World Journal of Innovative Research (WJIR)
ISSN: 2454-8236, Volume-9, Issue-2, August 2020 Pages 24-31

Comparative Analysis of Classical Test Theory and Item Response Theory in Estimating Item Difficulty of BECE Mathematics Objective Items In Makurdi-Nigeria

Adagba Solomon, Prof. Emaikwu S. O, Prof. Obinne A.D.E

***Abstract*— This study compares Classical Test Theory and Item Response Theory in estimating item difficulty of Basic Education Certificate Examination (BECE) Mathematics Questions in Makurdi-Nigeria. To carry out the study two research questions were posed and one hypothesis was formulated. The study adopted ex-post factor study design. The population of the study consists of 7743 Junior Secondary School Students in JS III from 127 Government approved public and private schools in Makurdi-Nigeria. A total of 1520 JS III students responded to the four (4) research instruments used for the study. Multistage sampling procedure was used for the study. First the researcher used purposive sampling technique; this is because in purposive sampling, specific elements which satisfy some predetermined criteria are selected. The predetermined criterion here is that elements drawn for the study must be students of JS III who are preparing to take the Basic Education Certificate Examination (BECE). Kuder-Richardson 20 formula (K -R20) was used to obtain reliability coefficients of 0.81, 0.84, 0.70 and 0.71 for the four instruments. BILOG-MG" was used to compute the item parameters of CTT and IRT (item difficulty index). Independent t-test was used to test the hypothesis formulated. The result of the study revealed that; majority of the estimates of the item parameters in CTT were outside acceptable range of 0.30 to 0.70 while IRT have fewer items outside the acceptable range of -2 to 2. CTT-based and IRT-based item difficulty estimates were not statistically comparable; there is statistically significant difference between item difficulty estimates of students" responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT model and IRT model. Based on the findings, the study recommended that; the examination bodies using multiple-choice test instruments should employ the use of both IRT and CTT statistics in test development validation processes. Benue State Examination Board should frequently organize workshops, seminars and conferences to train and retrain their staff and test developers on test development process. This will improve the quality of BECE test items for effective assessment of learner's ability at basic level of education among others**

***Index Terms*— CTT, IRT, Item Difficulty, Mathematics.**

I. INTRODUCTION

It is undisputable that education is a key to economic growth of a country as well as in science and technology. Therefore, science and technology education are very important and crucial factors for the development of any nation. There is no doubt that what distinguishes the developed nations from the developing nations of the world is the degree of science and technology prevalent in these nations and Mathematics is the fulcrum on which science and technology rotate. Mathematics is one subject that is an integral part of everyone's life and affects virtually every field of human endeavour. An average man needs Mathematics to survive no matter how rudimentary. There is no doubt about the fact that an individual can get on sometimes without knowing how to read and write, but can never push on smoothly without knowing how to count, measure, add and subtract. The many uses and applications of Mathematics in the home, office, in business, in industries, in agriculture, in decision making and even in governance abound and are innumerable. It is, thus, vitally important both to the nation and to the individual that all students receive a quality Mathematics education.

Mathematics is no longer important just in so far as it is a basic requirement for entry into institutions of higher learning. It is now more than ever before an essential ingredient in the education of every Nigerian child especially in this technological era (Anaduaka and Okafor, 2013).Unfortunately, students' achievement in this important subject has been consistently poor especially in the Basic Education Certificate Examination (BECE). BECE is the examination written by Nigerian students at the end of their upper basic education and it is used to measure the extent of knowledge and skills the students have acquired at that level of education. The result of this examination is also used as prerequisite for admission into senior secondary school where students go into their areas of interest with compulsory credit pass in Mathematic. These poor achievements of students are on the high side despite all efforts by the government and other stakeholders to boost students' achievement in the subject. It is therefore a clear indication that there are still problems yet unsolved (Musa and Dauda, 2014).

Over the years, Mathematics educators have not relented in searching for better ways of teaching the subject. There has consequently been a myriad of research studies that have sought to identify the numerous factors affecting the teaching and learning of Mathematics and address the problem of poor achievement of students in the subject. However, despite their findings and recommendations, the problem of poor achievement of students in Mathematics still persists.

Students' poor achievement in BECE Mathematics objective

Adagba Solomon, Department of Educational Foundations and General Studies, University of Agriculture, Makurdi, Benue State-Nigeria
Prof. Emaikwu S. O, Department of Educational Foundations and General Studies, University of Agriculture, Makurdi, Benue State-Nigeria
Prof. Obinne A.D.E,, Department of Educational Foundations and General Studies, University of Agriculture, Makurdi, Benue State-Nigeria

items over the years in MakurdiMetropolis-Nigeria has been attributed to the fact that the subject is difficult and that the syllabus not well covered. However, various factors affect students' achievements in Mathematics especially at the Basic Education Certificate Examination level. Prominent among these factors are the nature of the test items (item difficulty) and the learners' characteristics. Items that could not differentiate between high and low ability students brings about poor achievement among intelligent and low intelligent students in examination (Adegoke, 2013). The achievement of an examinee on a test item can be predicted (or explained) by the ability of the examinee and characteristics of the item which can be measured using Classical Test Theory (CTT) and Item Response Theory (IRT). Another major issue in the study is the property of invariance of person and item characteristics between CTT and IRT for objective measurement. The argument here is that CTT is based on test level statistics while IRT is based on item statistics. Measurement that changes in results or findings when used across different objects cannot contribute to the growth of science or to the growth of objective knowledge in any area (Osarumwense and Oyedeji, 2015).

In measurement theory, analysis based on CTT has been used over the years and is still useful nowadays in test construction, although the trend is definitely towards item response theory (IRT) that provides for sample free and item free measurement. It is presently common to refer to IRT as the "modern" method of item analysis, with the obvious implication being that CTT is not modern. Not modern does not mean that CTT is no more useful in measurement theory. A primary criticism of CTT is the instability of its item and person statistics, that is, item statistics derived with CTT such as item difficulty and discrimination, are dependent on the sample of respondents. Due to the instability of CTT item and test statistics, many researchers assumed that invariance characteristics of IRT parameter estimates makes it superior to CTT in educational measurements (Guler, Uyanik and Teker, 2014). However, the empirical studies especially in Africa on the superiority of IRT to CTT in measurement theory are very scarce to support this assumption. The empirical studies available, however, have primarily focused on the application in test equating and very few studies have compared CTT and IRT for item analysis and test design. It appears that the superiority of IRT over CTT in this regard has been taken for granted in the measurement community, and no empirical scrutiny has been deemed necessary. The empirical silence on this issue seems to be anomaly." The major criticism for CTT is its inability to produce item/person statistics that are invariant across examinee/item samples. This criticism has been the major impetus for the development of IRT models and for the exponential growth of IRT research and applications in recent decades (Awopeju and Afolabi 2016). Despite theoretical differences between IRT and CTT, there is a lack of empirical knowledge about how, and to what extent, the IRT and CTT based item and person statistics behave differently. The degree of invariance of item parameter estimates across samples, usually considered as theoretically superiority of IRT models in measurement theory should be investigated, using empirical studies (Adedoyin, 2010). The major issues in the study is the students' achievements in Mathematics especially at the Basic Education Certificate Examination, how comparable and invariant is classical test theory and item response theory in estimating test item difficulty of 2011-2014 Mathematics examination questions of Basic education Certificate Examination (BECE) in MakurdiMetropolis-Nigeria.

II. Objective of the Study

The purpose of this study is to compare Classical Test Theory (CTT) and Item Response Theory (IRT) in estimating test item difficulty of 2011-2014 Basic Education Certificate Examination (BECE) objective items in Mathematics in MakurdiMetropolis-Nigeria. Specifically, the study sought to ascertain;

i. The item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT model
ii. The statistical relationship between the CTT and IRT item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics

Research Questions

The following research questions were formulated to guide the study.

i. What are the item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTTand IRT model?
ii. What is the comparison between CTT-based and IRT-based item difficulty estimates?

Hypothesis

The following research hypotheses guided the conduct of the study.

i. There is no statistically significant difference between the CTT and IRT based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics.

III. METHODOLOGY

The ex-post facto design was used for the study in collecting and analyzing the data. These designs wereconsidered suitable for the study since occurrence of event in the research hadalready taken place. In the context of educational research, ex-post facto also known as 'after the fact' or 'retrospective' investigate possible cause-and-effect relationships by observing an existing condition or state of affairs and searching back in time for plausible causal factors. In this case, the researcher examined retrospect plausible causal factors on the item difficulty index and of Mathematics objective items for 2011-2014 Basic Education Certificate Examination BECE in Makurdi Metropolis. The researcher also adopted ex post facto design because, respondents are not randomly assigned to an experimental group or control group, they are purposefully put into a particular group based on some prior things they have. In this case, the respondents in this group must have been into JS III preparing for Basic Education Certificate Examination

BECE.The study area of this research is Makurdi Metropolis-Nigeria.The population of the study consists of 7743 Junior Secondary School Students in JS 111 from 127 Government approved public and private schools in Makurdi Metropolis-Nigeria who registered and prepared to write their Basic Education Certificate Examination (BECE). The sample size for the study consist of 1549 JS III students from Makurdi Metropolis. The sample size was determined using 20% of the total population.The instruments for data collection were Mathematics objective questions of 2011-2014 Basic Education Certificate Examination (BECE), developed by the Benue State Examination Board. In order to establish the internal consistency of the instruments for 2011-2014 Basic Education Certificate Examination (BECE), a trial testing was done on 30 students for each of 2011-2014 Basic Education Certificate Examination (BECE) Mathematics objective questions papers and were collected immediately, marked and scored since the items were dichotomous; they have correct/wrong answers. It was analyzed using Kuder-Richardson 20 formula (K - R_{20}). The reliability coefficients obtained for the four (4) instruments were 0.81, 0.84, 0.70 and 0.71. Based on the values, it shows that the instruments have high internal consistency. The item analyses were carried out using CTT and IRT frameworks. BILOG-MO was used to compute the item parameters of CTT and IRT item difficulty and two parameter model was used because the items fitted more in it. The difference between CTT and IRT item difficulty was established by conducting independent t-test. For the purpose of this study and in accordance with the set benchmark for CTT, all the calculated item difficulty (p)for CTT framework that falls within the range of 0.3 to 0.7 are taken to be appropriate. This was according to Ojerinde (2013) who stated that, items difficulty that are less than 0.3 for CTT framework are regarded to be too difficult while all the items with item difficulty greater than 0.7 are regarded to be too simple. For IRT framework, according to DeMars (2010), item difficulty (b) that falls within the range of -2.00 to +2.00 are taken to appropriate for two parameter model.

IV. Results

Research Question 1: What is the item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT and IRT?

Table 1:CTT and IRT Item Difficulty estimates for 2011-2014 Multiple-choice Objective Question

	2011		2012		2013		2014	
Item	***p***	*b*	***p***	*b*	***p***	*b*	***p***	*b*
1	**0.52**	**-0.11**	**0.42**	**0.47**	**0.54**	**-0.22**	0.97	-5.54
2	0.07	5.42	0.03	4.48	**0.44**	**1.01**	0.86	**0.00**
3	**0.66**	**-0.74**	**0.66**	**-0.73**	0.23	**1.28**	0.86	-2.17
4	0.00	**0.00**	0.08	4.18	0.23	3.37	0.21	**1.54**
5	**0.39**	**0.56**	**0.42**	**0.42**	**0.49**	**0.00**	0.52	**0.00**
6	0.07	5.42	0.19	2.29	0.74	**-1.35**	0.97	-2.20
7	**0.54**	**-0.05**	**0.55**	**-0.25**	0.18	**1.45**	0.66	-2.18
8	**0.73**	**-0.97**	**0.62**	**-0.62**	0.84	**-1.80**	**0.34**	**0.00**
9	0.23	3.65	**0.34**	**1.87**	0.51	**-0.14**	0.76	-2.64
10	0.21	**0.00**	0.13	**0.00**	0.22	**0.98**	0.73	**-1.94**
11	**0.54**	**-0.14**	**0.56**	**-0.11**	0.70	**-0.77**	0.10	**0.00**
12	0.93	**-1.62**	0.93	**-1.71**	0.54	**-0.27**	0.11	2.42
13	0.81	**-0.79**	0.82	**-1.23**	0.47	**0.33**	0.03	4.07
14	0.91	**0.00**	0.92	**0.00**	**0.33**	**0.00**	0.24	**0.00**
15	**0.67**	**-0.59**	0.71	**-0.83**	0.26	2.27	0.14	4.90
16	0.06	5.50	0.13	**1.93**	**0.38**	**0.73**	0.24	3.50
17	0.00	**0.00**	0.00	**0.00**	**0.41**	**0.66**	**0.34**	**0.96**
18	**0.40**	**0.25**	**0.50**	**0.01**	**0.31**	**0.86**	**0.35**	**0.00**
19	0.06	5.50	0.13	**1.93**	0.23	2.52	**0.45**	**0.00**
20	**0.60**	**0.37**	0.75	**-1.12**	**0.40**	**0.29**	**0.41**	**1.04**
21	0.73	**-2.20**	**0.67**	**-1.76**	0.23	**1.02**	**0.42**	**0.45**
22	**0.65**	**-0.80**	**0.57**	**-0.37**	**0.38**	**0.39**	0.24	**1.58**
23	0.87	**-1.32**	0.81	**-1.42**	0.57	**0.51**	**0.49**	**0.09**
24	1.00	**0.00**	1.00	**0.00**	0.23	**1.61**	0.17	3.21
25	**0.55**	**-0.20**	**0.50**	**-0.09**	0.16	4.17	0.14	**0.00**
26	**0.33**	**0.00**	**0.30**	**1.86**	**0.61**	**0.86**	0.24	**0.87**
27	**0.60**	**-0.49**	**0.65**	**-0.45**	0.27	**1.74**	0.17	**1.42**
28	0.00	**0.00**	0.00	**0.00**	0.11	**0.00**	0.27	3.15
29	**0.48**	**0.38**	**0.46**	**0.19**	**0.35**	**0.56**	**0.62**	**-1.39**
30	**0.67**	**-0.73**	0.71	**-1.00**	**0.34**	**0.41**	**0.49**	**0.08**
31	**0.61**	**-0.49**	**0.60**	**-0.51**	**0.32**	3.92	**0.49**	**0.09**
32	**0.61**	**-0.55**	**0.53**	**-0.24**	**0.64**	**-0.61**	0.14	**0.00**
33	**0.61**	**-0.39**	**0.63**	**-1.17**	**0.36**	**1.84**	**0.48**	**0.34**
34	**0.61**	**-0.38**	**0.48**	**0.04**	0.18	**1.14**	0.10	**0.00**
35	0.73	**-0.94**	**0.62**	**-1.06**	**0.56**	**-0.27**	**0.34**	2.98

36	0.51	**-0.07**	**0.61**	**-0.35**	**0.62**	**-0.72**	**0.38**	**2.00**
37	0.20	**0.00**	0.21	**0.00**	**0.34**	**1.04**	0.07	0.00
38	**0.47**	**0.05**	**0.45**	**0.10**	**0.65**	**-0.54**	**0.49**	**0.14**
39	0.88	-2.49	0.80	**0.00**	**0.65**	**-0.54**	0.17	2.55
40	0.13	2.87	0.19	**0.80**	**0.52**	**-0.19**	0.17	**0.00**
41	0.21	**0.00**	0.19	**1.44**	**0.41**	**0.33**	0.28	**0.00**
42	0.13	4.71	0.11	**1.23**	0.13	**0.00**	0.14	2.76
43	0.20	6.48	0.26	**1.44**	0.15	**0.00**	**0.31**	**0.61**
44	0.27	**0.00**	0.23	**1.81**	0.16	**1.71**	0.17	3.30
45	0.12	6.58	0.24	**0.93**	**0.40**	**0.43**	**0.31**	0.00
46	0.80	**-1.36**	0.77	**-1.38**	**0.51**	**-0.13**	**0.48**	**0.13**
47	**0.60**	**-0.42**	**0.40**	**0.64**	0.09	**0.00**	0.20	**0.00**
48	0.94	-3.29	0.80	**-1.47**	**0.70**	**-0.66**	**0.31**	2.50
49	0.39	**0.40**	**0.40**	**0.26**	0.24	**0.00**	**0.52**	**-0.18**
50	0.00	**0.00**	0.09	**1.95**	0.17	3.62	0.24	**1.58**
51	**0.46**	**0.08**	**0.52**	**-0.11**	0.14	**0.00**	0.17	0.00
52	**0.55**	**-0.20**	**0.39**	**0.49**	**0.32**	2.19	**0.63**	**-0.88**
53	0.12	**0.00**	0.21	**1.30**	**0.40**	**0.56**	**0.59**	**-0.30**
54	**0.34**	2.44	**0.45**	**0.22**	**0.36**	**1.21**	**0.59**	**-0.70**
55	**0.62**	**0.00**	**0.70**	**0.00**	0.29	**1.89**	**0.56**	**-0.41**
56	**0.54**	**-0.13**	**0.53**	**-0.03**	**0.37**	**1.70**	**0.48**	**0.16**
57	0.87	**-1.72**	0.80	**-1.27**	**0.45**	**0.02**	**0.49**	**0.10**
58	0.06	**0.00**	0.08	**0.00**	0.28	3.16	0.28	**1.44**
59	**0.35**	**1.75**	**0.36**	**0.74**	**0.65**	**-0.54**	0.86	**-1.29**
60	0.91	4.36	0.18	3.38	0.20	4.09	0 19	**3.61**

From Table 1, on the basis of the criteria set for the difficulty index (i.e. .30≤p≤0.70), or 30% to 70%, the 2011 Basic Certificate Examination (BECE) items which failed to satisfy the conditions were: 2, 4, 6, 9, 10, 12, 13, 14, 16, 17, 19, 21, 23, 24, 28, 35, 37, 39, 40, 41, 42, 43, 44, 45, 46, 48, 50, 53, 57, 58 and 60. Therefore, on the basis of the level set for difficulty, 31 items from the 2011 BECE examinationneeded modification or should be discarded.

On the basis of the criteria set for IRT difficulty index (i.e. -2≤b≤2), 13 items from the 2011 Basic Certificate Examination (BECE) items failed to satisfy the condition which are: 2, 6, 9, 16, 19, 39, 40, 42, 43, 45, 48, 54 and 60. Therefore, on the basis of the level set for difficulty 13 items were either too simple or too difficult for examinee and therefore which needed modification or should be discarded.

From 2012, on the basis of the criteria set for the difficulty index (i.e. .30≤p≤0.70), or 30% to 70%, the 2012 Basic Certificate Examination (BECE) items which failed to satisfy the conditions were: 2, 4, 6, 9, 10, 12, 13, 14, 16, 17, 19, 23, 24, 26, 28, 37, 40, 41, 42, 43, 44, 45, 48, 50, 53, 57, 58 and 60. Therefore, on the basis of the level set for difficulty, 28 items from 2012examination needed modification or should be discarded.

On the basis of the criteria set for IRT difficulty index (i.e. -2≤b≤2), 4 items from the 2012 Basic Certificate Examination (BECE) items failed to satisfy the condition which are: 2, 4, 6 and 60. Therefore, on the basis of the level set for difficulty 4 items were either too simple or too difficult for examinee and therefore which needed modification or should be discarded.

From 2013,on the basis of the criteria set for the difficulty index (i.e. .30≤p≤0.70), or 30% to 70%, the 2013 Basic Certificate Examination (BECE) items which failed to satisfy the conditions were: 3, 4, 7, 8, 10, 15, 19, 21, 24, 27, 28, 34, 42, 43, 44, 47, 49, 50, 51, 55, 58 and 60. Therefore, on the basis of the level set for difficulty, 22 items from 2013 BECEexamination needed modification or should be discarded.

On the basis of the criteria set for IRT difficulty index (i.e. -2≤b≤2), 8 items from the 2013 Basic Certificate Examination (BECE) items failed to satisfy the condition which are: 3, 15, 19, 25, 31, 50, 52, and 58. Therefore, on the basis of the level set for difficulty 8 items were either too simple or too difficult for examinee and therefore should be discarded.

From 2014, on the basis of the criteria set for the difficulty index (i.e. .30≤p≤0.70), or 30% to 70%, the 2014 Basic Certificate Examination (BECE) items which failed to satisfy the conditions were: 1, 2, 3, 4, 6, 11, 12, 13, 14, 15, 16, 22, 24, 25, 26, 27, 28, 32, 34, 37, 40, 41, 42, 44, 47, 50, 51, 58 and 60. Therefore, on the basis of the level set for difficulty, 29 items from 2014 BECE examination needed modification or should be discarded.

On the basis of the criteria set for IRT difficulty index (i.e. -2≤b≤2), 19 items from the 2011 Basic Certificate Examination (BECE) items failed to satisfy the condition which are: 1, 3, 6, 7, 9, 12, 13, 15, 16, 24, 28, 35, 37, 39, 42, 44, 45, 48 and 51. Therefore, on the basis of the level set for difficulty 19 items were either too simple or too difficult for examinee and therefore needed modification or should be discarded.

Research Question 2: What is the comparison between CTT-based and IRT-based item difficulty estimates?

Table 2: Summary of CTT and IRTdifficulty index for 2011-2014 BECE Mathematics Objective Items

Category	N	Good items	Poor items	Good items	Poor items	Good items	Poor items	Good items	Poor items

		2011		2012		2013		2014	
CTT	60	32	28	38	22	31	29	7	53
IRT	60	56	4	52	8	41	19	38	22

Table 2 above present the summary of differences in CTT and IRT difficulty index for 2011-2014 Basic Education Certificate Examination (BECE). The Table revealed that, out of 60 items for each test form, CTT has 32 good items compared to IRT with 56 good items because of their acceptable difficulty index (p= 0.30 *to* 0.70 and b= -2 to 2). CTT has the highest number of bad items (28 items) compare to IRT (4 items) with the difference of 24 bad items.

For 2012, the Table revealed that, out of 60 items for each test form, CTT has 38 good items compared to IRT with 52 good items because of their acceptable difficulty index (p= 0.30 *to* 0.70 and b= -2 to 2). CTT has the highest number of bad items (22 items) compare to IRT (8 items) with the difference of 14 bad items.

For 2013,the Table revealed that, out of 60 items for each test form, CTT has 31 good items compared to IRT with 41 good items because of their acceptable difficulty index (p = 0.30 *to* 0.70 and b= -2 to 2). CTT has the highest number of bad items (29 items) compare to IRT (19 items) with the difference of 10 bad items.

For 2014, the Table revealed that, out of 60 items for each test form, CTT has 7 good items compared to IRT with 38 good items because of their acceptable difficulty index (p= 0.30 *to* 0.70 and b= -2 to 2). CTT has the highest number of bad items (53 items) compare to IRT (22 items) with the difference of 31 bad items.

Hypothesis

There is no statistically significant difference between the CTT and IRT based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics.

Table 3: Independent t-test of Significance Mean Differences between CTT and IRT Based Item Difficulty Estimates

Parameters	N	Mean	Std	df	t	P-value	α	Remark
CTT	240	.3227	.24896					
IRT	240	.5399	1.71913	478	5.005	.000	0.05	**Significant**
Total	480							

P<0.05

Table 3 above revealed the independent t-test results of the mean difference between CTT and IRT Based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics. The result indicates a statistical significant mean difference between CTT and IRT Based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics ($t = 5.005$, $df = 478$, $p = .000<0.05$). This implies that there is statistically significant mean difference between CTT and IRT Based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics

Summary of Major Findings

i. The item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT revealed that majority of the items have poor item difficulty to measure student ability.

ii. Item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on IRT revealed majority of the items good item difficulty to measure student ability with little item modification.

iii. There is statistically significant mean difference between the CTT and IRT based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics.

V. Discussion of Finding

The results of the research questions and hypotheses were discussed according to the stated objectives.

Findings from research question 1 as presented in Table1revealed that item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematicsfor 2011-2014 examination in Makurdi-Nigeriabased on CTT have high poorly item difficulty indices to measure student ability. The result shows that, out of 60 items for 2011 BECE examination 31 have poor difficulty index, for 2012 BECE examination 28 items have poor difficulty index and, for 2013 BECE examination 22 items have poor difficulty index and for 2014 BECE examination 29 items have poor difficulty index respectively. From the findings, BECE Mathematics objective items for the 4 years have high poor items that should have been modified or deleted. The finding of the study supports the work of Ajeigbe (2018) who conducted a research on assessing quality of Osun State Mathematics multiple-choice items under 2-parameter model of item response theory. The study determined item difficulty of Osun State Mathematics Multiple-choice items under 2-parameter model of Item Response Theory. The result revealed that 9 (25.5%) items were easy which fell under $-3.00 \leq -1.00$; 25 (62.5%) items were moderately difficult within the range of $-1.00 \leq 1.00$; and 6 (15%) items were very difficult within the range of $1.00 \geq 2.00$. The results also revealed the classical statistics for total scores with means of 17.29 and 17.29 for 2-paramter with coefficient α (alpha) reliability coefficient of 0.89. The study concluded that the Mathematics items used in assessing students' cognitive level were found to have most of its performing averagely in terms of item quality.

Furthermore, finding revealed that the item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on IRT for 2011-2014 examination in Makurdi-Nigeria have good and poor item difficulty to

measure student ability as the case may be. The result shows that, out of 60 items for 2011 BECE examination 13 have poor difficulty index, for 2012 BECE examination 4 items have poor difficulty index, for 2013 BECE examination 8 items have poor difficulty index and for 2014 BECE examination 19 items have poor difficulty index respectively. From the findings, BECE Mathematics objective items for 2011 and 2014 have high poor items that should have been modified or deleted than 2012 and 2013. The findings showed that IRT item calibration produced more good items and fewer items for modification or deletion than CTT items calibration. The finding of the study Adegoke (2013), who conducted a study on Comparison of Item Statistics of Mathematics Achievement Test using Classical Test and Item Response Theory Frameworks. Results showed that item statistics obtained from IRT 2-parameter model appeared more stable than those from CTT. Moreover, for item selection process, IRT 2-parameter model led to deletion of fewer items than CTT model. This result implies that test developers and public examining bodies should integrate IRT model into their test development processes. Through IRT model, test constructors would be able to generate more reliable items than in the CTT model being currently used and ultimately the test scores of examinees will be more reliably estimated.

The finding from the hypothesis revealedstatistically significant difference between the CTT and IRT based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics. The result in Table 3 above revealed the independent t-test results of the mean difference between CTT and IRT Based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics. The result indicates a statistical significant mean difference between CTT and IRT Based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics ($t = 5.005$, $df = 478$, $p = .000<0.05$). This implies that there is statistically significant mean difference between CTT and IRT Based item difficulty estimates of students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics. The findings disagree with Osarumwense and Oyedeji (2015), they research on empirical comparison of methods of establishing item difficulty index of test items using classical test theory (CTT) empirically compare two methods of computing the item difficulty index of test items based on Classical lest Theory (CTT). After the item analysis, it was found that the 2010 Upper Basic Certificate Mathematics objective questions were within the recommended range of 0.30-0.70. The findings also showed a positive strong relationship between the item difficulty indices obtained by using the two methods. Finally, the findings revealed that there was no significant difference between the means of the item difficulty indices obtained by using the two methods.

VI. Conclusion

Based on the result, the following conclusions were drawn:

i. The two-parameter logistic model was successfully applied in the calibration of students'responses to Basic Education Certificate Examination (BECE) Objective Questions inMathematics based on CTT and IRT model.

ii. The students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT revealed that majority of the test items have poor item difficulty index to measure student ability. This is because CTT do not examined item characteristics in details like the IRT does, the validity and reliability of the test is based upon the total test scores regardless of students ability

iii. The IRT produced test itemswith good item difficulty index to measure student ability with little item modification. The high result of good test items in IRT as compared to CTT is of the fact that IRT focuses on item by item analysis and the validity of the test items is assessed for each item with the reliability calculated for each person's ability which varies across the continuum, having more precision at the center of performance distribution between the low and high ability students.

iv. The BECE test items were not good enough to measure students' ability as there was no application of CTT and IRT statistics in test development validation process to check the items validity and reliability to measure students ability which could be the cause of students' poor achievement in Basic EducationCertificate Examination (BECE) in Mathematics inMakurdi Metropolis.

v. The result of CTT and IRT were not comparable statistically in estimating item difficulty indexbut thus could be used as complementary procedures in the development of Basic EducationCertificate Examination (BECE) in Mathematics inMakurdi Metropolis.

VII. Recommendations

It is therefore recommended that:

i. The examining bodies using multiple-choice test instruments should employ the use of both IRT and CTT statistics in test development validation processes. This will ensure effective test development in that both statistics will complement one another.

ii. Benue State Examination Board should frequently organize workshops, seminars and conferences to train and retrain their staff and test developers on test development process. This will improve the quality of BECE test items for effective assessment of learner's ability at basic level of education.

iii. Benue State Government should employ qualified personnel to pilot the affairs BECE with expected professionalism. This can only be achieved if the government set aside nepotism and employ only measurement experts who could develop mathematics test items that meet up with the students ability, having more precision at the center of performance distribution between the low and high ability students

iv. Benue State Ministry of Education should set up a committee to monitor the conduct of BECE across the Local Government Areas in the State. The monitoring should cover the test development process where pilot testing must be emphasized. Conducting BECE below the standard for quality assurance should be banned.

VIII. Contribution to Knowledge

The study has contributed to knowledge in the following areas:

i. The study have successfully applied two-parameter logistic model in the calibration of students'responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT and IRT model.The two parameter logistic model revealed that, students' responses to Basic Education Certificate Examination (BECE) Objective Questions in Mathematics based on CTT have majority of the poor item difficulty index to measure student ability. This was because CTT do not examined item characteristics in details like the IRT does, the validity and reliability of the test is based upon the total test scores regardless of students ability

ii. The study further revealed that IRT produced test itemswith good item difficulty index to measure student ability with little item modification. The high result of good test items in IRT as compared to CTT is of the fact that IRT focuses on item by item analysis and the validity of the test items is assessed for each item with the reliability calculated for each person's ability which varies across the continuum, having more precision at the center of performance distribution between the low and high ability students.

IX. Significance of the Study

The findings from the study would be useful to examination bodies, teachers, test developers and researchers in the measurement of ability.

The result of this study would provide useful information to West African Examination Council, National Examination Council (NECO), and State Ministries of Education, on how to construct test items that differentiates between high ability and low ability students in Mathematics examinations.

This will enable teachers to understand the variability of item parameters in test development and individual test achievement ability based on test-level and item-level. It will guide teachers not to depend on students' scores in examination but also look at their performance at the item level to know whether there are items that all the students finds them difficult to answer in order to make amendment.

The result of the study is also expected to guide the test developers on the best and more economical method, in terms of finances and time, to use when computing test item parameters.

The research will serve as a reference point to other researchers in the area of item analysis. Copies of the research will be made available in the library and printed Medias for researchers use.

References

[1] Adedoyin, C. (2010). Investigating the Invariance of Persons Parameter Estimates based on Classical Test and Item Response Theories 2010. *An international journal on education science*,2, (2), 107-113

[2] Adegoke, B. A. (2013). Comparison of item statistics of Mathematics achievement test using classical test and item response theory frameworks. *Journal of Education and Practice*,4,(22), 87-96

[3] Ajeigbe, T.O. (2018). Assessing quality of Osun State Mathematics multiple-choice items under 2-parameter model of item response theory. *Nigerian Journal of Educational Foundations*, 17,(1), 27-37

[4] Alodiah, C. O. (2012). Item response theory (IRT) - A more desirable choice for quality assurance in Education. *Research in Education*, 18,(1), 255-258

[5] Anaduaka, U. S. &Okafor, C. F. (2013). Poor Achievement of Nigerian Students in Mathematics in Senior Secondary Certificate Examination (SSCE): What is not working? *Journal of Research in National Development*, 11,(2), 1-5

[6] Awopeju, O. A. &Afolabi, E. R. I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal*, 12,(28), 263-284

[7] DeMars, C. (2010). *Item Response Theory: Understanding Statistics.* New York: Oxford University Press, Inc.

[8] Guler, N., Uyanik, G. K. &Teker, G. T. (2014). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of Research on Education*, 2,(1), 1-6.

[9] Hambleton, R. K, &Swaminathan, H. (2005). Item response theory: Principles and applications. Boston, MA: Kluwer Academic Publishers.

[10] Jimelo, L. Silvestre. T. (2009). Item response theory and classical test theory; an empirical comparison of item/person statistics in a biological science test. *The International Journal of Educational and Psychological Assessment*,1, (1) 7-12.

[11] Lord, F. (1952). A theory of test scores. *Psychometric Monograph* (7), Psychometric Society.

[12] Lord, F. M. (1968). An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.

[13] Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment. 1(1), pp. 1-11*

[14] Musa, M. &Dauda, E. S. (2014). Trends analyses of students' Mathematics achievement Basic Education Certificate Examination (BECE) from 2004 to 2013: implication for Nigeria's vision 20:2020. *British Journal of Education*, 2,(7), 50-64.

[15] Nenty, H. J. (2004). From classical test theory (CCT) to Item Response Theory (IRT): An introduction to desirable transition. In O. A. Afemikhe& J. B. Adewalc (Eds), *Issues in Educational Measurement and Evaluation in Nigeria (371-384).* Ibadan: Educational Research and Study Group.

[16] Ojerinde, D. (2013). Classical Test Theory (CTT) Vs Item Response Theory (1RT): An evaluation of the comparability of item analysis result. *A guest lecture presented at the Institute of Education*, University of Ibadan on 23rd May.

[17] Ojerinde, D., Popoola, K, Ojo, F. &Ariyo, A. (2014) *Practical application of item response theory in large scale measurement.*Abuja*:* Marvelous Mike Press Ltd.

[18] Ojerinde, D., Popoola, K., Ojo. F., and Onyeneho, O. P. (2012). *Introduction to Item Response Theory: Parameter models, estimation and application.* Ibadan:Goshen Print media Ltd

[19] Olabode, J. O. &Adeleke, J. O. (2015). Comparative analysis of item local independence of WAEC and neco 2012 Mathematics test items. *Nigerian Journal, of Educational Research and Evaluation*, 14, (1) 66-73.

[20] Osarumwense, H. J. &Oyedeji, S. O. (2015). Empirical comparison of methods of establishing item difficulty index of test items using Classical Test Theory (CTT). *Journal of Educational Policy and Entrepreneurial Research (JEPER)*, 2,(10) 98-109.

[21] Peak, I. & Young, M. J. *(2005). Investigation of student growth recovery in a fixed-item linking procedure with a fixed-person prior distribution for mixed-format test data.Applied Measurement in Education*, 18, 199-215.

[22] Skompski, W. P., Jodoin, M. G., Keller, L. A. &Swaminathan, H. (2003). *An evaluation of item response theory equating procedures for*

capturing growth with tests composed of dichotomously scored items. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

[23] Spearman, C. (1904). General Intelligence, Objectively Determined and Measured. The*AmericanJournal of Psychology*, 15,(2), 201-292.

[24] Stage, C. *(2003).CIassical test theory or item response theory: The Swedish experience.* Umea: Kluwer Academic Publisher

[25] Umobong**,** M. E. (2004). Item Response Theory: Introduction objectivity into educational measurement. In O. A. Afemikhe& J. B. Adewale (Eds), *Issues in Educational Measurement and Evaluation in Nigeria (pp. 385-398).* Ibadan: Educational Research and Study Group.

[26] Wiberg, M. (2004).*Classical test theory vs. item response theory: An evaluation of the theory test in the Swedish driving-license test.* Umea: Kluwer Academic Publications.

[27] Zanon, C., Hutz, C. S., Yoo, H. &Hambleton, H. R. (2016). An application of item response theory to psychological test development. *Psicologia: Rejlexdo e Critica,* 29,(18) 1-10.