ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

examination bodies determines the level of instructional objective achieved (Reid, 2003; Fakeye, 2014).

Before the introduction of National Values Education at the BECE level, students' performance in the state and national BECE Social Studies and civic education subjects was at variance in Benue State. From 2012 to 2016, records show that students' performance in national BECE Social Studies and civic education at distinction level was constantly less than 5% every year. The report also shows that students' performance in national BECE Social Studies and civic education at credit level was less than 50% every year. More than 10% of the total candidates who participated in State BECE passed the subjects at distinction levels and 50% passed at credit level from 2012 to 2016. Benue State Examinations Board and NECO Records, (2012 – 2016) revealed the statistics of students' achievement in national (NECO) and Benue State BECEs Social Studies and civic education (National Values Education subjects) across the years.

Social Studies, civic education and security education existing as distinct subjects are integrated to form what is known as National Values Education in Basic Education Certificate Examinations (BECE). The aim of National Values Education curriculum at basic secondary school level is to inculcate into the Nigerian youths the idea of patriotism, national unity and consciousness, societal values, positive attitudes, religious tolerance, democratic principles, and security consciousness, among others, for national growth and development.

It has been observed that many researchers often use item difficulty and item discrimination interchangeably. However, these two item parameters are not the same. Item difficulty can be explained as the mental task required of an item to be answered correctly by an examinee based on their ability, whereas item discrimination is the capacity of an item to differentiate between examinees' ability (low and high scorers) based on their individual response.Item Response Theory (IRT) is basically used for assessing item parameters due to the fact that its model is used to estimate parameters of test items and abilities or latent traits of test takers, with the estimation of item parameters (difficulty and discrimination) and test taker ability are placed on the same scale.

Adegoke (2013) stated that in determining the quality of dichotomous test items, item difficulty and discrimination indices should be considered. He further observed that items that yield statistical value <0.4 are considered less difficult, while those with statistical values between 0.4 – 0.6 or 0.7, as the case may be, are considered having moderate difficulty level.Those that yield values >0.7 are considered too difficult. Also, the items with discrimination statistical values from 0.2 and above are considered as good items that can differentiate between high and low ability examinees. Therefore, it is important to examine how difficult and discriminating the test items set by Benue State examination board and NECO for Basic Education Certificate Examinations (BECE) in National Values Education (Social Studies, civic education and security education) in the state.

Therefore, it is expected that as a public examination body, Benue State Examinations Board (BSEB), should develop valid and reliable test items for assessment of learning outcomes in BECE subjects that have the same standard as the National Examination Councils (NECO). Non-conformity of state BECE tests to acceptable standard item parameters and test assumptions will produce invalid and unreliable scores. The scores produced using low test item quality will hinder decision-making on promotion and placement of students to the senior school level. The quality of results produced by public examination bodies determines the level of instructional objective achieved (Reid, 2003; Fakeye, 2014).

Test developers always try to construct items that provide unique information to measure testees' abilities (knowledge and skills). The most popular IRT models specify a single latent trait to account for all statistical dependencies among test items and all differences among test takers. This underlying ability or trait, typically denoted theta (θ), distinguishes items in respect

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

to difficulty and discriminates testee's ability or proficiency. The probability that a testee will provide a correct or an incorrect response to an item is a function of the testee's location on theta (θ) and one, two or three IRT parameter models chosen normally and which describes the relationship between the item and theta (Hambleton, & Rodgers, 1995). Item Response Theory is a set of models which, relate the likelihood of a particular reaction by an individual with a given trait level to the characteristics of the item designed to elicit the level to which the individual possesses that trait. IRT attempts to model the relationship between a testee's latent abilities and probability of the testee responding to a certain item correctly (XinmingAn & Yung 2014). According to Nenty (2015), IRT attempts to estimate the parameters involved, explain the process and predict the results of a given measurement.

According to Hambleton and Jones (1993), Item Response theory (IRT) is a universally used statistical theory by psychometricians to assess examinees' ability, item quality, and how the individual examinee's performance relates to the latent traits measured by the test items. The theory attempts to model the examinee's ability and the probability of correctly answering a particular item based on the item response pattern or format of a given test. Under IRT, the major concern is whether individual examinees answer each test item correctly or incorrectly, unlike CTT where the emphasis is on the raw test scores. IRT holds that the probability of a testee with given ability answering a random test item *i* with a given difficulty *bi* correctly is conditioned upon the ability of the testee and the item difficulty level. That is to say, if a person has high ability in a given subject, a testee will probably get an item with low difficulty level correct. Likewise, if a testee has a low ability in an item, he or she will probably answer the item wrongly (XinmingAn & Yung, 2014).

The probability of answering an item correctly is a function of both the examinee's ability level and latent trait and the properties of a given item. The latent trait is commonly denoted or represented theta (θ). The value of θ for a given person is known as the person's location. The item properties are parameters known as item difficulty and discrimination usually estimated under the IRT model. The difficulty parameter, or item location denoted by b, represents the location of an item on the ability scale. In designing research instrument to measure different latent trait levels, a researcher is expected to have items with difficulties across the full range of the trait. The second item parameter is discrimination, related to the steepness of a line slope on the Item Characteristic Curve (ICC). Discrimination of a test item is denoted by 'a'. This item parameter explainsthe extent to whichthe probability ofsuccess on an item changedfast with theta near the item difficulty. An item with a high discrimination index indicates a correlation between the latent trait and the probability of answering that particular item correctly. In other words, an item with a high discrimination parameter can differentiate better between low and high achievers. Adedoyin (2010) observed that the sloppier or flatter the ICC, the less an item will discriminate; hence the probability of correct response at the ability level is almost 50% with the high ability levels, likewise the steeper the curve the higher the possibility for an item to discriminate. It is expected that any duo standardised tests with the same objectives should possess the same or similar parameters when administered on a homogeneous population (testees).

According to Emaikwu (2013), the achievement test result, especially at the summative level, is an important parameter the society uses to adjudge the quality of its school products. The reason for testing is to examine the cognitive ability of the examinee in specific areas of human endeavour. However, to achieve this intended function in an examination, test items need to possess the necessary parameters and assumptions (difficulty, discrimination and guessing).

However, Ayanwale, Adeleke and Mamadelo (2018) carried out assessment of item statistics estimates of BECE 2017 Mathematics using CTT and IRT approaches and found that 33 items under CTT and 12 items under IRT had poor discrimination and difficulty indices and needed to be deleted. The researchers adopted

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

60 items of NECO BECE 2017 Mathematics objective test (paper 1) as instrument for data collection, administered it on 978 randomly sampled basic nine students from Osogbo and Olorunda Local Government Areas of Osun State. The collected data were analysed using JMETRIK and NOHARM.Based on their study findings, they suggested that NECO should adopt the use of IRT in standardising test items. The findings of this study is a proof that even NECO BECE as a national examination body employs poor psychometric properties for the assessment of examinees' ability. The identification of good items is done based on Demars (2010) specification who stated that good dichotomous items are characterized by: 'a' = ≥ 0.2 and 'b' = values ranging between -2 and 2.

It has been observed that learners who are exposed to the same curriculum specification have been performing at variance in BECE conducted by Benue State Examination Board and NECO. Also, there is increasing public outcry over the years concerning the low quality of BECE items use for measuring examinees' ability in the Benue State. The quality of test items used by the state examination board could be as a result of the poor test item parameters (difficulty and discrimination). It is doubtful whether BECE items particularly the one developed by Benue State Examination Board are standardised using the appropriate psychometric approach. Based on available literature, the researchers could not find existing studies that examined and compared Benue State and NECO BECE item difficulty and discrimination indices as a possible cause of variation in students' performance. Therefore, the need to analyse and compare BECE items developed by the two examining bodies employing IRT validity approach cannot be overemphasised. It is based on the uncertainty concerning the quality of Benue State BECE items that the researchers seek to investigate and compare the test parameters of Benue State and NECO BECE items 2017 in National Values Education in Benue State, Nigeria.

Research Questions

The following questions were developed for the study:

1. What is the difference between item parameters (difficulty and discrimination indices) of Benue state (BSEB) and NECO BECE 2017 National Values Education?
2. Is there significant mean difference between Benue state (BSEB) and NECO BECE 2017 National Values Education in terms of item:
   a. Difficulty parameters?
   b. Discrimination parameters?

Methodology

The study is a descriptive survey type because Benue State and NECO BECE multiple choice National values education 2017 were adopted and administered on a large sampled population of junior secondary school (JSS3) students. Adopting a descriptive survey is appropriate because it uses large and small populations (a universal set) by studying samples selected from a population to discover the relative incidence, manifestation, distribution and interrelationship between or among sociological and psychological variables (Kerlinger and Lee, 2000).The study target population comprises all Junior Secondary School Three (JSS3) students across 472 secondary schools in Benue State, 320 are State BECE registered schools while 152 are NECO-BECE registered Schools (Source: BSEB and NECO Records, 2017).

A multistage sampling approach was adopted: a simple random sampling technique was adopted to select two senatorial districts or zones: Benue North-West (Zone B) and Benue South (Zone C) which represents Tiv and Idoma tribes in Benue State. Benue North-west senatorial district (Zone B) was randomly selected out of two senatorial districts in Tiv geopolitical zone. At the second stage, a purposive-sampling technique was adopted to sample eight (8) local government areas (4 local government areas each) from the two selected senatorial districts. The choice of purposive sampling technique was appropriate as only local government areas with both state BECE and NECO-BECE registered schools were considered suitable for the study. A total of six (6) schools (four State BECE-registered schools and two NECO BECE-registered schools) per local government

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

area were selected for test administration. The use of stratified random sampling technique was based on the fact that the schools were classified under State BECE-registered and NECO BECE-registered secondary schools in Benue State. The study therefore, used a sample size of 1500 JSS3 students in Benue State.

Benue State Examination Board (BSEB) BECE National Values Education multiple-choice items 2017 and National Examination Councils (NECO)BECE National Values Education multiple-choice items 2017 were adopted for data collection. The psychometric properties (item difficulty and discrimination indices) of the two tests (BSEB and NECO) BECE National Values Education 2017 were compared. To have uniformed number of multiple-choice items from BSEB and NECO BECE, items were scaled to 45 each using curriculum objectives. The multiple-choice items comprised of a stem and four options (A – D) with a key and distractors. The correct response to an item was scored '1' and incorrect response or un-answered items were scored '0'. Formula for item scaling:

$$Ic = \frac{ni}{Ni} = i$$

Where: IC = item scaling
ni = the required number of observed items from individual topic
Ni = sum of existing items to be scaled.
i = sum of items expected

Data for the study were collected through direct visit and administration to schools. The researcher, with the aid of five research assistants visited and administered the two versions of adopted BECE National Values Education 2017 at the interval of two weeks to respective sampled schools. During the first administration, testees (students) were given identification numbers, and attendance was taken for identification and administration of the second test. In order to control for measurement error due to time variance during administration of the two tests, counterbalance approach was used. Counterbalance approach is a technique that enables a researcher to ensure fairness of time used for administration of multiple related tests.

The collected data were analysed using IRT Logistics Parameter and t-test statistics.

Results

The item parameters (discrimination and difficulty) under IRT were estimated where the testees' responses were modelled with their ability or theta (θ) using 3PL as a model that fits the data set. The identification of good items was done based on the specification that good dichotomous items are characterised by: 'a' = ≥ 0.2 and 'b' = values ranging between -2 and 2. (DeMars, 2010; Sansivieri, Wiberg & Matteucci, 2017)

For the test with unidimensionality:

$$b = \frac{-(d)}{a} \quad \text{.....Equation1}$$

Where: a = item discrimination; b = item difficulty; d = intercept.

For the test with multidimensionality, the item parameters were transformed to the same scale of Unidimensionality using the following formula:

$$Mdisc = \sqrt{a_1^2 + a_2^2} \quad \text{.....Equation2}$$

$$M\,diff = \frac{-(d)}{Mdisc} \quad \text{.....Equation3}$$

Where
a_1 = estimated ability for the 1st dimension (F1)
a_2 = estimated ability for the 2nd dimension (F2)
d. = intercept
Mdisc = discrimination for multidimensional test
Mdiff = difficulty for multidimensional test

1. What is the difference between item parameters (difficulty and Discrimination indices) of Benue State (BSEB) and NECO BECE 2017 National Values Education?

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 1a

Item Parameter Estimate of BSEB and NECO BECEs 2017 National Values Education with Unidimensionality and Multidimensionality

Item	BSEB Item Parameter with Unidimensionality				NECO Item Parameter with Multidimensionality					
	a	d	b	c	a1	a2	d	c	Mdisc	Mdiff
1	1.81	-2.22	1.23	0.20	-1.62	0.58	-0.41	0.00	1.72	0.24
2	16.22	-15.49	0.96	0.29	-0.29	-0.24	-1.45	0.00	0.38	3.82
3	5.28	-4.51	0.85	0.20	-0.77	0.21	-0.95	0.00	0.79	1.19
4	2.47	-2.50	1.01	0.25	-3.2	0.04	0.38	0.00	3.20	-0.12
5	7.78	-7.27	0.93	0.31	-3.41	0.07	-0.66	0.28	3.41	0.19
6	4.77	-4.43	0.93	0.36	-3.5	0.51	-1.69	0.23	3.53	0.48
7	3.67	-3.26	0.89	0.22	0.59	0.11	-1.95	0.00	0.60	3.27
8	3.23	-3.15	0.98	0.29	-1	-0.1	-0.73	0.00	1.01	0.72
9	1.36	-1.58	1.16	0.53	-2.21	0.02	-1.78	0.26	2.21	0.80
10	1.17	-0.47	0.40	0.12	-1.78	0.2	-0.87	0.11	1.79	0.49
11	2.82	-2.13	0.76	0.31	-1.89	-0.21	-0.63	0.15	1.90	0.33
12	1.05	-1.11	1.05	0.15	-1.34	0.11	-1.31	0.16	1.35	0.97
13	0.29	-0.40	1.38	0.00	-2.42	0.38	-0.76	0.10	2.45	0.31
14	0.99	-0.58	0.58	0.00	-2.97	0.38	-1.87	0.17	2.99	0.62
15	1.08	-0.64	0.59	0.09	-34.34	-2.56	-27.54	0.35	34.44	0.80
16	0.72	0.13	-0.18	0.00	-8.36	16.53	-40.21	0.22	18.52	2.17
17	0.80	-1.91	2.38	0.40	-0.21	-0.57	-0.66	0.00	0.61	1.08
18	0.79	-0.26	0.33	0.00	2.65	-0.68	-4.97	0.20	2.74	1.81
19	0.08	-0.35	4.33	0.02	-5.73	0.9	-4.20	0.29	5.80	0.72
20	0.36	-0.07	0.21	0.00	-1.57	-0.46	-2.91	0.21	1.63	1.78
21	0.27	-0.17	0.62	0.01	-2.13	-0.03	-1.40	0.25	2.13	0.66
22	-0.07	0.00	0.00	0.04	-1.05	-0.33	-0.74	0.07	1.10	0.67
23	0.46	0.52	-1.11	0.01	-0.31	-0.21	-0.64	0.00	0.37	1.73
24	1.35	-2.49	1.85	0.40	-3.21	-0.38	-3.91	0.21	3.23	1.21
25	0.24	0.56	-2.38	0.01	-1.43	-0.48	-0.51	0.11	1.51	0.34
26	3.56	-4.89	1.37	0.41	-1.39	0.06	-1.61	0.33	1.39	1.16
27	2.46	-3.29	1.34	0.42	-1.92	-0.5	-1.30	0.25	1.98	0.66
28	0.15	-0.05	0.35	0.02	-1.63	-0.02	-1.64	0.21	1.63	1.01
29	0.35	-0.31	-0.88	0.00	-1.19	-0.54	-1.19	0.16	1.31	0.91
30	0.02	-0.09	5.44	0.41	-6.4	2.13	-12.70	0.23	6.74	1.88
31	4.28	-4.01	0.94	0.36	-4.78	-1.56	-4.80	0.38	5.02	0.96
32	3.64	-5.00	1.37	0.39	-3.37	-1.85	-3.07	0.36	3.84	0.80
33	2.35	-5.51	2.35	0.41	0.58	0.13	-1.37	0.00	0.59	2.30
34	0.25	-0.32	1.25	0.01	-3.07	-2.73	-2.04	0.31	4.11	0.50
35	0.12	-0.53	4.48	0.01	-0.02	-0.3	-0.49	0.00	0.30	1.67
36	-0.69	-1.75	-2.55	0.48	-0.35	-0.48	-0.49	0.00	0.60	0.81
37	0.30	-0.03	0.76	0.39	-1.46	-1.88	0.15	0.05	2.38	-0.06
38	0.94	-1.33	1.42	0.44	-0.13	-0.65	-0.55	0.00	0.66	0.83
39	3.26	-5.06	1.55	0.39	-1.91	-1.99	-0.80	0.18	2.76	0.29
40	4.24	-4.37	1.03	0.39	1.21	-1.66	-3.86	0.23	0.50	1.88
41	2.16	-3.56	1.65	0.42	-3.06	-0.84	-4.26	0.30	3.18	1.34
42	-0.09	-1.38	-14.64	0.15	-1.86	-3.71	-2.04	0.12	4.15	0.49
43	3.09	-5.28	1.71	0.39	-26.65	-41.11	-27.18	0.32	48.99	0.55
44	2.95	-3.68	1.25	0.31	-1.93	-2.25	-1.62	0.09	2.96	0.55
45	0.35	-0.50	1.45	0.02	-2.72	-0.53	-5.78	0.19	2.77	2.09

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 1a presents the item parameter estimates of the BSEB (state) BECE 2017 National Values Education (NVE) test with one dimension. The test has 45 items, and the column labelled 'a' stands for item discrimination, column 'b' stands for item difficulty, column 'd' stands for intercept and column 'c' stands for guessing parameters. To find 'b' (item difficulty), Columns 'a' and 'd' are used with the above stated formula in (Eq1). From the table above, Benue State test with unidimensionality have 34 items (items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 20, 21, 23, 24, 26, 27, 29, 31, 32, 34, 37, 38, 39, 40, 41, 43, 44 and 45) out of 45 that are good, because their discrimination index ('a') = ≥0.2 and difficulty index ('b') ranged between -2 and 2. While 11 items are bad (5 items: 17, 19, 30, 33 and 35 are too difficult ('b' = >2); while 4 items (22, 25, 36 and 42) are too simple ('b' = <-2) and 6 items (19, 22, 28, 30, 35 and 42) have low discrimination power ('a' = < 0.2). The guessing value of the items ranges from 0.00 to 0.48. Therefore, out of 45 items of the BSEB BECE 2017 NVE test, 34 items are good to measure examinees' proficiency, while 11 items are bad (five items are too difficult, and six items are too simple).

Table 1a also presents the item parameter estimates of NECO BECE 2017 National Values Education (NVE) test with two dimensions. It also has 45 items, the column labelled 'a1' stands for ability in each item under the 1st dimension (F1), the column labelled 'a2' stands for ability under the 2nd dimension (F2). The column labelled 'Mdisc' stands for item discrimination, the column labelled 'Mdiff' represents item difficulty, while column 'c' stands for guessing parameters. To find 'Mdisc-' a1 and a2 are used as indicated (Eq2) above. Also, to find 'Mdiff', column'd' and 'Mdisc' are used as indicated under (Eq3) above. From the table above, 40 items (items 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, and 44) out of 45 have good parameters because their discrimination index ('Mdisc') = ≥0.2 and difficulty index ('Mdiff') range between -2 and 2. While five items are bad, (2, 7, 16, 33 and 45) are too difficult ('Mdiff' = >2), while no item is too simple ('Mdiff' = <-2) and no items have low discrimination power ('Mdisc' = <0.2). The guessing value of the items ranges from 0.00 to 0.38. Therefore, it means that out of 45 items of the NECO BECE 2017 NVE test, 40 items are good to measure examinees' proficiency, while five items are bad (too difficult).

Table 1b

Summary of Difference in Item Parameters between BSEB and NECO BECE National Values Education 2017

Test	N	No. of Good Items	No. of bad Items
State (BSEB)	45	34	11
NECO	45	40	5
Difference		6	6

Table 1b presents the summary of differences in item parameters of BSEB andNECO BECEs National Values Education 2017. Out of 45 multiple choice items, BSEB BECE 2017 has 34 good items while NECO has 40 good items because of their acceptable discrimination and difficulty indices ('a' or 'Mdisc.' = ≥ 0.2 'b' or 'Mdiff.' range from -2 to 2). BSEB BECE 2017 has the highest number of bad items (11 items) compared to NECO BECE 2017 (five bad items), with the difference of six good and six bad items.

1. Is there a significant mean difference in the quality of item parameters (Discrimination and Difficulty) of state and NECO BECE National Values Education 2017 terms of:

a. Difficulty Parameter?

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 2a

T-test Analysis of difference Between Item Difficulty Parameter of BSEB and NECO BECE National Values Education 2017

Parameter	BECE	Item	$\overline{X}$	S.D	S.M.E	df	t	p	Remark
Difficulty	BSEB	45	.6976	2.72664	.40646	88	.813*	0.003	Sig
	NECO	45	1.0422	.81260	.12114				

Table 2a shows the result of the analysis of the difference between item difficulty parameters of BSEB and NECO BECE 2017 national values education using t-test statistics. The result shows that there is a statistically significant difference in item difficulty parameter of state (BSEB) and NECO BECE 2017 national values education in Benue State (t_{88} = .813 , p = 0.003). This implies that BECE items in national values education developed by BSEB are less difficult compared to those developed by NECO in 2017. Students also perform better in state BECE than NECO BECE because of the test item's difficulty parameters.

b. Discrimination Parameter?

Table 2b

T-test Analysis of difference Between Item Discrimination Parameter of BSEB and NECO BECE National Values Education 2017

Parameter	BECE	Item	$\overline{X}$	S.D	S.M.E	df	t	p	Remark
Discrimination	BSEB	45	2.0596	2.79764	.41705	88	-1.661	.079	Not Sig
	NECO	45	4.3393	8.76911	1.30722				

Table 2b shows that there is no statistically-significant difference in item discrimination parameters of Benue State (BSEB) and NECO BECE's 2017 National Values Education ($t_{88} = -1.661, p = .079$). This implies that both BSEB and NECO BECE's 2017 items in National Values Education could not significantly discriminate between high and low ability test takers in Benue State. It also means that students' better performance in state BECE than NECO is not due to the test item discrimination parameters.

Discussion of Findings

The study's findings on research question one revealed there is a significant difference in the item parameters (difficulty and discrimination) of BSEB and NECO BECE 2017 National Values Education multiple choice items. The study discovered that the BECE multiple choice items developed by NECO in 2017 have high difficulty parameter compare to that of BSEB BECE 2017. The data model fit assessment of 1PL, 2PL and 3PL models revealed that the data set fitted three parameter logistics (3PL) model, based on the reason of parsimony. The result in Tables 1 reveals that NECO BECE 2017 National Values Education multiple choice items with two dimensionality have more items with high and moderate discrimination and difficulty indices to measure students' ability in the subject compared to state (BSEB) 2017 which have more simple items with low discrimination ability. The result reveals that the variation in examinees' performance in BSEB and NECO BECE's 2017 National Values Education to some extent was accounted for by the quality of item parameters developed by the two examining bodies. This clearly indicates that the BECE items developed by BSEB were not subjected to psychometric analysis for validity purposes.

This finding is in line with Oluseyi (2018), who found that students in Ekiti performed better in unified examination mathematics than West African Certificate Examination mathematics because the state items difficulty indices were lower than WACE. Also, it can be inferred from the findings that BSEB developed simple test

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

items that facilitated students' high grade while too difficult NECO items are responsible for students' low performance in National Values Education (NVE) in Benue State. The current study findings corroborated that of Ayanwale, Adeleke and Mamadelo (2018), who carried out the assessment of item statistics estimates of state and NECO BECE 2017 Mathematics multiple-choice items in Osun State using CTT and IRT approach and found that 33 items under CTT and 12 items under IRT had poor discrimination and difficulty indices and needed to be deleted. The researchers further found that NECO, as the national examining body in Nigeria, does not adopt IRT approach to validate her test items. This reaffirms the current study findings that NECO, as a national examining body, also have poor test items and needs to improve their test quality using the IRT approach for test validation.

The present study finding is also in line with Adedoyin and Mokobi (2013), who investigated the use of IRT psychometrics to examine the quality of junior certificate mathematics multiple-choice test items 2010 in Botswana and discovered that 12 items were poor items,ten items were classified as fairly good items that needed modification, while only one item was good. The present findings also agreed with Adegoke (2013), who compared item statistics of physics achievement test using the two theories (CTT and IRT) framework on 60 items and found more bad items using IRT 2PL model than using CTT. This further justifies why test developers at all levels should employ only IRT in their test development process because it produces better results in item parameters and ability estimates.

The study findings on research question two revealed a statistically significant difference between the item difficulty parameters of the state (BSEB) and NECO BECE's 2017 National Values Education multiple-choice items. The result in table 2a revealed that the state BECE 2017 National Values Education items with one-dimensionality have low difficulty parameter (difficulty indices) that enable students with low ability to pass compared to NECO BECE 2017 National Values Education items. The application of IRT in modelling testee's ability with item parameters revealed that Benue State Examinations Board has been using items with low item difficulty parameters in measuring students' ability in BECE subjects. This implies that BSEB does not follow the right test development process in the construction of the BECE items, particularly in National values Education.

The study findings also revealed that there is no statistically-significant difference in item discrimination parameters of BSEB BECE 2017 and NECO BECE 2017 National Values Education. The finding implies there is no significant difference in the ability of the test items developed by both state and NECO to discriminate between the high and low ability students (high and low achievers) in BECE. The inability of state BECE items to significantly differ from NECO BECE items to discriminate between the high and low ability learners creates doubt concerning the validity of BECE conducted by the two public examining bodies. This is because, for test items to be judged valid, such items must have acceptable difficulty and discrimination indices. The findings of the present study may not mean that NECO do not validate BECE items before administration, but it could be that the examination board uses Classical Test Theory (CTT) approach to estimate the validity of their test but not IRT approach as used by the researchers in this study.

The present study findings agreed with Ochima (2014), who observed that many public examinations bodies in Nigeria do not establish the appropriate psychometric properties of their achievement test before administration. Chima further stated that many examinations' boards validate their test items through moderation but not through the process of psychometric analysis.

The result of the present study contradicts that of Toyin (2015), who found that item parameters (discrimination and difficulty indices) using CTT approaches were appropriate (ranges 0.25 – 0.75) and similar across NECO, WAEC and NABTEB 2010 test. The study further discovered that the validity of test batteries used was fairly good as well as other psychometric

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

characteristics such as distractors across the public examination being studied. The finding supports Obemeata (2000) and Ochima (2015) findings that examiners need to properly develop achievement test before test administration to enhance the quality of assessment instruments in terms of test item parameters.

Recommendation

The following recommendations were made based on the findings:

- To have suitable test item parameters, BSEB should from time to time organize seminars, workshops, etc, to train her staff and test developers on the test psychometrics. This will improve the quality of BECE item for good quality measurement of learners' ability at the Basic Education level.
- State government examinations boards should endeavour to appoint or recruit qualified staff to pilot the affairs of State BECE effectively. The appointment of state examination board officials (Registrar/Secretary and Directors in charge of psychometric test unit) with professional knowledge and skills in test and measurement will help BSEB to develop test with good item parameters (difficulty and discrimination indices).
- Item Response Theory (IRT) approach should be used in establishing test psychometric properties by both BSEB and NECO. The use of IRT for validation of test items enhances the quality item parameters.
- Federal Ministry of Education should set up a department that will be responsible for quality assurance of BECE items developed particularly by state examinations boards across Nigeria.

Conclusion

This study found that the quality of Benue State Examinations Board's BECE item parameters are not the same with NECO BECE. It is also concluded that even NECO, as a national examining body, does not have the expected quality of BECE items that properly discriminate between high and low ability learners. This could be that the examining bodies adopt CTT approach for test validation. Therefore, the quality of BSEB and NECO BECE item difficulty and discrimination parameters accounted for variation in students' performance in Benue State.

References

Adamoah, Y.K., &Acquah, J. (2016). Determinants of students' performance in basic education certifcate examination (BECE) in the upper east region of Ghana: A Case Study of Kassena-Nankana West District. *American Journal of Research and Communication,* 4(10)91-107.

Adedoyin, O. O. (2010). Investigating invariance of person parameter estimates based on classical test and item response theories. *International Journal of Education* 2(2) 107–113.

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

Ayanwale, M. A., Adeleke, J. O. &Mamadelo, T. I. (2018). An Assessment of item statistics estimates of basic certificate examination through classical test theory and item response theory approach. *International Journal of Educational Research Review (IJERE)* 3(3) 93–105. www.ijere.com

Demars, C. (2010).Item response theory. Oxford University Press. *The Canadian Journal of Program Evaluation* 27(1) 126–129

Emaikwu, S.O. (2013). *Fundamentals of educational research and statistics*. Eagle-Dove publishers.

Fakeye, D.O. (2014). English Language proficiency as a predictor of academic achievement among EFL Students in Nigeria.*Journal of Education and Practice*, 5(9) 38 – 41

Hambleton, R.K. & Jones, R. V. (1993).Comparison of classical test theory and item response theory and their application to test development. *Educational Measurement: Issues*

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

and Practices,12(3) 38-47

Hambleton, R.K. & Rodgers, J. (1995).Item bias review.*practical assessment, research and evaluation*,4(6)155-186. Retrieved 18th November, 2012 http://pareonline.net/gettvn.asp?

Mokobi, T. &Adedoyin, O.O. (2014). Identifying location baised item in 2010 Botswana junior certificate examination Mathematics Paper one using the item response characteristics curves. *International Review of Social Sciences and Humanities*7.2: 63-82

Nenty, H.J. (2015, November 23-27). *Conjugal relationship between research and measurement*: A Keynote Address Delivered at the 1stEARNiA Conference in Cameroon.

Obemeata, J.O. (2000). *Principles of Essay and Multiple Choice Test Construction*, Stirling-Holden Publishers (Nig) Ltd.

Ochima, C.O. (2015). Assessment of Benue State Examinations Board Test in Social Studies. *Journal of Practical Assessment, Research and Evaluation (JPARE)*, 10(4) 155 - 165

Oluseyi, A.E. (2018). Equating the State Unified and West African School Certificate Mathematics Examinations Items.*American International Journal of Contemporary Research* 8(4) 100 – 106.

Reid, W.A. (2003). Curriculum as instructionalized learning: implication for theory and Research. *Journal of Curriculum and Supervision* 1(19) 29 – 43

Toyin, M.D. (2015). Analyzing psychometric properties of Mathematics in public examinations in Nigeria.*Research on Humanities and Social Sciences* 5(7) 24 – 30

XinmingAn, & Yung, Y. F. (2014). *Item response theory: what it is and how you can use the IRT Procedure to Apply It*. SAS Institute Inc