rise to misconception, fear, phobia , anxiety and " I cannot do it syndrome" among mathematics students. These issues have exerted a great impact on students' academic achievement in mathematics examination.

NERDC (2012) had also asserted that Psychometric properties of examinations are the internal and/or external attributes inherent in tests upon which an assessment of candidates is based. These properties include the facility and difficulty indices, the discrimination index, the power of distracters, validity and reliability indices. In a mathematics examination the examinees are required to choose one response among set of many alternatives.

A multiple choice item question comprises of a stem, a key and distracters. The stem is the direct question or incomplete statement while the key and the distracters are collectively known as alternatives or options. The key is the correct answer to an item within the alternatives while the distracters are the remaining alternatives that are not the answer (Opara2016). Student's performance is very important to a teacher because it is used to measure outcome. Ukwuije and Opara (2012) suggested that maximum performance measurement is to obtain what a person can do under ideal condition that is determining a person's abilities. Here the testing condition is carefully controlled and the student is told to do his best. A quality testing is a panacea to the issues of measurement, assessment and evaluation; as they provide enough evidence to accurately make decision and improve educational practice and efficiency.

Testing procedures and tools should be reliable within the framework of a test theory or model of application. Testing procedures determines the credibility of the scores provided by items and the validity aids in estimating the abilities of the examinees.

Test developers are concerned about the quality of test items and how examinees respond to them. Writing test items needs a lot of precision. A test item must focus the attention of the examinee on the principle or construct upon which the item is based. Ideally, students who answer a test item incorrectly will do so because their mastery of the principle or construct in focus was inadequate or incomplete. Any characteristics of a test item which distract the examinee from the major point or focus of the item, reduces the effectiveness of that item. Any item answered correctly or incorrectly because of extraneous factor in the item, results in misleading feedback to both the examinee and examiner.

Thus, items in a given test are characterized by (i) discrimination index (a),(ii) difficulty index (b),(iii) distracter index (c).Item difficulty, otherwise known as item facility, is the proportion of examinees who answer a particular item correctly, the item difficulty is an indication of the extent to which an item is difficult to the respondents. An item difficulty

level is determined by estimating the percentage of examinees that get the item right when it is administered. The higher the difficulty index, the easier the item. The lower the index, the more difficulty the item. Discrimination index, closely related to difficulty index, is item discriminative index. Kpolovie *(*2010*)* argued that discriminative index is the extent to which an item accurately differentiates between the best-able and least-able examinees on the entire items that constitute the test .Some items may be found to exhibit negative indices. In this case, the item tends to penalize more of the strong examinees than the weak ones. This is an abnormal behaviour, thus such an item should be re-assessed. The discriminating index is the best rough measure of the extent to which an item discriminates between examinees who earn high scores on the whole test and those who do not. It is thus a measure of whether a given item contributes towards the general direction of the other items of the test. This is, however, based on the assumption that most of the items are contributing significantly towards the efficacy of the test to discriminate between strong and weak examinees. The effectiveness of distracters, an item with an incredible incorrect alternative should not be taken to form the final test. An incredible distracter is that which is so transparently wrong, incorrect or unattractive that virtually none of the examinees ticking the incorrect alternatives opted for it. A test constructor should endeavour to determine how effective the distracter is in attracting responses. An effective distracter is one that distracts those respondents who are not sure of the correct answer. A distracter should thus be effective in reducing guess work.

OBJECTIVE

The objective of this study is to find the psychometric properties of a 50 itemed student's performance in mathematics via estimates of difficulty, discrimination and distracter indexes.

Research Questions

The following research questions were slated to guide the findings of this study.

1. Which items estimate properly thed i f f i c u l t y i n d e x o n s t u d e n t ' s performance in Mathematics?
2. Which items discriminate on student's performance in Mathematics?
3. Which items distracter on student's performance in Mathematics?

METHODOLOGY

Instrumentation research design was used for the study, it involves the development and estimation of difficulty, discrimination and distracter indexes of research instrument in Mathematics.

Multi stage sampling procedure was adopted for the study. Simple random sampling technique was then used to select eight schools out of fourteen schools. Stratified random sampling was also used to draw 900 SS2 Students from the population of 1312 SS2 Students to whom were administered the 50 item instrument. The instrument used for the study was Mathematics Achievement Test (MAT). The MAT was validated by three experts in measurement and evaluation and two subject specialists. The reliability of the instrument was determined using Kuder-Richardson formula 20 (KR20) and the reliability coefficient was 0.79. The student's performance (scores) arranged from upper to lower scores and graded as 31 – 50 upper group and 1 – 30 as lower group. 25% of the multiple choice answer papers were randomly selected from upper and lower group respectively which consist of 36 scripts to estimate the difficulty, discrimination and distracter indexes. Based on the estimation formula, items would be satisfied fit at 0.3 to 0.6, using manual method.

Results.

Research Question 1. Which items estimate properly the difficulty index on student's performance in Mathematics?

S/no.	Upper	Lower	Total	Difficulty Index
1.	12	18	36	0.83
2.	11	4	36	0.41
3.	11	1	36	0.33
4.	14	6	36	0.56
5.	11	5	36	0.44
6.	13	8	36	0.58
7.	13	11	36	0.67
8	9	1	36	0.28
9	12	4	36	0.44
10	15	5	36	0.56
11	17	16	36	0.92
12	9	8	36	0.47
13	10	4	36	0.39
14	7	1	36	0.22
15	12	7	36	0.53
16	14	15	36	0.81
17	13	4	36	0.47
18	16	14	36	0.83
19	16	15	36	0.86
20	13	8	36	0.58
21	18	15	36	0.91
22	10	3	36	0.36
23	8	3	36	0.30
24	8	1	36	0.25
25	17	13	36	0.83
26	10	7	36	0.47
27	14	7	36	0.58
28	16	12	36	0.78
29	15	7	36	0.61
30	14	9	36	0.64
31	11	1	36	0.33
32	16	16	36	0.89
33	15	18	36	0.91
34	16	15	36	0.86
35	15	11	36	0.72
36	16	14	36	0.83
37	9	6	36	0.42
38	16	14	36	0.83
39	13	9	36	0.61
40	15	8	36	0.64
41	15	8	36	0.64
42	18	15	36	0.92
43	15	14	36	0.81
44	16	17	36	0.92
45	17	17	36	0.94
46	17	17	36	0,94
47	17	17	36	0.94
48	17	10	36	0.75
49	17	9	36	0.72
50.	16	10	36	0.72

Table 1b. Difficulty index (B) distribution of items of student's performance in Mathematics.

Items	Items
1,4,6,7,10,11,15,16,18,19,20,21,25,27,28,29,30,32,33, 34,35,36,37,39,40,41,42,43,44,45,46,47,48,49,50	2,3,5,8,9,12,13,14,17,22,23,24,26,31, 38.
35 items retained	15 items poor/not retained.

From the above table, it was observed that 35 items were retained, 18 items not retained from the 50 item students' performance in Mathematics.

Research Question 2. Which items discriminate on student's performance in Mathematics?

Table 2.

S/no.	Upper	Lower	Total	Discriminatory Index
1	12	18	18	1.67
2	11	4	18	0.83
3	11	1	18	0.67
4	14	6	18	1.11
5	11	5	18	0.89
6	13	8	18	1.17
7	13	1	18	1.33
8	9	1	18	0.55
9	12	4	18	0.89
10	15	5	18	1.11
11	17	16	18	1.83
12	9	8	18	0.94
13	10	4	18	0.78
14	7	1	18	0.44
15	12	7	18	1.05
16	14	15	18	1.61
17	13	4	18	0.94
18	16	14	18	1.67
19	16	15	18	1.72
20	13	8	18	1.17

21	18	15	18	1.83
22	10	3	18	0.72
23	8	3	18	0.61
24	8	1	18	0.5
25	17	13	18	1.67
26	10	7	18	0.94
27	14	7	18	1.16
28	16	12	18	1.55
29	15	7	18	1.22
30	14	9	18	1.17
31	11	1	18	0.67
32	16	16	18	2
33	15	18	18	1.83
34	16	15	18	2
35	15	11	18	1.44
36	16	14	18	1.44
37	9	6	18	1.67
38	16	14	18	0.83
39	13	9	18	1.67
40	15	8	18	1.22
41	15	8	18	1.27
42	18	15	18	1.83
43	15	14	18	1.61
44	16	17	18	1.83
45	17	17	18	1.89
46	17	17	18	1.89
47	17	17	18	1.89
48	17	10	18	1.5
49	17	9	18	1.44
50	16	10	18	1.44

Table 2b. Discrimination power (A) of the items on students' performance in Mathematics.

Items on high positive discrimination 2.00 and above	**Items on moderate positive discrimination 1.50-1.99**	**Items on border line positive discrimination** 1.00-1.49	**Items on low to zero positive discrimination** 0.50-1.00	**Items on negative discrimination** 0.00-0.49
2	17	16	14	1

The table above shows that 2 items are on high positive discrimination, 17 items are on moderate positive discrimination, 16 items are on borderline positive discrimination, 14 items are on low to zero positive discrimination and 1 item is on negative discrimination.

Research question 3. Which items distracter on student's performance in Mathematics
TABLE 3a: Estimation of distracter index (c) with 50 multiple choice item of five options ABCDE in student's performance in Mathematics.

ITEM	**A**			**B**			**C**			**D**			**E**			**TOTAL**
	H	**L**	**DI**	**H**	**L**	**DI**	**H**	**L**	**DI**	**H**	**L**	**DI**	**H**	**L**	**DI**	
	2	2	0.00	5	1	0.22	*	*	*				1	1	0.00	18
							*	*	*	2	3	-0.11				18
	1	1	0.00	2	2	0.00	*	*	*							18
	*	*	*	2	1	0.11				0	4	-0.22				18
	0	1	-0.11	*	*	*	1	3	-0.22	0	1	-0.11				18
	0	2	-0.22	2	0	0.22	1	0	0.11	*	*	*				18
				1	1	0.00				0	3	-0.33	*	*	*	18
	2	4	-0.22	2	0	0.22	*	*	*	0	1	-0.11				18
	3	2	0.11	0	1	-0.11				0	1	-0.11	*	*	*	18
	0	1	-0.11	0	2	-0.22	*	*	*	2	1	0.11				18
	0	1	-0.11	1	3	-0.22				*	*	*	0	2	-0.22	18
				*	*	*	0	1	-0.11				1	3	-0.22	18
	0	2	-0.22	1	0	0.11	0	1	-0.11	*	*	*	1	3	-0.22	18
	1	3	-0.22	*	*	*	0	2	-0.22							18
	*	*	*				0	2	-0.22				5	2	0.33	18
	0	3	-0.33				1	0	0.11	*	*	*				18
	0	2	-.022				*	*	*				1	1	0.00	18

*	*		0	1	-0.11	0	2	-0.22				1	0	0.11	18
			2	2	0.00	1	2	-0.11	*	*	*				18
0	1	-0.11	1	4	-0.33	*	*	*	0	1	-0.11				18
1	0	0.11	*	*	*	1	1	0.00	0	2	-0.22				18
0	1	0.11	1	2	-0.11										18
1	1	0.00	1	1	0.00	6	4	0.22				*	*	*	18
0	2	-0.22	2	2	0.00	*	*	*				0	1	-0.11	18
						1	3	-0.22	4	2	0.22	*	*	*	18
*	*	*	1	1	0.00	0	1	-0.11	5	3	0.22	3	4	-0.11	18
6	3	0.33	*	*	*							1	0	0.11	18
0	2	-0.22	*	*	*	5	2	0.33	3	0	0.33				18
*	*	*	0	1	-0.11	0									18
0	1	-0.11				2	3	-0.11	*	*	*				18
0	3	-0.33	6	2	0.44	1	2	-0.11	*	*	*	1	0	0.11	18
5	2	0.33	*	*	*							0	2	-0.22	18
6	2	0.44	0	1	-0.11	0	4	-0.44	*	*	*	1	0	0.11	18
*	*	*	1	3	-0.22							4	2	0.22	18
4	4	0.00	0	1	-0.11	1	1	0.00	*	*	*				18
*	*	*	9	8	0.11										18
*	*	*	2	0	0.22	2	1	0.11							18
			3	3	0.00	0	1	-0.11	*	*	*	1	0	0.11	18
0	2	-0.22				0	1	-0.11				*	*	*	18
0	2	-0.22	0	1	-0.11	*	*	*				0	1	-0.11	18
5	0	0.55	*	*	*							0	1	-0.11	18
									*	*	*				18
			1	0	0.11	*	*	*	0	1	-0.11				18
*	*	*	1	2	-0.11										18
0	1	-0.11	0	1	-0.11							*	*	*	18
0	1	-0.11	0	3	-0.33	*	*	*							18
			2	2	0.00				*	*	*	0	1	-0.11	18
			*	*	*				1	1	0.00				18
0	2	-0.22	0	1	-0.11	3	0	0.33				*	*	*	18
*	*	*	2	1	0.11	0	2	-0.22				0	1	-0.11	18

TABLE 3b: Distracter Index (C) of the Items of Students Performance in Mathematics

Negative distracter item	Positive distracter item	Zero distracter item
2d, 4a, 5a,c,d, 6a, 7d, 8a, d, 9b, d, 10a,b, 11a, b, e, 12c, e, 13a, c, e, 14a, b, 15c, 16a, 17a, 18b, c, 19c, 20a,b, d, 21d, 22b, 24a, e, 25c, 26c, e, 28a, 29b, 30a, c, 31a, c, 32e, 33b, c, 34b, 35b, 38c, 39a, c, 40a, b, e, 41e, 43d, 44b, 45a, b, 46a, b, 47e, 49a, b, 50c, e	1b, 4b, 6b, c, 7b, 8b, 9a, 10a, 13b, 15e, 16c, 17e, 18e, 19b, 21a, c, 22a, 23c, 25d, 26d, 27a, e, 28c, d, 31b, e, 32a, 33a, e, 34e, 36b, 37b, c, 38b, e, 41a, 43b, 49c, 50b	1A, E, 3a, b, 23a, b, 24b, 26b, 35a, c, 47b, 48d

The table above shows that 42 items are not effective distracter,30 items are effective distracter and 8 items has equal distracter.

DISCUSSION

For 50 item Mathematics student's performance, the difficulty index ranges from 0.08 of item 31 to 0.50 of item 42. 35 items were to be retained for the final testing whereas 15 items are poor/not retained. These items should be discarded. This agrees with Toland (2008) that the accuracy of student's error of estimate of b parameter under 3pl depends on the amplitude of the parameter being estimated.

The estimate of the discrimination index shows that it ranges from 0.00 of items 25,26,33,34,38,42,43,48,49 to 0.33 of items 41; 41 items have moderate positive discrimination; 7 items are borderline positive discrimination and 8 items had negative discrimination. All other items belong to low to zero discrimination. The negative discrimination revealed that lower ability students perform better on items 1, 6, 24, 27, 28, 31, 32, and 37 than the higher ability students. The items belonging to low to zero discrimination shows that equal number of higher ability and lower ability students answer the items correctly. The indication is that the items on negative discrimination direction does not discriminate, therefore be discarded (Opara 2016).

The distracter index ranges from 0.00 to 0.40 from the result presented. 42 items are on negative distracter, 30 items on positive distracter and 8 items on zero distracter. The indication reveles that negative distracter attracts higher ability students than low ability students. It further shows that the distracter is not effective. Positive distracter attracts low ability students than higher ability students showing effective distracter. The zero distracter items attracts both the higher ability students and low ability students. It is thus, equally distracted. The relationship between the psychometric properties and its associated estimate could be attributed to factors described by Toland (2008) as test length, underlying ability or other factors. In summary, 35 items were retained based on the consideration of the difficulty index estimation.

CONCLUSION

The findings from the study shows that 35 items from the 50 itemed multiple choice questions on student's performance in Mathematics were considered as having their difficulty index lying within the acceptable limit. It also observed that the estimate depends purely on the psychometric properties. There is a strong relationship between the item defaulting and the trait of estimate item information provided by the difficulty and discrimination index of items.

The estimation of distracter index is determined by the number of item distracter retaining or rejecting an item as a function of the interaction of all the items of the test. Some of the items require a review, to constitute a fairly good item for a final test.

RECOMMENDATIONS

The following recommendations were made

1. That examination bodies should store verified psychometric properties test items in

their computer base bank for sale to users.

2. Test developers should be well trained in the art of estimating difficulty, discrimination and distracter indexes procedure for credible, valid and reliable test item construction.

References

Afolabi, A. O. (2010). Comparison of Private and Public Schools Products Performance in Mathematics and English Language from Educational Perspective. The Primary School Educator, 2(1), 1-6

Federal Republic of Nigeria (2004). National Policy on Education. Lagos; NERDC

Federal Republic of Nigeria (2012). Science, Technology and Innovative Policy. NERDC

Kpolovie, P. J (2010) Advanced Research Methods, Owerri Springfield Publisher ltd.

Nnaka, C.U &Anaekwe, M.C (2004). Application of Research Findings in Science Technology and Mathematics (STM). Education to Enhance Classroom Instruction: The Place of Cooperative Learning Strategy. Science Teachers Association of Nigeria proceedings of the 45th Annual Conference

Opara, I.M (2016). Test Construction andM easurement: Concepts and Applications, Career Publishers, Owerri.

Toland, M.D. (2008). Determining the accuracy of item transactions, 19:3, 1028 – 10 – 29. Retrieved from http://www.rash.org/rmt/rmt193c.htm Trials Based on Item Response Theory, controlled clinical trials, 24 – 390 – 410. Trust Publishers, pp. 66 – 83.

Ukwuije R.P.I and Opara. I.M. (2012). Test and Measurement for Teachers, 3rd edition, Port Harcourt: Chadik Printing Press.