British Journal of Multidisciplinary and Advanced Studies:
Education, Learning, Training & Development, 5(1),180-193, 2024
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Published by the European Centre for Research Training and Development UK

Verification of Dichotomous Scored Junior Secondary Schools Certificate Examination in Mathematics Achievement Test Using Classical Test Theory

Abe, Thomas Olabode (Ph.D)
Counseling Psychology Department,
College of Education, Bamidele Olumilua University of Education Science and Technology Ikere

Olojo, Oludare Jethro (Ph.D)
Department of Science Education, School of Education,
Bamidele Olumilua University of Education, Science and Technology, Ikere Ekiti, Nigeria

Olofin, Samuel Oluwaseyi (Ph.D)
Department of Science Education,Faculty of Education,Ekiti State University, Nigeria

doi: https://doi.org/10.37745/bjmas.2022.0437 Published March 1, 2024

Citation: Abe, T.O., Olojo, O.J., and Olofin, S.O. (2024) Verification of Dichotomous Scored Junior Secondary Schools Certificate Examination in Mathematics Achievement Test Using Classical Test Theory, British *Journal of Multidisciplinary and Advanced Studies*: Education, Learning, Training & Development, 5(1),180-193

ABSTRACT: *In educational assessment there are two major frameworks by which items contained in a test can be examined. These are classical test theory and item response theory. In this study, the author prove the applicability of the classical test theory to the assessment of the quality of multiple- choice junior secondary mathematics test items. Fifty multiple –choice general mathematics test items were administered to four hundred (400) students randomly selected from twenty (20) Junior Secondary schools three (JSS III) students. Ten (10) schools each were purposively selected from private and public schools in Ado local government area of Ekiti state, Nigeria. Three general questions were generated and answered, while two hypotheses were formulated and tested using t-test statistic at 0.05 level significant. Results showed that, out of the 50 items in the test on the basis of difficulty parameter, five (5) items (10%) failed to meet the set criteria for item qualify while on the basis of item discriminating index five items were of poor quality. However, 45 items based on the established standards were considered 'good' items. Finally on the basic of performance, there was no significant differences between students in private and public schools, and also between male and female students from the twenty schools considered for this study which led to the upholding of the two hypotheses formulated for the study at p < 0.0.5. It is recommended that, general mathematics papers, that are to be used to examine students' learning outcomes in Junior School Certificate Examination should be properly validated.*

KEYWORDS: Dichotomous Scored, Achievement Test and Classical Test Theory

British Journal of Multidisciplinary and Advanced Studies:
Education, Learning, Training & Development, 5(1),180-193, 2024
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Published by the European Centre for Research Training and Development UK

INTRODUCTION

Assessment is an indispensable part of educational process. A major purpose of assessment in educational settings is to measure students achievement in order to make a variety of decisions such as promoting them to the next class determining their present level of learning and the extent to which they are ready for next learning experiences. Okpala, Onocha and Oyedeji (1993) while Abe (2002) posits that, Assessment in the context of post compulsory schooling is concerned primarily with making judgments about the extent to which students have achieved outcomes. To assess students' level of cognitive achievement in a given school subject, the two major types of tests are the objective test and essay test. However, there are variants of each, including multiple choice test items for objective test, in which in this study emphasis was on multiple choice test items.

In educational measurements, there are two frameworks by which a test and the items it contains can be studied. These are Classical Test Theory (CTT) and Item Response Theory (IRT). But this paper is set out on the usage of CTT to discuss item discrimination, item analysis and item selection from the perspective of CTT. This is followed by a presentation of an illustrative empirical study using the CTT framework on the tests items of 2021 May/June JSSCE Mathematics paper.

Objective of this Study

The objectives of this study are focus on:

i) The item parameters of JSSCE Mathematics achievement test using the CTT

ii) The test items that are bad in terms of their difficulty; and

iii) The test items that are bad in terms of the discriminating power

Research Questions

To guide the study, the following questions were answered. These are three general:

1) What are the items parameters (difficulty levels and discrimination indices) of the mathematics achievement test using the CTT models

2) On the basis of discriminating indices, which of the items are bad

3) On the basis of difficulty levels, which of the items are bad

Hypothesis

Two hypotheses were formulated for the study from the five questions generated for this study and tested at 0.05 level of significant or Tolerable limits of error.

Ho1: There is no significant difference between the achievement of male and female students when examined with JSSCE.

British Journal of Multidisciplinary and Advanced Studies:
Education, Learning, Training & Development, 5(1),180-193, 2024
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Published by the European Centre for Research Training and Development UK

Ho2: There is no significant difference between the achievement of private and public students when examined with the JSSCE.

Literature Review

Classical Test Theory

This Theory tries to explain the link among the observed score, the true score and the error score within that theoretical framework models of various forms have been formulated by spearman (1904, 1910, 1913), the CTT model is defined as X= T+E, it is a simple linear model linking the observable test score (X) to the sum of two unobservable (or often called latent) variables; that is true score (T) and error score (E). However there are two unknowns in the equation (X and E) and this makes it not easily solvable unless some simplifying assumptions are made.

Boyle and Radocy (1986) and Asmus and Radocy (1992), also opined that X is the observed score, T is the true score and E represents error while within the framework of CTT, true scores are a theoretical concept defined as the mean observed score (i.e expected value) across infinite replications of a measurement procedure Error scores are the difference between the observed score and the true score.

Item Difficulty

This is a measure of individual test question difficulty. It is the proportion of test takers who answered correctly out of the total number of test takers. That is to say, the rate of pool of examinees on an item is used as the index for the item difficulty, symbolically, it is given as item difficulty index p =

$$\text{Index p} = \frac{\text{Number of students who get the item right}}{\text{Total number of student who tried the item}}$$

The difficulty index of an item is computed by dividing the number of pupils passing the item by the total numbers of pupils in the combined upper lower groups i.e p – value = NR/NT x 100% or NW/ NT x 100% ……..(i)

Also, the following formula is used to find difficulty level as

$$\text{P - value i.e difficulty level} = \frac{R_u + R_L}{N_U + N_L} \quad \ldots\ldots\ldots\ldots\text{(ii)}$$

$$\text{Or P - Value} = \frac{U + L}{2N} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(iii)}$$

Where;

R_u = The number of Students in the upper group who responded correctly.

R_L = The number of students in the lower group who responded correctly.

N_u = Number of Students in the upper group

N_L = Number of Students in the lower group

While blue print of the item Difficulty is given below:

0.00 - 0.20 very difficult

0.21 - 0.49 difficult

0.50 - 0.79 middle difficult

0.80 - 0.89 easy

0.90 - 1.00 very easy

Item Discrimination

The ability of an item to discriminate between higher examinees and lower ability examinees is known as item discrimination. There are several methods being used in CTT to assess item discrimination. These include: (a) finding the difference in the preparation of high – achieving and low –achieving Students who scores the item correctly and (b) biserial correlation or point – biserial correlation between a dichotomously scored item and scores in the total test. The use of the difference between the proportion of high – achieving examinees or testees that scored the item correctly and the proportion of low – achieving examines or testees that scored the item correctly necessitates splitting the examinees or testees into two groups. However instead of Splitting the groups into 50-50 usually the 27% or 25% of the two contrasting groups are used (Kelly 1939, Alonge 1989 & 2004 and Courvile, 2004)

$$\text{D Value} = \frac{\text{U 25\% - L 25\%)}}{\text{N 25\%}} \quad \text{----------(4)}$$

While Kelly (1939) and Cureton, (1957) proposed another technique of finding the discriminating power of an item. The technique involves dividing the total examinees or testees into 27% upper and 27% lower to make room for better spread of the distribution.

Then

$$\text{D- Value} = \frac{\text{(U 27\% - L27\%)}}{\text{N 27\%}} \quad \text{…… (5)}$$

This sometimes referred to in many measurement quarters as the estimation of 'α tetra Choric' Kelly (1939) and Cureton (1957).However, D Values range from 1 to -1. A positive index depicts

British Journal of Multidisciplinary and Advanced Studies:
Education, Learning, Training & Development, 5(1),180-193, 2024
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Published by the European Centre for Research Training and Development UK

that α higher proportion of the group answered the item correctly. While α negative item discrimination index indicates that larger proportion of the group answered the item correctly. Although this method gives a fairly stable index of discrimination, it is problematic; in that the process of its computation ignores so much data.

In fact it omits the data of a lot of people (e.g 46% of the examinees information regarding the exact scores in the high - achieving groups and in the low – achieving groups. In other to correct this problem, Hambleton and Jones (1993) posited the point biserial correlation r_{pbi} for item j, a computationally simplified Pearson's between the dichotomously scored item j and the total X

That is;

$$r_{pbi} = \frac{(U_j - U_x)}{\sigma_x} \sqrt{(p_i / q_j)} \quad \text{----------(5)}$$

Where U_j, is the mean total score among examinees who have responded correctly to item j.

U_x is the mean total score for all examinees, P_j is the item difficulty index for item j.

$q_j = (1 - P_j)$ and σ_x is the standard derivation of the examinees total score. This method is better than the later in that in its computation, scores of all the examinees are used.

The above is called Item - Total correlation coefficients which is to correlate performance on the items (scored as either 0 or I) with the total test score. The total test score is usually the total numbers of items answered correctly (Unadjusted) or the total number of items answered correctly omitting the item being examined (adjusted).

Item Analysis and Item Selection

Item analysis within the framework of classical test theory consists of:

1) Determining Sample: Specific parameters and (2) detecting items based on the statistical Criteria set. A poor item is identified by an item difficulty value that is too high ($P > 0.70$) or too low ($P < 0.30$) or a low item discrimination such, that $r_{pbj} \leq 0.20$. However, items are selected on the basis of these characteristics. The choice of level of difficulty and discrimination is usually governed by the purposes of the test and the anticipated ability distribution of the group for whom the test is intended. For example, norm referenced test are developed to differentiate between examinee with regard to their competence in Mathematics. That is, such test is designed to yield a broad range of scores maximizing discrimination among all testes or examinees taking the test. That is to say, items are usually chosen to have a medium level and narrow range of difficulty.

Several studies have been conducted in the area of test item analysis in Nigeria, such as (Abe & Omole 2019, Melibemu, 2015, Enu, 2015, Adegoke, 2013) was in the area of physics, chemistry and General Mathematics at Senior Secondary School Levels. However, Adegoke (2013) opined that the use of CCT Framework is the Simpler and easier to apply than IRT Probably this might

be the reason why its test statistics are still prevalently used in test construction process in Nigeria and indexed in many developing African countries.

RESEARCH METHODS

This study adopted descriptive research design of survey type which involved strictly of verification of dichotomously scored JSSCE Mathematics Achievement Tests using classical Tests Theory. It is a survey type which aimed at collecting data on and describing in a systematic manner the characteristics features or facts about a given population (Champion, 1970, Nworgu 1991 & 2006, Gay 1996, Adeyemi 2007; Abe 2014). The target population for the study consisted of all public and private secondary schools in Ado local government area of Ekiti state totaled fifty eight (58) secondary schools which comprises of thirty two (32) private and twenty six (26) public secondary schools. According to Chang Hanson and Harris (2001), stable estimates of CTT item difficulty and discrimination can be found with a sample size of 150 to 200. Therefore a sample of four hundred (400) students was randomly selected from twenty (20) Junior secondary schools three (JSS III) purposively selected from the fifty eight secondary schools. That is ten (10) schools each were purposively selected from private and public schools.

The author adopted the Junior Mathematics Achievement Tests (JMAT) designed and constructed by the state Ministry of Education. It is assumed that, the items in the JMAT must have been validated by the tests experts in the ministry of education session. However, the researcher discovered there were tests experts with mathematics background in the ministry. While it is pertinent for independent tests experts to assess the psychometric properties of the tests items of public examining bodies such as ministry of education as approved in the 1981 National policy of Education as reviewed in 2004 National policy of Education in Nigeria. More importantly, independent test experts can assess if these items are good and have appropriate difficulty and discriminating indices.

The essence of this is to offer valid suggestions to Ekiti State Ministry of Education on how to improve on the quality of test items that they are to award grades and certificates to students at Junior Secondary level. The (JMAT) comprised of 50 multiple choice items with four answer choices/option (A-D). The test items covered the whole junior Secondary School general Mathematics syllabus. The 50 multiple – choice items Junior Mathematics Achievement Tests for the 2021 Junior School Certificate Examination (JSCE) was administered on the students in the sampled Schools.

The test items were scored dichotomously as either correct or incorrect with correct answer as I and incorrect 0. After scoring the data dichotomously, the classical item analysis of the CAT was carried out. The item psychrometric properties such as item difficulty, item discrimination indices and both construct validity and reliability were determined using the Microsoft EXCEL SPSS and BILOG MG – 3. The dichotomously marked test items were then transferred into Microsoft EXCEL. The data were then converted into ASCII file which was then used for the analysis using BILOG MG – 3 since the emphasis was on CTT framework only phase 1 of the output of the BILOG as examined and t-test statistic was used to test the two hypotheses at $p < 0.05$ level of significance or tolerable limits of error.

British Journal of Multidisciplinary and Advanced Studies:
Education, Learning, Training & Development, 5(1),180-193, 2024
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Published by the European Centre for Research Training and Development UK

RESULTS

The results are presented in the order of general questions and the testing of the hypotheses

Research Question 1: What are the items parameters (difficulty levels and discrimination indices) of the mathematics achievement test using the CTT model?

Table 1 depicts the classical Test Theory model item parameters of JSSCE Mathematics Achievement Test. Numbers of the items in the achievement test (i.e items 1 – 50) and the items statistics of JSSCE Mathematics Achievement Test. Items difficulty (p) and items discrimination (phi) generated by using the SPSS.

Table 1: Items Parameters of JSSCE Mathematics Achievements Test Using CTT Models

Items	Item difficulty (p)	Remark	Items Discrimination (phi)	Remark
1	0.66	Good	0.35	Appropriate
2	0.88	Good	0.32	Appropriate
3	0.88	Good	0.16 *	Inappropriate
4	0.80	Good	0.28	Appropriate
5	0.98	Good	0.28	Appropriate
6	0.84	Good	0.36	Appropriate
7	0.30	Good	0.24	Appropriate
8	0.92	Good	0.28	Appropriate
9	0.54	Good	0.12 *	Inappropriate
10	0.50	Good	0.48	Appropriate
11	0.68	Good	0.32	Appropriate
12	0.14 *	Bad (Low)	-0.04 *	Inappropriate
13	0.38	Good	0.32	Appropriate
14	0.92	Good	0.28	Appropriate
15	0.86	Good	0.44	Appropriate
16	0.60	Good	0.44	Appropriate
17	0.18 *	Bad (Low)	0.24	Appropriate
18	0.42	Good	0.32	Appropriate
19	0.30	Good	0.44	Appropriate
20	0.82	Good	0.36	Appropriate
21	0.94	Good	0.44	Appropriate
22	0.92	Good	0.36	Appropriate
23	0.92	Good	0.40	Appropriate
24	0.52	Good	0.44	Appropriate
25	0.82	Good	0.40	Appropriate
26	0.60	Good	0.28	Appropriate
27	0.40	Good	0.40	Appropriate
28	0.74	Good	0.48	Appropriate
29	0.46	Good	0.48	Appropriate
30	0.78	Good	0.40	Appropriate
31	0.68	Good	0.40	Appropriate
32	0.12 *	Bad (Low)	0.04 *	Inappropriate

33	0.58	Good	0.44	Appropriate
34	0.34	Good	0.44	Appropriate
35	0.18 *	Bad (Low)	0.20	Inappropriate
36	0.94	Good	0.32	Appropriate
37	0.80	Good	0.44	Appropriate
38	0.74	Good	0.48	Appropriate
39	0.90	Good	0.36	Appropriate
40	0.50	Good	0.44	Appropriate
41	0.34	Good	0.28	Appropriate
42	0.70	Good	0.12 *	Inappropriate
43	0.34	Good	0.40	Appropriate
44	0.36	Good	0.48	Appropriate
45	0.30	Good	0.36	Appropriate
46	0.48	Good	0.60	Appropriate
47	0.78	Good	0.36	Appropriate
48	0.24 *	Bad (Low)	0.40	Appropriate
49	0.67	Good	0.40	Appropriate
50	0.76	Good	0.44	Appropriate

Research Question 2: On the basis of difficulty levels, which of the items are bad?

From the table 1 above , on the basis of item selection criteria of difficulty levels of ($0.30 \leq p \leq 0.70$) item 3 with (p = 0.16), item 12 with (p= 0.14), item 17 with (p= 0.18), item 32 with (p= 0.12) and item 35 with (p= 0.18) were bad and very difficult to satisfy the condition.

The blue print strictly adhere to in identifying, whether an item is bad or good is given bellow:

0.0 – 0.20	Very difficult
0.21 – 0.49	Difficult
0.50 – 0.79	Middle difficult
0.80 – 0.89	Easy
0.90 – 1.00	Very easy

Hence items 3, 9, 12, 17, 32 and 35 fell into very difficult or not understood by the students. While items 7 with (p = 0.30), item 13 with (P = 0.38), item 18, with (p = 0.42), item 19 with (p = 0.30), items 15, 16, 11, 9, 24, 34, 37, 40, and 50 with (p = 0.44), item 19 with (p = 0.46), items 34, 41, 43, both have (p = 0.34), item 44, with p = 0.36, item 45 with p = 0.30 and item 46 with (p = 0.48) fell with the categories of difficult items from the table of blue print. While items 1, with (p = 0.66), item 9 with (p = 0.54), item 10 with (p = 0.50), item 11 with (p = 0.68), item 16 with (p = 0.60), item 28 with (p = 0.74), item 30 with (p = 0.78), item 31 with (p = 0.68) item 33 with (p = 0.58), items 38 with (p = 0.74) item 40 with (p = 0.50), item 42 with (p = 0.70), item 47 with (p = 0.78), item 49 with (p = 0.67) and item 50 with (p = 0.26). Were moderately difficult

however, items 2 and 3 have (p = 0.88), items 4 and 37 have p = 0.80), item 6 with (p = 0.88), item 15 with (p = 0.86), and items 20 and 25 have (p = 0.82). They fell into easy items while items 5 with (p = 0.94), items 8, 14, 22, and 23 have (p = 0.92), items 21 and 36 have (p = 0.94), and item 39 with (p = 0.90) were all very easy to solve or check by the students at Junior Secondary level.

Research Question 3: On the basis of discriminating indices which of the items are bad or in appropriate from the table 1, the discriminate index ranges between 0.04 and 0.60 which showed that forty five (45) items were appropriate while only five (5) items were in appropriate in terms of their discriminating index and therefore the need for further modification before they can be useful.

Hypotheses Testing

Ho1: There is no significant difference between the achievement of Male and Female students when examined with the JSSCE.

Table 2: t- test analysis showing the achievement of Male and Female students

Variable	V	Mean	SD	df	tcal	p
Male	204	32.00	6.58	398	1.47	0.298
Female	196	28.80	7.22			

p > 0.05, NS. = Not significant

Table 2 shows mean difference of 3.20, with t-cal value 1.47, which is not significant because the p value of 0.298 was greater than 0.05. This implies that there was no significant difference between male and female students' achievement test, hence the null hypothesis was not rejected at p < 0.05 level of tolerable limit of error.

Ho2: There is no significant difference between the achievement of Private and Public school students when examined with the JSSCE.

Table 3: t – test analysis of students' achievement in Private and Public schools

Variable	N	Mean	S.D	df	t-cal	p
Private	200	28.76	7.87	398	0.99	0.529
Public	200	30.76	6.28			

p > 0.05, NS = Significant

Table 3 shows mean difference of 2.00, with t-cal value 0.99, which is not significant because the p value of 0.529 was greater than 0.05. This implies that there was no significant difference

between private and public schools achievement test, hence the null hypothesis was not rejected at $p < 0.05$ level of tolerable limit of error.

Summary of Major Findings

The analysis of the items contained in JMATS of the difficulty levels of the items ranged between 0.12 (item 32) and 0.94 (item 36). The analysis shows that, only five (5) items out of fifty (50) items were bad and items 12, 17, 32 and 35 were good respectively, similarly, on this basis of discriminating indices, item 3, with item discriminating index of (0.12) that of item 12 discriminating index of (0.04), item 32 with index (0.04) and item 35 with discriminating index of (0.12) were bad. The study also reveal that, only the very high achieving students that could get the correct answers. Also there was no significant difference between achievements of students in both private and public schools, also between female and male students which led to the upholding of the two hypotheses at $p < 0.05$.

DISCUSSION

The following study vividly revealed that, 45 out of the 50 items in the Junior Secondary School Certificate Examination (JSSCE) were appropriate in terms of difficulty indices. While it was also show that 5 out of the 50 items were not appropriate (bad), hence, they will require further modification before they could be used. This implies that JSSCE was very appropriate as used to assess the students' Junior Secondary School Certificate Examination in Ekiti State Ministry of Education in attestation to what is obtainable in the 6,3,3,4 system of Education as stated in National Policy of Education as whether the students are promotable to Senior Secondary School or go to Technical Schools.

Also, the finding equally revealed that, 45 out of the 50 items in the Junior Secondary School Certificate Examination (JSSCE) were appropriate in terms of discriminating index while 5 others need modification in order for them to be useful in the assessment of the students. In the perspective of psychometricians, a goods item should be moderately easy and moderately difficult. Also a good item should discriminate appropriately between high-achievers and low achievers apart from being component of valid and reliable tests.

The implication of this finding was that out of the 50 items that were contained in the JMAT, five items (10%) of the items were bad. These items did not function appropriately. There was every probability that, these items will affect the final score of an average students in the test. When items contained in the tests are good and of inappropriate psychometric properties the scores of the students are likely to be distributed normally, if the items are not good and of in appropriate psychometric properties the distribution of scores of examinees or testes tend to be skewed. The need to analyse test items have been that emphasized time and time again. This is because such analysis shows or reveals the psychometric properties of test items. However it assists in identifying good and bad items. This corroborate the findings of (Adegoke & Isokoya 2018, Fakayode 2017, Metibemeu 2016, Bichi & Bichi 2016, Adegoke & Bichi 2015, Enu 2015, Adegoke 2013). If test items contained in tests are not properly analysed, the results or test scores emanating from such tests will be invalid and unreliable. Therefore analysis of test items contained

in Junior Secondary School Mathematics test examined in this study shows that five items were faulty. Any decision that might have been taken with these items was questionable. This is because these items were difficult and were not in the range of criteria for assessing good items. This is also was in line with the studies of Abe and Omole, (2019) and Adegoke and Osokoye (2018).

Past studies on analysis of dichotomously scored test items using classical test Theory (CTT) and its item response theory (IRT) models have shown that some test items of the public examining bodies could be faulty. For examples in the study carried out by Bichi and Bichi (2016) in their study in analysis of dichotomously score chemistry test of Kano state Qualifying Examination using item Response framework found out that 15 items out of 40 items were bad. Fakayode (2012) while compassing CTT and IRT approaches in equation test scores for WAEC Mathematics test for June and November 2015, found out that 13 items were bad among the June Mathematics test and 15 were bad Items.

Fakayode (2017) also reported these bad item bad inappropriate difficulty and discriminating indices. The items were either too difficult or could not discriminate properly between high achieving students and low achieving students. Also the finding also revealed that at $p < 0.05$ there was no significance difference between;

- the achievement of male and female students when assessed with JSSCE

- there was no significant difference between the achievements of students in private and public schools when assessed with JSSCE hence the two null hypotheses formulated were upheld, this study corroborates with the study of Abe and Omole (2019)

CONCLUSION

Analysis of the test items contained in the JSSCE test examined in the study depicts that five items were faulty. They were bad items that require modification as the items might have been uncleared or misunderstood by the testees. While any decision that might have been taken with these five items was question able. This is because the items were difficult and were not in the range of criteria for assessing good items. However the discriminating indices were considered appropriate for 45 items out of the 50 items while 5 items were found in appropriate and thus do not fit into the set of items for the JSSCE unless it is written for better clarity. While the study showed that students' gender and schools type have no significance difference on their academic achiever in the JSSCE at $p < 0.05$

Recommendations

Based on the findings of this, test developers and psychometricians should endeavour to analyse their test items before they are administered on candidates most especially in national or state examinations of this kind that requires certification for the students to proceed to the next level of education as enshrine in national policy of education 1981 and 1987.

REFERENCES

Abe, T.O (2012) Improving Internal Assessment Scores of JSS Mathematics through Scaling Moderation. *Journal of Counseling and Applied Psychology* 1(1), 89-96

Abe, T.O (2014) The effect of teachers' qualifications on students' performance in Mathematics, *Sky Journal of Educations Research* 2 (1) 10 - 14

Abe, T.O & Gbenro O.S (2014) A comparison of students' attitudinal variable towards Mathematics between Private and Public Secondary Schools. *Journal of Educational Policy and Entrepreneurial research. (JEPER)* 1(1): 32 - 39

Abe, T.O & Omole E.O (2019) Difficulty and Discriminating Indices of Junior Secondary School Mathematics Examination; A case study of Oriade Local Government, Osun State. *American Journal of Education and Information Technology* 3(2), 37 – 46

Adegoke, B.A (2013) Comparison of item statistics of Physics Achievement Test using Classical Test and Item Response Theory frameworks. *Journal of Education and Practice* 4 (22), 87 – 96

Adegoke, B. A & Bichi, A.A (2015) Application of classical test theory to analysis of dichotomously scored chemistry achievement test. *Nigeria Journal of Applied Psychology*, 2(1), 23 – 31

Adegoke, B.A & Osokoya, M.M (2018) Analysis of Dichotomously Scored Further Mathematics of Achievement Test using Classical Test Theory. A Festschrift in honour of professor Michael Olubusuyi Fayemidagba, Published by: Unilorin Mathematics Education Group, University of Ilorin, Ilorin. PP 224 – 232.

Adeyemi, T.O (2004). *Research Methods and these writing in Educational studies*, Lagos: New Haven Pub.

Alonge, M.F (1989). *Measurement and Evaluation in Education and Psychology.* 1st edition, Ado - Ekiti Adebayo publishing and printing press (Nigeria) Ltd.

Alonge M.F (2004) *Measurement and Evaluation in Education and Psychology.* 2nd edition, Ado – Ekiti.

Asmus, E.P & Radocy, R.E (1992) Quantitative Analysis. In R. Colivell (Ed.), *Handbook of research in music teaching and learning* (Pp. 141 – 183). New York. NY: Schirmer books.

Boyle, J.D, & Radocy, E.E (1986) *Measurement and Evaluation of musical experiences* New York: Schirner books.

Champion, D.J (1970) *Basic Statistics for Social Research* Tennessee USA, Chandler Publishing company Pp 286 – 287

Courville, T.R (2004) *An Empirical Comparison of Item response Theory and Classical Test Theory item/person Statistics.* Unpublished doctoral thesis. Texas A & M University.

Cureton, E.E (1950) Validity. In E.F linguist (Ed), *Educational measurement.* Washington D.C: American Council on Education. Texas A & M University.

Enu, V.O (2015) The use of item response theory for these validation and calibration of Mathematics and Geography items of joint command schools promotion Examination (JCSPE) in Nigeria..

Fakayode, O.T (2017) Relative effectiveness of classical test and item response theories in equating West Africa Examinations council Mathematics test scores for June and November 2015. Post field seminar presented at the Institute of Education University of Ibadan.

Federal ministry of Education (1981) *National Policy on Education* Lagos.

Federal Ministry of Education (1987) *National Policy on Education, Lagos.* 3rd Edition, NERDC Yaba, Lagos, 7, 44 – 45.

Gay, L.R (1996) *Educational Research: Competencies for Analysis and Application Fifth Edition.* New Jersey: Merrill, Prentice – Hall Incipper Saddle River, A. Simon & Schuster Company, 249 – 251.

Hambleton. R & Jones, R (1993) Comparison of Classical Test Theory and Item response Theory and their applications to Test Development. *Educational Measurement: Issues and Practices,* 12, 38 – 47.

Kelly, T.L (1939) Selection of Upper and Lower groups for the validation of Test Items. *Journal of Educational Psychology* 30, 17 – 24.

Metibemu, M.A (2016), Comparing Classical Test Theory and Item response Theory in the Development, Scoring and Equating of Secondary School Physics Test in Ondo state. Nigeria, Unpublished Doctoral Thesis. University of Ibadan, Nigeria.

Nworgu, B.G (1991) *Educational Research basic issues and Methodology,* 1st Edition, Ibadan Wisdom Publisher Ltd. Pp 58 – 59.

Nworgu, B.G (2006) *Educational Research basic Issues and Methodology*, 2nd Edition, Nsukka, Enugu by University Trust Publishers Pp 77 – 78.

Okpala P.N, Onacha C.O & Oyedeji O.A (1983). *Measurement in Education,* Jattu Uzairne, Edo – state sterling – Horden Publishers (Nig) Ltd.

Osokoya, M.M (2007). Assessment procedure and students' locus of control as Determinants of achievement in Chemistry. *Journal of International and Society for Teacher Education,* 11(2), 52 – 57.

Spearman (1904). The proof and measurement of association between two things. *American Journal of Psychology,* 15, 72 – 101.

Spearman (1910). Correlation calculated from faulty data. *British Journal of psychology*, 3, 271 – 295 doi : 10.1111/ j. 2044 – 8295. 1910 thoo206.x.

Spearman (1913) Correlations of sum and differences. *British Journal of Psychology,* 5: 417 – 426 doi: 10.1111/ j 2044 – 8295. 1913. Thooo72X.