IJMS 2020 vol. 6 (2): 38 - 57

International Journal of Multidisciplinary Studies (IJMS)

Volume 6, Issue 2, 2019

Comparison of Item Parameter Estimates of Traditional True Score Theory and Modern Mental Theory in Oyo State Junior Secondary School Certificate Mathematics Examination

Asowo A. P* and Afolabi E. R. I
Department of Educational Foundations and Counselling
Obafemi Awolowo University, Ile-Ife, Nigeria

ABSTRACT

The study compared the Traditional True Score Theory (TTST), and Modern Mental Test Theory (MMTT) estimated item parameter indices to the ability of examinees in Junior Secondary School Certificate Examination (JSSCE) in Mathematics. This is with the view of providing empirical justification on the appropriateness of decisions made statistically from psychometric tests. The study adopted a descriptive survey design. A sample of 600 students was randomly selected from a population of 95,419 students who sat for the 2016 JSSCE Mathematics Paper 1 in Oyo State, Nigeria. An adoption version of the 2016 Oyo State JSSCE Mathematics Paper 1 titled Mathematics Test (MT) instrument was utilized for information aggregation. Data collected were analyzed using SPSS and BILOG-MG software packages. The effects indicated that the difficulty indices of Mathematics test items based on TTST ranged from 0.00 to 0.65. The discrimination indices ranged from 0.00 to 0.41. Fifty-two items (86.67%) on the MT items had moderate item difficulty ($0.200 \leq p \leq 0.620$). On the other hand, 42 (70%) on the items discriminate poorly ($0.10 \leq D \leq 0.29$). Regarding MMTT results, item difficulty parameter ranged between 0.216 and 7.988 for 2PLM while the discrimination parameter ranged between 0.100 - 0.729 respectively. Furthermore, there was a negative relationship between the difficulty indices of MMTT and TTST models ($r = -0.702$, $p < .01$). The concomitance correlation for discrimination was positive (0.646, $p <. 01$). The study concluded that the Oyo State 2016 JSSC Mathematics examination was of a moderate psychometric quality irrespective of the theoretical measurement model used in the appraisal.

KEYWORDS: *Keywo Traditional True Score Theory, Modern Mental Test Theory, Item Difficulty, Item Discrimination*

Corresponding author: Asowo, A. P, Email: ayobodeasowo@gmail.com

1 INTRODUCTION

In the recent past, the teaching and learning of mathematics in Nigerian secondary schools have witnessed a serious interference with students' failed efforts at credential levels of instruction. Topmost of this failed performance of students in an external test is essential at the Basic Education Certificate Examinations (BECE) and Senior Secondary Certificate Examinations (SSCE) levels in Nigeria. This failed performance of students' in mathematics certificate examinations in Nigeria, a nation that needs mathematics for its growth, maturation, and productivity, deserves the total attention of measurement experts and psychometricians for a potential reversal. In tests and measurement, more than a few studies have been adopted which have been accurate in predicting students' performance based on their pattern of responding to test items in public examinations at the senior secondary school degree. This is done without paying full attention to the accomplishment of memory address which is supposed to reflect the aims of mathematics curriculum in basic education certificate examinations in Nigeria.

Scores arising from the Junior Secondary School Examination have been hitherto analyzed using the Traditional True Score Theory (TTST) and the Modern Mental Test Theory (MMTT) statistical procedures. Such analyses have focused on item parameter estimates to provide a levelled playground for examinees of equal ability using standardized tests as a precondition for decision making and placement. Among the uses of test for decision making, schools decide who is to be promoted into the next class, external examining bodies decide who is to be certificated as a precondition for higher learning or preferment, higher institutions decide who is to be admitted into the desired course of study and organization agencies decide who is suitable to be employed for the job. Likewise, the essence of using test scores is to assess and evaluate the abilities and competence of individual examinee that is contending for the best junior secondary school grade at the credential level. This is where a levelled playground is achieved by truthfully interpreting examinee test scores as far as possible in making accurate conclusions about test-takers behavior on the exam.

In tests and measurement, to ascertain the quality of an instrument whether it is valid, reliable and consistent over time, an investigation on item characteristics and measurement models are indispensable. By looking into test scores using various measurement models, Anikweze (2010) described the Traditional True Score Model (TTSM) popularly known as the Classical Test Theory (CTT) as the fundamental consciousness of the individual examinee expected observed score (X) that included two major elements that are vital in interpreting the examinee test score as the true score and the error score. The theory emphasizes item calibration through item analysis that would form the basis for shaping examinees' academic achievement. TTST is best utilized in traditional testing situations where diverse groups or individuals that are related by gender, school placement, school type, beliefs, race, ethnicity, or social values formed members of a cluster. For instance, candidates seeking to proceed for a higher level of education, or promotion are administered equivalent sets of test items for placement. Here, true test score would depend largely on the calibration of examinee characteristics on the test, where item difficulty alters would depend on whether the candidates taking the test with assumed equal ability level possess a significant ability on the test or not.

In theory, the magnitude of the TTS model item statistic relies on the ability-distribution of test-takers with the disadvantage of being a sample-dependent test. To simply put it, the TTST provides easier ways of fixing the item characteristics of a test but may not be significant enough in providing exact data about the conduct of the individual examinee on a trial. This implies that the examination agencies set up to control the quality of education at JSSCE level in Nigeria would have to subject the whole instrument to simple item calibration analysis to determine the realness, truthfulness, item difficulty, and discrimination of items on the test. In the same view, the examination agencies set up to control the quality of education at JSSCE level in Nigeria would likewise have to subject the whole instrument to MMT models related to independent item and person statistics of individual examinee scores of the set of items administered for certification. Therefore, the interpretation of score underneath the two frameworks would assist the examination agencies to understand that the normal effect of all examinees that took the test using the TTS model is the same for all test takers on the test irrespective of their ability level, while the MMT models focus on individual test-takers responses to each of the items on the test explaining different ability level on the scale. This implies that differences are most paramount in the calibration analyses underlying each of the theories.

Conversely, Adams and Wieman (2010) explained the Modern Mental Test Theory (MMTT) additionally recognized as the Item Response Theory (IRT) is a modern procedure used in generating a mental response function curve (i.e. the probability of examinees with equal ability level to answer an item correctly on a test) for each item to create a scaled score for the whole test primarily based on what is regarded about each item. Osterlind (2012) elucidated that the modern mental true test score theory is an approach in tests measurement that positioned sets of multifaceted statistics about examinees' cognition and how the appraisal of individual examinee behavior is interpreted. Similarly, Ayanwale (2017) described the mental test score theory as an attempt to model the ability of test-takers with the probability of answering an item correctly based on the pattern of responses to all the items that constitute the test. Theoretically, in tests and measurement, the understanding of the modern mental test theory is based on two basic assumptions. The first assumption positions on how an intelligent person should have the opportunity of performing exceptionally well on the set of items than a less opportune test-taker on the same set of items. The second assumption similarly positions that; any intelligent person should not be careless in getting the items correctly on the scale and continuously be more likely to perform exceptionally well on easier items than on items that are difficult on the test. To effectively estimate the responses of the individual examinee to their ability on a particular test item, items and person parameter estimates are taken into consideration to describe the behavior of each examinee on the test. There are four commonly used MTST models – 1PL/ Rasch, 2PL, 3PL, and 4PL. Each of these models' centers on a probability function that predicts the chance that a person with a higher ability will be able to answer a question with a given difficulty and error variance. Also, for 3PL and 4PL models, there is an increased pseudo-guessing parameter and carelessness parameter respectively. These models according to Baker (2001) provide mathematical equations with the probability of correct response of individual examinee ability that is defined by a particular item representative function (IRF). The one-parameter model also called the Rasch model is only interested in the difficulty level of the items by assuming that all items discriminate equally among the test takers. The two-parameter model assumes that the probability of correct response of examinee should vary in terms of location of items: difficulty (bi) and discrimination of items (ai) with no guessing. It can be expressed as $P(\theta) = \frac{1}{1+\theta^{-a(\theta-b)}}$

Similarly, the three-parameter model or lower asymptote assumes that an examinee can get an item correctly by chance. Hence, the three-parameter logistic model is represented by a – the discrimination power of an item, b – the difficulty of an item and c- the likelihood of guessing making low and high ability learners have the same probability of getting an item correctly. It can be expressed as $P(\theta) = c + (1- c) \frac{1}{1+\theta^{-a(\theta-b)}}$

The four-parameter model referred to as the upper asymptote assumes that high ability students may answer easy items incorrectly due to carelessness or items that are excessively biased. It can be expressed as $P(Xi = 1 | \theta i : a, b, c, d)$ $\frac{1}{1+e^{1.7a(\theta-b)}}$.

(Rupp 2009; Ojerinde, 2013).

As a procedure in tests and measurement, Traditional True Score Theory (TTST) and Modern Mental Test Theory (MMTT) signify two contrasting measurement frameworks for studying test statistics. Although, these theories are generally consistent and complementary with the view of solving the same measurement problems using different methods.

Abedalaziz and Leng (2013) described the relationship that occurred between the TTST and MMTT approaches in analyzing test item characteristics as the item difficulty (p-valve or intercept) and item discrimination parameters (slope). The scholars stated that the TTST and MMTT models can be used to complement each other or separately in describing the behavior of test-item characteristics. The intercept or level of difficulty would allot measurement experts to administer a valid and reliable test with items with the same difficulty level by looking at individual examinees' ability levels that are generated on the continuum. Likewise, item discrimination or slope would allot measurement experts compare the rate to which high/low scores between high/low ability test takers have answered the items on the test correctly based on subject content. Similarly, item analysis as the umbrella name for both parameters according to Adedoyin and Mokobi (2013) is a process that examines students' responses to items on the test and to assess the quality of the test as a whole. However, considering item analysis under the TTST framework and its potential to produce intercept and slope indices for test items in the process of test development. Its ability to detect items that function differently or items that are biased towards more than one groups of examinees is equally imperative. Evidence from test validation could suggest that a test developed under the TTST framework may be fair enough in assessing the proficiency of examinees in mathematics certificate test. Unfortunately, the Traditional Test Score Theory (TTST) is encumbered with inadequacies. Among these are what Fan (1998) summarized as sample dependency of observed scores for item statistics. To the scholar at the item level, the TTST model is relatively simple and does not require a complex theoretical model as the MTST required before relating individual examinee ability to succeed on a particular item. These limitations of the traditional test score theory have made the measurement community shift ground to the development of the modern mental test score theory (MMTST).

Although the MTST is different in the aspect of an item by item analysis from the TTST, many examination agencies in Nigeria, such as the West African Examinations Council (WAEC), National Examinations Council (NECO), National Business and Technical Examination Board (NABTEB) and the Joint Admission and Matriculation Board (JAMB) still use the Traditional Test Score Theory (TTST) for test development and item analysis. Based on this basis, many researchers have empirically assessed the comparison between the MTST and TTST item parameter estimates using different data sets. Some of them include: Fan (1998); Adams and Wieman (2010); Anikweze (2010); Osterlind (2012); Abedalaziz and Leng (2013); Ojerinde (2013); and Metibemu (2016). The scholars found that both TTST and MTST statistics were comparable and could complement each other. Sometimes, the correlation between the intercept of TTST and MTST is high.

The comparison of slopes between both methods could also be large, while the spread across intercept may be insignificant. Despite all scholars' arguments on the comparison of item parameter estimates, few of the studies have considered the use of TTST and MTST frameworks in estimating item parameters of examinees in junior secondary school certificate examination in Nigeria. Therefore, the major focus of this study was to analyze the item characteristics of Oyo State JSSC Mathematics examination using both TTST and MTST methods to check if both methods can be used to complement each other or separately in reflecting the quality of items constructed in the Mathematics test, and whether the models can be used to accurately interpret the true behavior of individual examinee true test score for satisfactory decision making.

The specific objectives of the study were to:

a. examine the item parameter estimates of students' responses to Oyo State Junior Secondary School Certificate (JSSC) Mathematics Examination based on TTST model;
b. examine the item parameter estimates of students' responses to Oyo State JSSC Mathematics Examination based on the MMTT model; and
c. test and compare whether there is a significant relationship exists in the Oyo State item parameter estimates of students' responses to the JSSCE Mathematics Examination based on TTS and MTS Theories.

2 RESEARCH METHODOLOGY

The study adopted a descriptive survey research design. The population comprised all students that sat for the Oyo State junior secondary school certificate (JSSC) Mathematics examination paper 1 (August/September 2016) in Nigeria. A sample of 600 candidates was randomly selected from a total population of 95,419 candidates who took the examination. The instrument for this study was titled "Mathematics Test" (MT). It was an adapted version of the Oyo State August/September (2016) JSSCE Mathematics paper 1. It was a dichotomous four option multiple-choice examination consisting of 60 items that were based on the junior secondary school mathematics curriculum in Nigeria. The instrument for the study was a standardized test developed by the exams section of the department of curriculum and evaluation at the Ministry of Education in Oyo State. The consistency coefficient of students' responses to the 60 multiple-choice mathematics items using Cronbach's Alpha coefficient was found to be 0.734, (n = 600) which was interpreted to be reliable and considered appropriate for the study. Data analyzed were subjected to SPSS and BILOG-MG.

3 RESULTS & DISCUSSION

The Unidimensionality of JSSCE Mathematics

In analyzing item response theory parameter estimates, unidimensionality is one of the basic assumptions that should be considered as it states that one item should measure one construct, whereby other assumptions hold. Unidimensionality is based on the eigenvalues greater than one extracted from the inter-item correlation matrix of multiple-choice items not less than twenty items. An item is said to be unidimensional when a single factor accounts for a significant amount of the total test score variance about individual items. In this study, the Mathematics Test (MT) with 60 multiple-choice items were used to establish the main factor that represents the underscoring construct of the MT items. Factor analysis was used to establish the unidimensionality of the 60 items and the results were presented below. The Cronbach's alpha corroborated the outcome generated by factor analysis with a high internal consistency of 0.80, which indicate that the Oyo State JSSCE Mathematics was unidimensional by measuring only one construct.

Table 1: Total Variance Explained by the Outcome of Factor Analysis

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.715	9.525	9.525	5.715	9.525	9.525
2	2.483	4.139	13.664	2.483	4.139	13.664
3	2.026	3.376	17.040	2.026	3.376	17.040
4	1.600	2.667	19.707	1.600	2.667	19.707
5	1.567	2.612	22.319	1.567	2.612	22.319
6	1.484	2.474	24.792	1.484	2.474	24.792
7	1.457	2.429	27.221	1.457	2.429	27.221
8	1.398	2.330	29.550	1.398	2.330	29.550
9	1.375	2.291	31.842	1.375	2.291	31.842
10	1.328	2.213	34.055	1.328	2.213	34.055
11	1.278	2.129	36.185	1.278	2.129	36.185
12	1.267	2.112	38.296	1.267	2.112	38.296
13	1.227	2.044	40.340	1.227	2.044	40.340
14	1.219	2.032	42.373	1.219	2.032	42.373
15	1.171	1.951	44.324	1.171	1.951	44.324
16	1.163	1.938	46.262	1.163	1.938	46.262
17	1.122	1.870	48.132	1.122	1.870	48.132
18	1.089	1.816	49.948	1.089	1.816	49.948
19	1.083	1.805	51.752	1.083	1.805	51.752
20	1.071	1.784	53.536	1.071	1.784	53.536

21	1.043	1.738	55.275	1.043	1.738	55.275
22	1.022	1.703	56.978	1.022	1.703	56.978
23	.992	1.654	58.632			
24	.965	1.608	60.240			
25	.949	1.581	61.821			
26	.929	1.549	63.370			
27	.926	1.543	64.913			
28	.911	1.518	66.431			
29	.871	1.452	67.883			
30	.848	1.413	69.296			
31	.842	1.403	70.699			
32	.832	1.387	72.086			
33	.820	1.366	73.452			
34	.796	1.326	74.778			
35	.780	1.299	76.078			
36	.774	1.290	77.367			
37	.750	1.250	78.617			
38	.723	1.204	79.822			
39	.719	1.198	81.020			
40	.690	1.151	82.171			
41	.665	1.109	83.279			
42	.652	1.086	84.365			
43	.644	1.074	85.439			
44	.626	1.044	86.483			
45	.596	.993	87.476			
46	.589	.981	88.457			
47	.578	.963	89.420			
48	.564	.941	90.361			
49	.562	.936	91.297			
50	.557	.929	92.226			
51	.540	.901	93.126			
52	.525	.875	94.002			
53	.512	.854	94.856			
54	.491	.819	95.675			
55	.488	.814	96.488			
56	.452	.754	97.242			
57	.448	.747	97.989			
58	.414	.689	98.679			
59	.401	.668	99.347			
60	.392	.653	100.000			

Source: Researcher's Analysis

Outcomes in Table 1 showed the initial value of 5.715, which accounted for 9.525% of the total variance. This could also be explained that the initial value of

5.715 was twice the second value 2.483. The extraction of principal component analysis after the interaction of communalities between the items showed twenty-two components of eigenvalues greater than 1 as revealed in the Scree plot (Figure 1). The interaction of communalities accurately explained 9.525, 4.139, 3.376, 2.667, 2.612, 2.474, 2.429, 2.330, 2.291, 2.213, 2.129, 2.112, 2.044, 2.032, 1.951, 1.938, 1.870, 1.816, 1.805, 1.784, 1.738, and 1.703% of variance accounted for by each component to the total variance in all of the items on the MT. Additionally, for the 60 multiple-choice Mathematics test items, with regards to the eigenvalue greater than 1, the entire percentage variance was 56.978. From the results of PCA, it is apparent that the items were unidimensional because the first factor (9.525) extracted exceeded the second factor (4.139) by a reasonable distance. Also, the scree plot (Figure 1) was examined to confirm the unidimensionality of the Mathematics test.

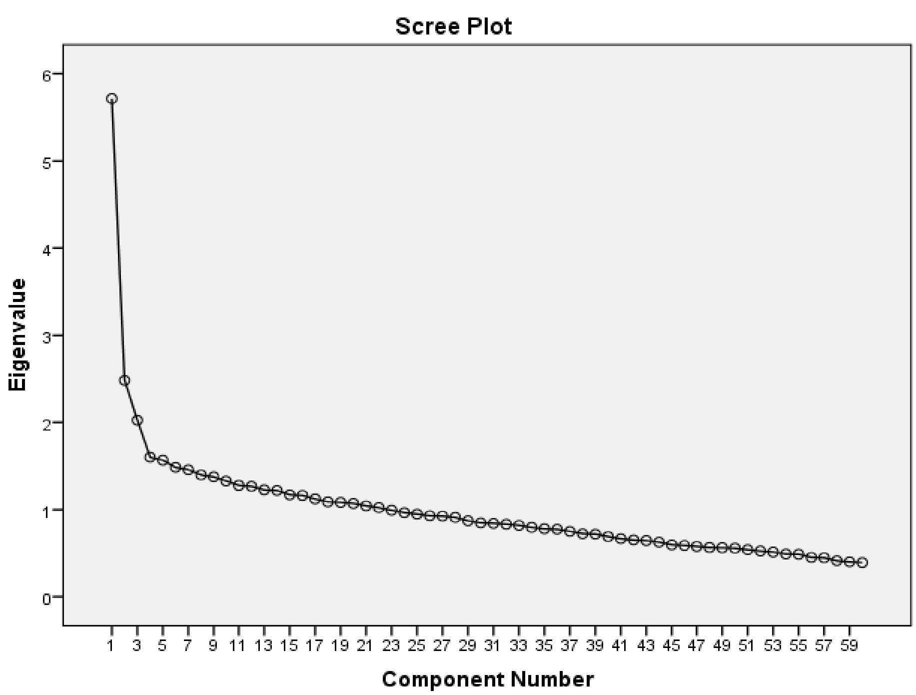

Fig 1: Scree plot of MT

Figure 1 is the scree plot for the MT. The factor analysis that was performed on the mathematics test items using the extraction method of the principal component analysis showed that the distance between the largest component 5.715 and the next component 2.483 is large. The ratio (5.715 ÷2.483 = 2.302) of the first components was large enough and substantially greater than one. On the model, there were twenty-two factors greater than eigenvalue of one than the other extracted communality factor. This showed how one - dimensional the test is considering the first factor higher than the other factor. The results showed that the 60 multiple-choice mathematics items fulfil the assumption of unidimensionality as factor analysis results were in line with the set of conditions used for assessing unidimensionality proposed by McBride & Weiss (1974), Orlando, Sherbonve and Thissen (2001). According to the scholars, dichotomously scored items are one-dimensional when the first loading for all items is significantly greater than one. Hereafter, the test for unidimensionality confirms the realness and consistency of the test constructed by the Oyo State Ministry of Education in Nigeria.

Research Objective One: examine the item parameter estimates of students' responses to Oyo State Junior Secondary School Certificate (JSSC) Mathematics Examination based on TTST model.

To answer this research objective, students' responses were scored using the key of "1" for a correct answer and "0" for a wrong answer to the items on the MT. The scores for each respondent to the 60 items were further analyzed using traditional test score theory item statistics to determine both intercept and slope indices of the test as well as their respective item-total correlation coefficients. The results were summarized in Table 2.

Table 2: Intercept index (P) of 2016 Oyo State JSSCE Mathematics Test Items

Intercept Index (P)	Items	Decision
$P \geq 0.90$	Nil	Very easier items
$0.20 \leq P \leq 0.62$	1-12, 14-19,21,23-26,28-32,34-52,55,57-60	Ideal items
$P < 0.20$	13, 20, 22, 27, 33, 53, 54, 56	Very difficult items

Source: Instructional Assessment Resources (IAR 2011)

Table 2 revealed that in 2016 Oyo State JSSCE Mathematics test items, 86.7% (52 items) were acceptably good because the items met the ideal standard of intercept index ranging between 0.20 – 0.62 while 13.3% (8 items) on the MT were not acceptable because they were very difficult items considering their intercept index below 0.20. However, no item on the MT is very easy, given its slope index to be above 0.90. These imply that most of the items on the MT under the TTS model had the ideal difficulty index. The result implies that the difficulty level of the students is constant which makes it challenging to differentiate between low and high ability learners who have mastered the concerning content well.

.

Table 3: Slope index (d) of 2016 Oyo State JSSCE Mathematics Test Items under TTST

Slope Index (D)	Items	Decision
D > 0.40	24, 53	Very Good items
$0.30 \leq D \leq 0.39$	2, 4, 9, 11, 16, 17, 19, 23, 30, 38, 39, 45, 46, 58	Good items
$0.10 \leq D \leq 0.29$	1, 3, 5 - 8, 10, 12, 14, 15, 20, 25, 26, 28, 29, 31, 32, 34 - 37, 41 - 43, 48 - 51, 56, 57, 59, 60	Poor items
$0.01 \leq D \leq 0.10$	13, 21, 22, 27, 40, 44, 47, 52, 54, 55	Very Poor items
Negative	18, 33	Ambiguous

Source: Ebel (1965) in Ovwigho (2013)

Table 3 revealed that 3.39% (2 items) were very good, 23.33% (14 items) were good. However, 53.33% (32 items) were poor, 16.67% (10 items) were very poor, and 3.39% (2 items) had a negative slope. This implies that there were more easy items on the test which made it difficult to differentiate between low and high ability students.

Tables 2 and 3, therefore, showed that the item parameter estimates of Oyo State JSSC Mathematics examination using TTST model were moderately difficult (item difficulty). This implies that the items on the test are neither too difficult nor too easy for the students. The results under the traditional test score theory are that the test fluctuated between students who have without a doubt learned the subject content and students who have not. It indicates that the test could not explain how high and low ability students behave on the test (item discrimination) accurately.

Research Objective Two: examine the item parameter estimates of students' responses to Oyo State JSSC Mathematics Examination based on MMTT model

The MTST analysis was conducted by means of BILOG-MG 3.0. All MMTT parameter estimations were obtained using the marginal maximum likelihood (MML) method with normal past distribution, which is a default for BILOG-MG.

Table 4: Intercept Index (b) of 2016 Oyo State JSSCE MT Items for Two PLM

Intercept Indices	Items	Decision
< 0.20	1, 12, 15, 45	Very Difficult items
0.20 – 0.69	2 – 6, 9, 11, 14,16, 21, 23, 24, 46	Medium items/ ideal items
0.70 - 0.90	7, 8, 10, 13, 17, 19, 20, 22, 25 – 44, 47 – 60	Easy items

Source: Instructional Assessment Resources (IAR 2011)

Table 4 showed that 6.78% (4 items) was considered difficult, 22.03% (13 items) were considered ideal, while 71.19% (42 items) were easy items. This implies that there were easier items on the test. The results imply that the test did not differentiate between students that have been taught the subject content well and students who have not.

Table 5: Discrimination index (a) of 2016 Oyo State JSSCE MT Items for 2PLM

Item Slope	Items	Decision
0.01 – 0.34	6 -10, 13, 14, 19, 21, 25 - 28, 31, 33 - 44, 47, 49, 51, 52, 54 - 56, 58, 60	Poor items
		Good items
0.35 – 0.64	1 - 5, 9, 11, 12, 15, 17, 20, 22, 23, 29, 30, 32, 39, 45, 46, 48, 50, 53, 59,	
0.65 - 1.34	Nil	Moderate
1.35 – 1.69	16, 24,	High items
>1.70	Nil	Very high items

Source: Instructional Assessment Resources (IAR 2011)

Table 5 showed that 3.38% (2 items) had high slope index, 38.33% (23 items) had a good slope index, 57.63% (34 items) had very poor slope index. This implies that the test might have increased the performance of low ability student considering the test poor slope index. The result implies that the mental test score theory approach reflected the behavior of individual students on the test. It showed how high ability learners were careless in responding to easy items and how low ability students had the advantage of guessing the correct answers right on difficult items.

Tables 4 and 5, therefore, showed that the item parameter estimates of Oyo State JSSC Mathematics Examination using MMTT model are easy items (item difficulty) and moderately discriminated well among low-and high-level ability students (item discrimination) due to other irrelevant factors related to the construct of the test. The result implies that the mental test score theory is efficient and adequate in explaining how high and low ability students behave on the test as they move from item to items along the scale.

Research Objective Three: test and compare whether there is a significant relationship exists in Oyo State item parameter estimates of students' responses to the JSSCE Mathematics Examination based on TTS and MTS Theories.

To answer this research objective, item statistics obtained from both TTST and MTST analysis were pooled together to test whether the two models are comparable and can be used interchangeably. Also, the Pearson correlation coefficient was adopted in testing the relationship between item intercept and slope indices obtained based on TTST and MTST.

Table 6 presents the item statistics of both TTS and MMT theories. The left-hand phase gives the traditional item statistics (intercept p), and slope (d) obtained from SPSS version 20. The right hand gives the slope (a) and intercept (b) parameters of the mental test score theory obtained from phase 2 of BILOG-MG. Table 6 also showed that the b-values increases while p-values decreases. This implies that there is an inverse significant relationship between item intercept of the MTS model and p-value of the TTS model.

Table 6: Items are grouped according to various discrimination levels based on MTST (a, b) and TTST (p, d) statistics.

	TTST Statistics		**MMTT 2PLM**	
Item No	**p- values**	**d- values**	**a – values**	**b – values**
1	0.21	0.20	0.554	0.158
2	0.52	0.36	0.637	1.692
3	0.44	0.27	0.487	-0.216
4	0.50	0.31	0.440	0.502
5	0.6	0.18	0.394	-0.616
6	0.31	0.18	0.319	1.696
7	0.21	0.16	0.309	2.565
8	0.33	0.17	0.212	2.079
9	0.47	0.32	0.432	0.335
10	0.23	0.17	0.314	2.663
11	0.48	0.32	0.481	0.272
12	0.48	0.29	0.406	0.197
13	0.00	0.00	0.342	1.979
14	0.38	0.18	0.200	1.473
15	0.37	0.21	0.351	1.097
16	0.41	0.38	0.729	0.464
17	0.28	0.32	0.510	1.388
18	0.20	-0.10	-	-
19	0.38	0.32	0.500	0.818
20	0.02	0.18	0.214	2.526
21	0.38	0.03	0.611	0.652
22	0.14	0.07	0.368	3.215
23	0.48	0.36	0.477	0.209

24	0.45	0.41	0.706	0.433
25	0.39	0.14	0.260	1.263
26	0.24	0.10	0.223	3.388
27	0.18	0.05	0.184	4.918
28	0.31	0.25	0.291	1.723
39	0.65	0.23	0.417	-0.860
30	0.35	0.35	0.551	0.834
31	0.20	0.14	0.248	2.948
32	0.36	0.20	0.354	1.089
33	0.10	-0.02	0.161	7.988
34	0.35	0.17	0.259	1.580
35	0.35	0.24	0.342	1.165
36	0.24	0.21	0.318	2.385
37	0.36	0.27	0.347	1.043
38	0.36	0.31	0.316	1.209
30	0.29	0.33	0.466	1.410
40	0.39	0.02	0.258	1.240
41	0.22	0.15	0.206	3.222
42	0.33	0.23	0.282	1.572
43	0.29	0.17	0.345	1.803
44	0.35	0.09	0.100	3.222
45	0.51	0.34	0.363	0.013
46	0.43	0.33	0.440	0.440
47	0.34	0.07	0.122	3.473
48	0.31	0.29	0.473	1.126
49	0.20	0.10	0.153	4.965
50	0.22	0.24	0.504	1.880
51	0.30	0.15	0.121	3.503
52	0.37	0.03	0.339	1.147
53	0.02	0.30	0.587	1..517
54	0.18	0.09	0.354	1.484
55	0.32	0.08	0.148	3.155
56	0.16	0.11	0.230	4.569
57	0.25	0.25	0.211	3.123
58	0.40	0.31	0.317	0.935
59	0.30	0.23	0.358	1.614
60	0.33	0.23	0.329	1.510

Source: Researcher' Analysis

Table 7: The relationship between item intercept and item slope of MTS and TTS Theories

Correlation Coefficient of TTST and MMTT Item Intercept	Correlation Coefficient of TTST and MMTT Item Slope
-0.702	0.646

** Correlation was significant at 0.01 level (2-tailed)

It was presumed from Table 6 that there was a negative significant relationship (r = -0.702) exists between b- intercept index of TTST and p – intercept index of TTST as reflected in Table 7. The table also showed that the b-values increases while p-values decreases, which is statistically significant ($p < .01$), it can be inferred that a negative significant relationship exists between the two indices. Likewise, from Table 7, there was a simultaneous positive relationship between the slope of MTST and TTST models. This indicated that as a-values increases, d - values is also increasing. This implies that the association of relationship between the 'a' (intercept) and the d-values (slope) is high and positive (r = 0.646) which showed that the relationship between the models is significant ($p < .01$). The result implies that both TTST and MTST are related to each other and can be used interchangeably, but the mental test score theory is more suitable in reflecting the true ability of students in a test.

Discussion:

From the study, it is understandable that the analysis of the 2016 Oyo State JSSCE dichotomously scored items is one - dimensional using factor analysis and scree plot methods. It was established that the second eigenvalue clearly exceeded the second factor of the sixty items with a high correlation coefficient that indicated unidimensionality. The results revealed that the 60 multiple-choice mathematics items fulfil the assumption of unidimensionality as the model factors were in line with the set of conditions used for assessing unidimensionality by Mcbride and Weiss (1974), Orlando, Sherbonve and Thissen (2001). Therefore, with the results from factor analysis, the assumption holds to a good extent that items on the MT are unidimensional and moderately valid.

Specifically, findings of research objective one showed that the item parameter estimates of students' responses to Oyo State JSSC Mathematics examination based totally on the TTST model are moderately difficult (item difficulty) however could no longer discriminate well among the students (item discrimination). This means that high ability examinees may have a lower chance of answering an item correctly on the test which may be due to carelessness, prejudice, or confusing items on the test, and low ability examinees may have a higher chance of answering an item correctly on the test which may be due to guessing, or testwiseness. This could also be as a result of sample dependency of the test items or maybe the test has been poorly developed by test developers.

Furthermore, findings from research objective two showed that item parameter estimates of students' responses in Oyo State JSSC Mathematics Examination based on the MTST model are easy items (item difficulty) and are well discriminated among low- and high-level ability students (item discrimination). This indicated that the items were able to function well among low ability examinees which are a disadvantage to high ability learners because if test items are extremely difficult, low ability examinees would be unlikely to answer the items correctly, or if the items are too easy, most examinees including those with low ability level would have a moderate chance of answering the items correctly, but such items are often needed to adequately achieve the objectives of the course content. Also, this finding laid credibility on the observation of test experts such as Hambleton and Jones (1993) and Ojerinde (2013) that despite the popularity of traditional item statistics as an integral part of standardized test and measurement technology, it is encumbered with so many limitations such as examinee and item sample dependency. This made the estimates of TTST not generalizable across populations.

The findings further showed that TTST-based discrimination index is comparable with the MTST–based discrimination parameter. It showed that a high and positive relationship exists between item discrimination of MTS and TTS models. This means that TTST and MTST can be used interchangeably in item calibration for students' assessment. This supports the findings of Adegoke (2013) that the association between a-values and point biserial correlation should be high and positive. Lastly, the result showed that as b-values increases, the p-values decreases, which is statistically significant ($p < .01$), it can be inferred that a negative significant relationship exists between the two indices. This study is in line with Stages (2003) and Wiberg (2007) whose findings revealed that the association between p-values of TTST and b-parameter of MTST were high and negative.

These findings were consistent with the earlier studies of Fan 1998; Adedoyin, Nenty and Chilisa 2008; Nukhet 2002; MacDonald and Paunonen 2002; Courville 2005; Osterlind 2012; Abedalaziz, and Leng 2013; Ojerinde 2013 and Metibemu 2016 that item statistics from the two contrasting frameworks are quite related and MTST can be used to complement each other or used separately.

4 CONCLUSION & RECOMMENDATIONS

Based on the findings, the study showed that the 2016 Oyo State Mathematics test is unidimensional and moderately valid. The study also showed that item parameter estimates of students' responses in Oyo State JSSC Mathematics examination based on the TTST model are moderately difficult (item difficulty) but could not discriminate well among the students (item discrimination). The study showed that item parameter estimates of students' responses in Oyo State JSSC Mathematics examination based on the MTST model are easy items (item difficulty) and discriminated well among low- and high-level ability students (item discrimination). The study further showed that a high and positive relationship exists between item discrimination of MTS and TTS models. Finally, the study showed that there is a significant relationship between the b and p values. The study, therefore, concluded that the TTTST and the MMTTS theory item parameter estimates were comparable and can be used to complement each other or used separately. The study as a consequence recommended that examination agencies in Nigeria and Oyo State ministry of education must embrace the use of both the TTST and MTST approaches in test development and item analysis to have high-quality test items with reliable and valid results.

REFERENCES

Abedalaziz, NCH. & Leng, 2013, 'The Relationship between ctt and irt approaches in analyzing item characteristics', *The Malaysian Online Journal of Educational Sciences*, vol. 1 no. 1. pp. 52-58.

Adams, WK & Wieman, CE 2010, 'Development and validation of instruments to measure learning of expert-like thinking', *International Journal of Science Education*', vol. 33, no 9, pp. 1289 -1312.

Adedoyin, OO & Mokobi, T 2013, 'Using irt psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items', *International Journal of Asian Social Science*, vol. 3 no. 4, pp. 992-1011.

Adedoyin, OO, Nenty, HJ & Chilisa, B 2008, 'Investigating the invariance of item difficulty parameter estimates based on ctt and irt', *Educational Research and Review, vol.* 3, no. 2, pp. 83-93.

Adegoke, BA 2013, 'Comparison of item statistics of physics achievement test using classical test and item response theory frameworks', *Journal of Education and Practice*, vol. 4 no. 22, pp. 87-96.

Anikweze, CM 2010, '*Measurement and evaluation: for teacher education',* 2nd edn, SNAAP Press Ltd, Enugu.

Ayanwale, MA 2017, Efficacy of item response theory in score ranking and concurrent validity of dichotomous and polytomous response mathematics achievement test in Osun State, Nigeria, PhD thesis, Institute of Education. University of Ibadan.

Baker, FB 2001, 'The basic of item response theory: test calibration', *ERIC Clearing House on Assessment and Evaluation*, University of Maryland, College Park, MD, pp. 136-330.

Courville, TR 2005, An empirical comparison of item response theory and classical test theory item/person statistics, PhD thesis, Texas A & M University.

Ebel, RL 1965, 'Measuring educational achievement content standard test scores', *Educational and Psychological Measurement, no. 12, pp. 15-25.*

Fan, X 1998, 'Item response theory and classical test theory: an empirical comparison of their item/person statistics', *Educational and Psychological Measurement*, no. 58, pp. 357-385.

Hambleton, R & Jones, R 1993, 'Comparison of classical test theory and item response theory and their applications to test development', *Educational Measurement: issues and Practice*, vol. 12, no. 3, pp. 38-47.

Instructional Assessment Resources 2011, *'Item analysis*', Retrieved November 9, 2013 from University of Texas at Austin, Instructional Assessment Resources, IAR. Available from: http://www.utexas.edu/academic/ctl/assessment/iar/students/report/itemanalysis.php.

McBride, JR & Weiss, DJ 1974, '*A word knowledge item pool for adaptive ability measurement',* Research report no. 74, pp. 2- 37. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods programs, (AD 781894).

McDonald, RP & Paunonen, S 2002, 'A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory', *Journal of Educational and Psychological Measurement*, no. 62, pp. 921-943.

Metibemu, MA 2016, Comparison of classical test theory and item response theory in the development and scoring of senior secondary school physics tests in Ondo State, PhD thesis. Institute of Education. University of Ibadan.

Nukhet, C 2002, 'A study of raven standard progressive matrices test's item measures under classic and item response models: an empirical comparison', *Journal of Faculty of Educational Science*, vol. 35 no. 1- 2, pp. 71-79.

Ojerinde, D 2013, '*Classical test theory (ctt) vs item response theory (irt): an evaluation of the comparability of item analysis results*', A Guest lecture presented at the institute of Education, University of Ibadan on 23rd May.

Osterlind, SJ & Wang, Z 2012, *'Item response theory in measurement, assessment and evaluation for higher education.* In C. Secolsky & D. B. Denison eds., handbook on measurement, assessment and evaluation*',* pp. 208-215, New York, NY: Routledge.

Orlando, M, Sherbouve, CD & Thissen, D 2001, 'Summed- score linking using item response theory: application to depression measurement', *Psychological Assessment*, vol. 12 no. 3, pp. 354-359.

Ovwigho, BO 2013, 'Empirical demonstration of techniques for computing the discrimination power of a dichotomous item response test', *Journal of Research and Method in Education*, vol. 3 no. 2, pp. 12-17.

Rupp, AA 2009, 'Item response theory modeling with bilog-mg and multilog for windows', *International Journal of Testing*, vol. 3 no. 4, pp. 365-384.

Stage, C 2003, '*Classical test theory or item response theory: the Swedish Experience',* (No. 42). Kluwer Academic Publisher, Umea.

Wiberg, M 2007, 'Classical test theory vs. item response theory: an evaluation of the theory test in the Swedish driving-license Test*', Educational Measurement (No. 50). ISSN 1103-2685, Umea universities: Kluwer Academic Publications*