TEST CHARACTERISTICS OF BIOLOGY MULTIPLE CHOICE ITEMS FOR 2015 AND 2016 NATIONAL EXAMINATIONS COUNCIL SENIOR SCHOOL CERTIFICATE EXAMINATION

[1]AYANDA, V. S. & [2]FALEYE , B. A.

[1&2]Department of Educational Foundations and Counselling,
Faculty of Education,
Obafemi Awolowo University,
Ile-Ife, Nigeria
honvictor2013@gmail.com, **bamidelefaleye@yahoo.com**

Abstract

The study investigated the item parameters of Biology multiple choice test items of the 2015 and 2016 National Examinations Council (NECO) Senior School Certificate Examination (SSCE). The study adopted the ex-post facto research design. The population for the study comprised the candidates that sat for Biology during the 2015 and 2016 NECO SSCE in Osun State. A sample of 1,457 and 2,875 candidates' responses for year 2015 and 2016 SSCE Biology test were selected using multistage sampling procedure. Data were collectedfrom Information Technology Services (ITS) unit of NECO using a hard-disk drive. Data collected were coded and analyzed using Multi-dimensional Item Response Theory (MIRT) package of R for statistical analysis. The results showed that for 2015 NECO Biology SSCE multiple choice test, four items were too difficult (average a = 2.23) while 13 of the items were too easy (with average a of 0.12). Findings on the item parameters of the 2016 NECO Biology SSCE multiple choice test showed that two items were too difficult (with average a = 15.98) while six of the items were too easy (average a = 0.20). However, all the test items possessed acceptable degree of discrimination (with b ranging between 0.70 and 4.59). The study concluded that 2015 and 2016 NECO Biology multiple choice SSCE items was multidimensional. However, 2015 NECO Biology items had more easy items than 2016 NECO Biology items while both 2015 and 2016 Biology discriminate effectively and the guessing parameters is significant.

Keywords: Item Parameters, Item discrimination, Item difficulty, Item Response Theory, Item Characteristics.

Introduction

Test developers are basically concerned about the quality of test items and how examinees respond to them. Writing test items is a matter of precision. A test item must focus the attention of the examinee on the principle or construct upon which the item is based. Ideally, students who answer a test item incorrectly will do so because their mastery of the principle or construct in focus was inadequate or incomplete. A good test instrument is expected to possess basic qualities which include validity, reliability, credibility, civility, availability and utility (Nwana, 2000, in Ukwuije, 2012). However, the validity and reliability of any test depend ultimately on the characteristics of its items. These characteristics are item difficulty, item discrimination and guessing. A test should have a well determined difficulty level (not too difficult or too simple), discriminating power (how it differentiates between good and weak students), and distractor index (how it discourages guessing).

Test theories enable the prediction of outcomes of tests by identifying parameters of item difficulty, item discrimination and the ability of test takers. According to Nwaobia (1990), item analysis (difficulty and discriminating indices) is concerned with ascertaining the worth of the test items. A good test should be able to differentiate the brilliant students from the dull ones.This can only be realizable when carefully constructed tests are constructed, administered, marked and scored. Tests that are too difficult or too simple rarely make effective evaluation possible.

The importance of difficulty indices, discrimination power and guessing index in test items cannot be overemphasized. According to Schumacher (2010), Classical Test Theory (CTT) utilizes traditional item and sample dependent statistics i.e. item difficulty and item discrimination. In classical theory the two statistics that form the cornerstone are item difficulty and item discrimination. Difficulty index actually tells us how easy the item was for the students in that particular group. The higher the difficulty index the easier the question and the lower the difficulty index the more difficult the question while discrimination power of test items on the other hand is the ability to discriminate between the brilliant students and poor students.Item response theory on the other hand, has provided a better approach to estimating item characteristics which include difficulty, discrimination and the guessing parameters.

Testing has become one of the most important parameters by which a society adjudges the products of her educational system. The essence of testing is to reveal the latent ability of examinee (Emaikwu, 2012). Testing has been fully accepted in most modern societies as the most objective method of obtaining information for decision making in schools, industries and government establishments. It is now used for admission, recruitment, promotion, placement, evaluation, guidance, research and teaching purpose among others (Emaikwu, 2011).

Tests appear therefore to be a mechanism commonly used to measure cognitive abilities an individual possesses. Test is a set of stimuli presented to obtain responses from an individual that could be used to measure certain cognitive traits or characteristics possessed. They are the instruments for measuring the affective or non-intellectual aspects of behavior of an individual. A test can be studied from different perspectives and the items in the test can be evaluated according to different theories. Two of such theories are the classical test theory (CTT) and the item response theory (IRT). These theories are the two major frameworks that are used in educational measurement to develop, evaluate and study test items. These frameworks are based on different assumptions and use different statistical approaches. They are not only to develop, evaluate, or determine the reliability and validity of tests but also to holistically improve the quality of test items. Classical Test Theory was originally the leading framework for developing and analyzing standardized tests. Later, IRT was developed to complement the role of CTT.

The major advantage of CTT is its relatively weak theoretical assumptions, which make CTT easy to apply in many testing situations (Hambleton & Jones, 1993). While CTT has proven very useful test development, the two statistics that form its cornerstones, item difficulty and item discrimination are both sample dependent. In particular, because the classical test theory model lacks information regarding how an examinee is predicted to perform on a particular item, it cannot accommodate tests that target an examinee's proficiency level (Hambleton, Swaminattham & Rogers, 1991).On the other hand, item response theory has become an important complement to CTT in development, interpretation

and evaluation of tests and test items. The interest in IRT grew out of a combination of the concerns on the limitations inherent in CTT and the availability of computing systems. IRT has strong mathematical basis and depends on complex algorithms that are more efficiently solved via computer. It describes the relationship between an examinee's test performance and the traits assumed to underlie such performance on an achievement tests as a mathematical function called item characteristics curve (ICC) (Hambleton & Swaminathan , 1995).

There are two major forms of test/examination taken in Nigeria and they are classified according to who conducts the examination. In internal examinations the items/questions are constructed by the classroom teachers and administered by the teachers in their individual institutions, while external examination is an examination conducted by national bodies and states examination bodies. In Nigeria, there exists a number of national examination bodies involved in standardized testing and they include National Examinations Council (NECO), West African Examinations Council (WAEC), National Business and Technical Examination Board (NABTEB), and Joint Admissions and Matriculation Board (JAMB). These bodies examine candidates of various backgrounds all over the country.

The essence of public examinations is to provide uniform assessment to all candidates who are exposed to a given curriculum. Public examination handles large scale testing programmes of candidates taking the examinations they registered at the same time. The councils use standard scores to report the performance of the candidates. The aims of secondary school education in Nigeria, as stated in the National Policy on Education (FRN, 2013), are to prepare the individual child for useful living in the society; and for higher education. It is at the end of six years in the secondary school that the senior secondary school certificate examinations are taken. One of the science subjects which students are expected to pass in the School certificate examinations by the students is Biology.

Several factors have been identified by researchers as being responsible for the persistent poor performance recorded in biology especially at Senior Secondary Certificate Examinations. Some of these include lack of teachers, lack of educational facilities like laboratories, textbooks, overloaded syllabuses, laziness, poor attitude and lack of interest on the part of the students, poor teaching methods by teachers, large class size. Dinah (2013). These factors amongst others have drawn the attention of researchers towards Biology as a subject in the secondary school.

The Biology test items are divided into practical, essay and multiple choice questions, and the researcher is majorly interested in investigating the Biology multiple choice items presented to candidates by National Examination Council, because the item parameters can be adjudged using multiple choice items, so as to determine the item location or item threshold parameters, the item slope as well as the lower asymptote value. Apart from the foregoing, multiple choice items are by far the most commonly used test item format in testing, whether school-based or public assessment (Popham, 2002)

The establishment of NECO, which was seen by many as an attempt to reduce the burden of WAEC and mitigate the burden of testing large number of candidates, unfortunately led to concerns by some that creditability issues would inevitably arise (Afemikhe, 2002). Research and reports over the years have shown that the level of students' performance was not steady but fluctuates yearly with poor performance having prevalence. The difference in performance may be attributed to the variation in psychometric properties of the test items, among others. Osun State Government has invested hugely to physical development of its secondary schools in recent times. Unfortunately, this has not translated to improvement in

students' performance in the SSCE. While many have attributed the candidates' poor performance to change in policies and poor remuneration, Okuntade (2017), little attention has been given to the characteristics of test items that are administered on the students. There is therefore need to analyze the item characteristics of NECO for year 2015 and 2016.

The main purpose of the study was to investigate the item parameters of the 2015 and 2016 NECO's SSCE Biology multiple choice test items. In a bid to achieve the purpose, the following research questions were raised.

a) What is the dimensionality, discrimination, difficulty and guessing parameters of the 2015 NECO's SSCE Biology multiple choice test items?
b) What is the dimensionality, discrimination, difficulty and guessing parameters of the 2016 NECO's SSCE Biology multiple choice test items?

Methodology

The study adopted the ex-post facto research design. The population of the study comprised all the candidates' that sat for Biology during the 2015 and 2016 NECO SSCE in Osun State. The population of candidates that sat for NECO SSCE for year 2015 and 2016 were 969,991 and 1,022,474 respectively while candidates that sat for Biology items in Osun State in 2015 were 4,027 and 18,922 in 2016. A sample of 1,457 and 2,875 candidates' responses for year 2015 and 2016 SSCE Biology test were selected using multistage sampling technique. The instruments used were 2015 and 2016 NECO's SSCE Biology multiple choice items. Data was extracted from NECO Master File of the Information Technology Services (ITS) Department's database on candidates who sat for NECO Biology items for year 2015 and 2016 respectively. Data collected was analysed using MIRT package of R language environment for statistical analysis.

Results

Research Question One: *What is the dimensionality, discrimination, difficulty and guessing parameters of the 2015 NECO's SSCE Biology multiple choice test items?*

To estimate the item parameters of 2015 NECO Biology, IRT measurement framework was adopted. In order to choose which of the two classes of IRT models: Unidimensional IRT models and Multidimensional IRT for the estimation of the item parameters, the dimensionality of the tests was assessed. The dimensionality of the tests was assessed using nonlinear factor analysis. To achieve this, Normal Ogive Harmonic Analysis Robust Method (NOHARM; Fraser & McDonald, 2003) software package was used. The results are presented as follows.

Table 1: Dimensionality of 2015 NECO Biology

DIM	GFI	RMSR CRITERION	RMSR	DIFF IN RMSR	REDUCTION IN RMSR	PERCENT REDUCTION IN RMSR
1	0.963874	0.063033	0.011787			
2	0.977408	0.063033	0.009321	0.002466	0.209197	20.91966
3	0.98391	0.063033	0.007866	0.001455	0.156074	15.60743
4	0.986791	0.063033	0.007128	0.000739	0.093931	9.393115

Table 1 presents the assessment of the assumption of dimensionality in 2015 NECO Biology test. The table shows that when 1-dimension was hypothesized to underlie the data set, the data showed a good fit (Goodness of Fit index, GFI > than 0.95), Root Mean Square of Residuals, RMSR, 0.01179 was less than the criterion. However, when the fitness of the 1-dimension model to the data set was compared to that of 2-dimension model to the data, Table 1 showed that more than 10% reduction in RMSR was recorded. This result showed that the 2-dimension model fitted the data better than 1-dimension model. This revealed that the data set violated the assumption of unidimensionality. To isolate the optimal number of dimensions underlying the test, the fitness of 3-dimension model to the data set was compared to that of the 2-dimension model. The table showed that more than 10% reduction in RMSR was recorded when 3-dimension was hypothesized to underlie the data set. Furthermore, the four-dimension model was fitted to the data and compared to the 3-dimension model. Although, the GFI for the 4-dimension model was greater than 0.95 and the RMSR was less than the criterion, the percentage in reduction of RMSR from 3-dimension model to 4-dimension model was less than 10%. These results showed that 3 dimensions underlie the performance of the examinees on the 2015 NECO Biology test. Consequently, the data set was model using Multidimensional Item Response Theory (MIRT) model.

To obtain the item parameters of the 2015 NECO Biology test, multidimensional 3-parameter logistic model was used. This was executed using MIRT package of R language and environment for statistical computing. The result is presented in the Table 2

Table 2: Item parameters estimates of 2015 NECO Biology test

Item	a1	a2	a3	D	c	Mdisc	Mdiff
1	-1.81	0.37	0.50	-0.87	0.21	1.91	0.45
2	-0.41	-0.35	1.40	1.38	0.00	1.50	-0.92
3	0.12	0.08	1.90	0.55	0.35	1.91	-0.29
4	-1.84	0.13	0.02	0.06	0.00	1.85	-0.03**
5	-0.55	-0.42	1.31	1.81	0.01	1.49	-1.22
6	-2.31	0.09	-0.04	0.64	0.00	2.31	-0.28
7	-2.19	0.27	-0.23	0.61	0.00	2.22	-0.28
8	-1.81	0.24	0.17	-0.17	0.00	1.83	0.09**
9	-2.48	0.32	-0.10	0.65	0.00	2.50	-0.26
10	-1.61	0.32	0.10	-0.22	0.00	1.64	0.13**
11	-1.57	0.33	0.37	-0.71	0.00	1.65	0.43
12	0.03	-0.02	-1.00	-2.60	0.00	1.00	2.59*
13	-1.85	0.35	-0.15	0.58	0.00	1.89	-0.31
14	-0.96	-0.11	1.05	0.75	0.00	1.43	-0.53
15	1.19	0.06	-0.12	-2.13	0.00	1.20	1.77*
16	-1.50	0.38	0.29	0.29	0.00	1.57	-0.18
17	-2.17	0.12	-0.47	1.34	0.00	2.23	-0.60
18	-0.84	-0.30	1.25	1.87	0.01	1.53	-1.22
19	-1.76	0.57	0.19	-0.08	0.00	1.86	0.05**
20	-0.23	0.33	1.36	0.56	0.07	1.42	-0.39
21	-1.69	0.34	-0.20	0.60	0.00	1.74	-0.34
22	-0.36	-0.18	1.25	1.15	0.00	1.32	-0.87
23	-1.52	0.79	-0.24	-0.68	0.00	1.73	0.39
24	-2.26	1.20	-0.09	-0.69	0.03	2.56	0.27
25	-2.10	0.90	0.26	-0.36	0.01	2.30	0.16
26	-1.97	0.58	0.04	0.31	0.00	2.05	-0.15
27	-2.26	0.22	-0.03	0.00	0.00	2.27	0.00**
28	-1.53	0.62	-0.01	-0.25	0.00	1.65	0.15
29	-1.13	-0.51	-0.57	-0.87	0.00	1.36	0.64
30	-2.05	0.64	-0.03	0.48	0.00	2.15	-0.22
31	-2.12	0.32	-0.23	0.61	0.00	2.16	-0.28
32	-1.96	0.37	0.11	-0.24	0.00	1.99	0.12**
33	-1.71	0.51	0.17	0.20	0.00	1.79	-0.11**
34	-1.23	-0.56	0.60	1.33	0.00	1.48	-0.90
35	-0.81	-0.17	0.78	0.68	0.03	1.14	-0.60
36	-1.23	-0.91	-0.51	-0.84	0.00	1.61	0.52
37	0.26	-0.04	-0.92	-2.18	0.00	0.96	2.27*
38	-1.87	0.29	-0.10	0.18	0.00	1.89	-0.10**
39	-0.68	-0.18	1.23	0.97	0.02	1.42	-0.68
40	-1.41	-0.54	0.12	1.51	0.00	1.52	-0.99
41	-1.19	0.07	0.69	-0.20	0.27	1.37	0.14**
42	-1.09	0.05	0.76	0.52	0.05	1.33	-0.39
43	-0.66	0.25	0.32	-0.12	0.00	0.77	0.15
44	-1.23	-0.05	0.44	0.12	0.03	1.31	-0.09**
45	-0.88	0.04	0.49	0.33	0.00	1.01	-0.33
46	-1.26	0.07	0.23	-0.03	0.07	1.28	0.03**
47	-1.34	-1.02	-0.27	-0.76	0.05	1.71	0.44
48	-1.80	-0.55	0.03	0.59	0.34	1.88	-0.31
49	-1.76	-0.60	0.23	1.44	0.01	1.88	-0.77
50	-1.40	-0.31	0.58	0.86	0.01	1.55	-0.56
51	-1.32	0.32	-0.04	-0.07	0.00	1.36	0.05**
52	-2.02	-0.16	-0.47	1.42	0.00	2.08	-0.69
53	-0.34	0.00	-0.62	-2.02	0.00	0.70	2.86*
54	1.71	0.18	0.28	0.30	0.00	1.74	0.17
55	-1.77	-1.28	-1.13	0.54	0.00	2.46	-0.22
56	-2.15	-0.26	-0.48	0.76	0.00	2.22	-0.34
57	-0.49	-1.30	-0.09	0.15	0.00	1.39	-0.11
58	0.05	-0.90	0.45	0.03	0.00	1.01	-0.03**
59	-1.24	1.18	0.00	-1.85	0.02	1.71	1.08
60	-1.96	0.00	0.00	0.12	0.00	1.96	-0.06**
Average	**-1.36**	**0.02**	**0.17**	**0.15**	**0.03**	**1.68**	**-0.01**

Table 2 presents the item parameters of the 2015 Biology test based on the three dimensions underlying the test. The column labelled a1, a2 and a3 represent the discrimination parameter of the items at dimension 1, 2 and 3 respectively. While column labelled d is the intercept or easiness parameter and c is the guessing parameter. These values were used to calculate the unidimensional equivalent of the multidimensional item discrimination

$$\text{Mdisc} = \left(\sum_{k=1}^{m} a_{ik}^2\right)^{1/2} \quad \ldots\ldots(1)$$

and difficulty parameters

$$\text{Mdiff} = \frac{-d_i}{MDISC_i} \quad \ldots\ldots(2)$$

Using Mdisc and Mdiff values as obtained in Table 2, four (12, 15, 37 and 53) of the sixty Biology test items were relatively high difficult items while 13 of the items (4, 10 19, 27, 32, 33, 38, 41, 44, 46, 51, 58 and 60) were relatively too easy. However, the result as presented in Table 2 showed that none of the items failed to effectively discriminate.

Research Question 2*: What is the dimensionality, discrimination, difficulty and guessing parameters of the 2016 NECO's SSCE Biology multiple choice test items?*

Table 3: Dimensionality of 2016 NECO Biology

DIM	GFI	RMSR CRITERION	RSMR	DIFF IN RMSR	REDUCTION IN RMSR	PERCENT OF REDUCTION IN RMSR
1	0.919237	0.028411	0.013358			
2	0.966653	0.028411	0.008583	0.004774	0.357422	35.7422
3	0.974765	0.028411	0.007467	0.001117	0.130101	13.01015
4	0.979859	0.028411	0.006671	0.000796	0.106608	10.66081
5	0.983181	0.028411	0.006096	0.000575	0.086199	8.619914

Table 3 shows the dimensionality of 2016 NECO Biology test. The table shows that when 1-dimension was hypothesized to underlie the data set, the data showed an acceptable fit (Goodness of Fit index, GFI > than 0.90), Root Mean Square of Residuals, RMSR, 0.0134 was less than the criterion, 0.0284. However, when the fitness of the 1-dimension model to the data set was compared to that of 2-dimension model to the data, the table showed that more than 10% reduction in RMSR was recorded. Additionally, a good model-data fit was recorded for the 2-dimension model (GFI = 0.97, RMSR 0.005). This result showed that the 2-dimension model fitted the data better than 1-dimension model. This revealed that the data set violated the assumption of unidimensionality. To uncover the number of dimensions underlying the test, the test data was fitted to increasing number of dimension until the percentage reduction in RMSR was less than 10%. At this point the iterative procedure stops. The table showed that the percentage reduction of RMSR on comparing 2-dimension and 3-dimension models was greater than 10%. The same trend was recorded when 3-dimension and 4-dimension models were compared. But when 4-dimension and 5-dimension models were compared, the table showed that the percentage reduction in RMSR was less than 10%.

These results showed that the 4-dimension model fitted the data set. This implies that four dimensions underlie the performance of the examinees in the test. Thus, MIRT was used in the calibration of the test.

To obtain the item parameters of the 2016 NECO Biology test, multidimensional 4-parameter logistic model was used. This was executed using MIRT package of R language and environment for statistical computing. The result is presented in the Table 4.

Table 4: Item parameters of 2016 NECO Biology test

Item	*a*1	*a*2	*a*3	*a*4	*D*	*c*	Mdisc	Mdiff
1	-1.30	0.04	0.46	-0.17	0.82	0.00	1.63	-0.51
2	4.34	-1.97	-0.97	-0.49	-6.64	0.15	4.59	1.45
3	-0.89	-1.44	-0.38	0.15	1.16	0.16	1.46	-0.80
4	-0.68	-1.74	-0.18	-0.49	0.99	0.23	1.83	-0.54
5	-1.35	-1.65	2.33	-0.60	-3.83	0.17	3.09	1.24
6	-1.08	-2.38	-0.61	0.28	0.24	0.13	2.38	-0.10**
7	-0.67	-1.42	-0.47	-0.48	1.02	0.16	1.33	-0.77
8	-0.41	-0.94	0.09	-0.71	2.15	0.00	1.32	-1.63
9	-0.38	-1.34	-0.58	-0.57	-0.11	0.17	1.04	0.10**
10	-0.79	-1.63	-0.08	-0.47	1.39	0.10	1.83	-0.76
11	-0.62	-1.42	-0.22	0.12	0.39	0.10	1.40	-0.27
12	-0.44	-1.36	-0.14	-0.53	0.74	0.13	1.44	-0.52
13	-0.67	-1.14	0.14	-0.50	1.16	0.00	1.51	-0.77
14	-0.37	-1.95	0.49	0.73	-0.33	0.14	2.33	0.14**
15	-2.04	-0.50	0.27	-0.39	0.25	0.02	2.26	-0.11**
16	-1.59	-0.28	0.39	-0.97	0.43	0.02	2.08	-0.21
17	-1.37	0.12	0.08	-0.74	-0.49	0.00	1.61	0.30
18	-0.87	-1.71	-0.41	-0.13	0.15	0.22	1.69	-0.09**
19	-1.04	-1.87	-0.50	-0.03	0.50	0.16	1.88	-0.26
20	-0.68	-1.41	0.14	-0.58	1.07	0.06	1.75	-0.61
21	-0.13	-1.01	0.07	-0.52	1.27	0.00	1.20	-1.06
22	-0.60	-1.56	0.03	-0.52	1.73	0.12	1.77	-0.97
23	-0.41	-1.30	0.36	-0.50	1.34	0.01	1.68	-0.80
24	-0.33	0.35	0.73	0.19	-2.77	0.00	1.31	2.11*
25	-0.40	-1.17	0.25	-0.56	0.56	0.01	1.53	-0.36
26	-0.40	-1.24	0.57	-0.67	1.94	0.01	1.81	-1.07
27	2.61	-1.41	-1.33	1.43	-4.11	0.07	2.86	1.44
28	-0.67	-1.69	0.15	-0.69	0.91	0.16	2.02	-0.45
29	-0.58	-1.82	0.23	-0.77	1.75	0.14	2.17	-0.80
30	1.32	-2.15	1.52	0.03	-0.57	0.20	3.06	0.19
31	-0.29	0.00	0.40	-0.77	-0.37	0.00	1.22	0.30
32	-13.85	-5.86	-6.50	8.39	-29.19	0.10	16.84	1.73*
33	0.49	-1.75	0.64	0.22	-1.04	0.21	2.15	0.48
34	-0.49	-1.55	-0.05	-0.14	1.11	0.05	1.60	-0.70
35	-3.22	0.14	0.07	-0.15	-1.23	0.10	3.25	0.38
36	-1.29	-2.18	-0.67	0.17	0.48	0.20	2.26	-0.21
37	-17.08	-3.10	4.07	8.67	-17.66	0.17	19.61	0.90
38	-0.80	-1.64	-0.52	0.26	0.56	0.16	1.54	-0.37
39	-0.45	-1.19	-0.31	0.10	0.58	0.00	1.01	-0.57
40	-0.51	-1.57	-0.05	-0.14	0.57	0.08	1.63	-0.35
41	-0.49	-2.21	0.52	1.33	-0.91	0.15	2.81	0.32
42	-0.85	-1.89	0.00	-0.20	0.91	0.19	2.08	-0.44
43	-0.26	0.52	0.99	-0.41	-1.21	0.00	1.58	0.77
44	-0.55	-1.19	0.07	-0.26	0.70	0.06	1.39	-0.50
45	-0.55	-1.25	0.37	-0.28	1.65	0.00	1.63	-1.01
46	-0.56	-1.52	-0.32	-0.26	0.47	0.09	1.43	-0.33
47	-0.96	-1.79	0.39	-0.29	1.23	0.22	2.23	-0.55
48	-0.65	-1.24	0.87	-0.14	2.50	0.00	1.93	-1.30
49	-1.21	0.37	0.66	-0.45	0.44	0.00	1.77	-0.25
50	0.03	-1.11	0.95	0.53	0.97	0.00	1.85	-0.52
51	-0.23	-1.09	0.14	-0.22	1.50	0.00	1.25	-1.20
52	-0.57	-1.88	-0.38	0.19	-0.62	0.24	1.77	0.35
53	-0.72	-1.65	-0.69	0.24	0.10	0.21	1.38	-0.08
54	-0.88	-1.94	-0.39	0.08	-0.01	0.13	1.94	0.00**
55	-0.38	-1.09	0.11	0.20	1.19	0.00	1.26	-0.94
56	-0.62	0.38	1.21	0.21	-1.12	0.00	1.73	0.65
57	-0.49	-1.48	0.03	-0.23	0.95	0.11	1.60	-0.59
58	-0.41	-0.79	0.42	0.00	1.86	0.00	1.28	-1.46
59	-0.60	-1.20	0.00	0.00	1.01	0.01	1.34	-0.76
60	-1.35	0.00	0.00	0.00	-0.37	0.02	1.35	0.27
Average	-1.04	-1.33	0.05	0.13	-0.56	0.09	2.38	-0.19

Table 4 presents the item parameters of the 2016 Biology test based on the four dimensions underlying the test. The columns labelled *a*1, *a*2, *a*3 and *a*4 represent the discrimination parameter of the items at dimension 1, 2, 3 and 4 respectively. While column labelled *d* is the intercept or easiness parameter and *c* is the guessing parameter. Using Mdisc and Mdiff values as obtained in Table 4, two (24 and 32) of the sixty 2016 Biology items were relatively high difficult items while six of the items (5, 9, 14, 15, 18 and 54) were relatively too easy. However, the result as presented in Table 4 showed that none of the items failed to discriminate effectively.

Discussion of Findings

The study investigated the characteristics (item parameters) of Biology test items of the 2015 and 2016 NECO's SSCE. The results from the assessment of the assumption of dimensionality in 2015 NECO Biology multiple test items using nonlinear factor analysis showed that 1-dimension, 2-dimension and 3-dimension models were hypothesized to underlie the data set, the data showed a good fit using Goodness of Fit index (GFI), Root mean square of residuals (RMSR). However, when the 1, 2 and 3-dimension models were compared based on their GFI and RMSR values, the result showed that 3-dimension underlie the performance of the examinees on the 2015 NECO Biology test. Meaning the test has more than a latent trait in which it checks. Consequently, the data set was model using Multidimensional item response theory (MIRT) model. The results of the item parameters of the 2015 NECO Biology test using the multidimensional 3-parameter logistic model showed the values of discrimination parameter, intercept or difficulty parameter and the guessing parameter of the items in the test. Findings showed that four out of the sixty 2015 Biology items were relatively difficult items while thirteen of the items were relatively too easy. However, findings showed that none of the items failed to effectively discriminate.This finding is supported by the finding of Anigbo (2015) which say the psychometric properties (difficulty, discrimination indices and distracter patterns) of NECO Economics Objective test packages for 2009, 2010 and 2011 possess acceptable qualities. In agreement with Ogbonna (2015), the National Examinations Council (NECO) should be applauded to great extent for good work done and maintenance of standards.

The assessment of the assumption of dimensionality in 2016 NECO Biology test using nonlinear factor analysis showed that 1-dimension, 2-dimension and 3-dimension models were hypothesized to underlie the data set, the data showed a good fit using Goodness of Fit index (GFI), Root Mean Square of Residuals (RMSR). However, when the 1, 2 and 3-dimension models were compared based on their GFI and RMSR values, the result showed that 3- dimension underlie the performance of the examinees on the 2016 NECO Biology test. Consequently, the data set was model using Multidimensional item response theory (MIRT) model. Findings on the item parameters of the 2016 NECO Biology test using the multidimensional 3-parameter logistic model also showed the values of discrimination parameter, intercept or difficulty parameter and the guessing parameter of the items in the test. Moreover, two out of the sixty 2016 Biology items were relatively difficult items while six of the items were relatively too easy. However, the result also showed that none of the items failed to effectively discriminate. This finding is contradicted by the finding of Olatunji (2007) that WAEC SSCE multiple choice tests have better discriminating indices than NECO SSCE multiple choice tests in Economics.

Conclusion/Recommendation

The study concluded that 2015 and 2016 NECO Biology multiple choice SSCE items was multidimensional. However, 2015 NECO Biology items had more easy items than 2016 NECO Biology items while both 2015 and 2016 Biology discriminate effectively and the guessing parameters is significant**.** The study recommended that the examination bodies using multiple-choice test instruments should ensure that the dimensionality, difficulty, discrimination and guessing parameter are consistent over years so as not to be biased to candidates who take the exam in a particular year with higher difficulty.

Reference

Abiri, J. O. (2006). Elements of Evaluation Measurement and Statistical Techniques in Education. Ilorin: Library and Publication committee, University of Ilorin, Nigeria.

Afemikhe, O. A. (2002). *Assessment Policy in the first twelve years of formal Educational in Botswana and Nigeria. Mosenodi 14*(1), 53-60.

Anigbo, L. C. (2015). *Item Analysis of National Examination Council Senior School Certificate Examination Economics Objective Tests. International Journal of Innovative Education Research.3(4):23-30, Oct – Dec. 2015. ISSN: 2354-2942*

Dinah C. S. (2013). *Factors which influence academic performance in biology in Kenya: a perspective for global competitiveness. International Journal of Current Research, 5(12), pp 4296-4300*

Emaikwu, S. O. (2011). Evaluation of student's ability in schools. Being a paper presented at a workshop on teaching practice on Friday, 29th July in the College of Agricultural & Science Education, Federal University of Agriculture Makurdi, Benue State.

Emaikwu, S. O. (2012). *Conducting examination in Nigerian Universities: The quandary of missing scripts and scores*. *International journal of Academic research in progressive education and development. October 2012, Vol. 1, No. 4. ISSN: 2226-6348*

Federal Republic of Nigeria (2013). *National Policy on Education*. Lagos: NERDC Press.

Hambleton, R. K. and Jones, R. W. (1993). *Comparison of classical test theory and item response theory and their applications to test development. Educational measurement:Issues and practice,* 12(3): 38-47.

Hambleton, R. K. and Swaminatthan, H. (1995). *Item response theory: Principles and application.* Boston: Kluwer.

Hambleton, R. K., Swaminatthan, H. and Rogers, H. J. (1991). *Fundamentals of item response theory:* Newbury Park, CA: Sage.

Nwana O.C. (2000). *Evaluation of schools and college examinations in the Nigerian education system.*2002 Fafunwa foundation lecture.

Nwaobia, E. M. (1990). Monograph on *Introduction to Advanced Measurement and Evaluation.Nssuka:* University of Nigeria.

Ogbonna, A. G. (2015). Content Validity and Item Analysis of National Examinations Council Senior Secondary School Certificate Examination Economics Achievement Test. *M.Sc. (Ed) Dissertation*. Enugu State University of Science and Technology, Enugu

Olatunji, D. S. (2007). *Effects of Number of Options on Psychometic Properties of Multiple Choice Tests in Economics*. Unpublished M.Ed Thesis, University of Ilorin, Ilorin.

Popham, W. J. (2002). *Classroom Assessment: What Teachers Should Know.* Boston, M.A.: Allyn and Bacon.

Okuntade, T. F. (2017). Variables responsible for poor scholastic execution in Osun State optional schools: Perception of 2014 Batch A and C corps Members posted to Ilesha East: Grin.com,

Ukwuije, R. P. I. (1996). *Test and Measurement for Teachers.* Port Harcourt: CELWIL Publishers, 107 – 110.