RESEARCH ARTICLE

COMPARATIVE VALIDATION ANALYSIS OF BIOLOGY MULTIPLE-CHOICE TESTS OF THE NATIONAL EXAMINATION COUNCILS IN SOUTHWESTERN NIGERIA FROM 2015 TO 2017

Taiwo Oluwafemi AJEIGBE
Department of Educational Foundations and Counselling,Faculty of Education
Obafemi Awolowo University, Ile-Ife, Nigeria
taiaje@oauife.edu.ng or ajetaiolu@gmail.com

Abstract

The study compared the validity and reliability of Biology multiple-choice tests conducted by the National Examination Councils (NECO) across six states of the Southwestern Nigeria for three consecutive years (2015 to 2017). It also compared parameters a, b, and c of the NECO Biology Multiple-choice examinations across the three years. These were with a view to ascertaining the stability of the item parameters of Biology Multiple-choice items used by the National Examination Council to assess the students. The study adopted ex-post facto research design, since the secondary data was used for the study. The population for the study comprised 363,807students who sat for the NECO Biology paper III from 2015 to 2017 in Southwestern Nigeria. A proportionate sample of 10% was taken across the states from 2015 to 2017. This consisted of 36,383 {(Male=17641 (48.5%) and Female=18742 (51.5%)} students. The instrument for the study consisted of 60 NECO Biology paper III used in 2015 to 2017 and the responses to those questions. The responses to those questions were coded as "1" for right option and "0" for wrong option. The data collected was analysed using Exploratory Factor Analysis (EFA) for validity, Cronbach's Alpha and Spearman Brown Split-Half to assess the reliability, while one-way analysis of variance was use for comparing students' scores. The results of the component matrix revealed factor loading ranges from 0.611 to 0.721, which was above the minimum value of 0.6 recommended, as well as a dominant factor underlying the construct was recorded variance greater than 20% in two years of the three consecutive years reviewed in the study. The results also showed high coefficients of reliability using Split-Half yielded 0.94 (94%), 0.91 (91%), 0.95 (95%) and Cronbach's Alpha yielded 0.92 (92%), 0.91 (91%) and 0.93 (93%) for the three years respectively. Finally, there was no statistical significant difference in the item difficulty ($F_{(2,177)}$= 0.979; p>0.05), discrimination ($F_{(2,177)}$= 0.573; p>0.05) and guessing ($F_{(2,177)}$= 0.984; p>0.05) parameters across the three years. The study concluded that the NECO Biology Paper III was consistently stable in terms of validity, reliability and parameter estimates.
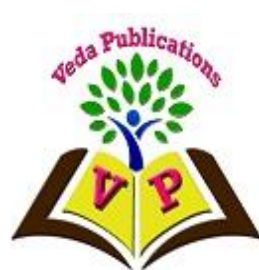
Introduction

Tests are usually means of assessing students' cognition at every level of educational endeavour. At primary, secondary and tertiary education, the use test as an instrument for assessing learning outcome cannot be overemphasized. As much as test is important to the teachers, it is as well important to the students. On the part of the teachers, it serves several purposes such as, assessing teaching methods, diagnostic, placement/promotion, and so on. The students on the other hand will not likely compromise the important of test, since their academic progress depend on their performance in test. Measurement error in test score interpretation may be eminent especially when student are conscious of the fact that academic grade is not attached to given correct or incorrect responses to test items (DeMars, 2000) and such behaviour may threaten the validity of the interpretation afterwards (DeMars, 2000, Meijer and Sijtsma, 2001).

Basically, test can be classified in to two, namely essay and objective test. The focus of this paper is multiple-choice test, which is one of the types of objective test. The multiple-choice test has gained wider popularity and acceptability in every facet of education both locally and internationally. It popularity could be attributed to its widely sampling of content taught during a period of instruction. The use of multiple-choice test is not limited to educational settings, firms and companies also make of multiple-choice test in aptitude testing, promotion test, in civil commission, it is very relevant. It popularity has made assessment easier in various subjects such as Mathematics, English Language, Chemistry, Physics, Biology, Economics, Agricultural Science, Yoruba, Igbo, Hausa, Government, Commence to mention but a few.

Developing multiple-choice tests require a lot of skills, apart from the rigour embedded in it. In developing multiple-choice, it important to identify the stem structure, the key and the distractors. The stem should not be too worded, no ambiguity and should not give clue to the testees in locating the key. Also, the distractors should be not the redundant, that is it should be plausible, so as to really differential between the high and low performing students. All of these are very germane to developing valid and reliable multiple-choice test items in any subject-matter.

Ajeigbe,T.O.

Developing valid and reliable multiple-choice test items is of essence to test developers, especially tests that are met for public consumption. This is important to avoid measurement error that may be inherent in poorly constructed multiple-choice items. The process of developing valid and reliable multiple-choice items is the responsibility of any examination bodies. For instance, in Nigeria, there are many examination bodies that make use of multiple-choice tests in assessing students, such as the National Examination Councils (NECO), West African Examination Council (WAEC), National Business and Technical Examinations Board (NABTEB) and Joint Admissions and Matriculation Board (JAMB) to mention but a few.

The National Examination Council on which this study is predicated was the fallout the act of the Abdulsalam Abubakar military administration, promulgation of a decree, in April 1999, that created the National Examinations Council (NECO). The creation of the examination body (NECO) was not free of criticisms and controversies. Some stakeholders in the education sphere applauded its arrival, while some were full of scepticisms about its strength to conduct valid and reliable examination that will be accepted world-wide. The mandate of NECO was to take over the responsibilities of the National Board for Educational Measurement (NBEM), which had been created, in 1992, by the Ibrahim Babangida administration, although its enabling decree was promulgated in 1993. However, the conduct of the Senior School Certificate Examinations (SSCE), which had, hitherto, been the exclusive preserve of the West African Examinations Council (WAEC) was made an additional responsibility of the new examination outfit. Hence, NECO was to take exclusive charge of the conduct of the SSCE for school based candidates while WAEC was to take charge of the same examination for private candidates.

Assessing construct validity of the Biology multiple-choice items across the years under review in this study, the factor analysis as proposed by Nunnally and Berstein (1994), being an advanced correlational statistical procedure is which used to identify unobserved or latent variables called factors which are predicted by a theory. He further ascertained the factor analysis is most commonly used in the development of measuring devices in which the goal of the researchers is either to confirm (confirmatory factor analysis) or identify (exploratory

Ajeigbe,T.O.

factor analysis) factors included within a measure which is said to operationally define a theory (Nunnally & Berstein , 1994).

An important aspect of the item analysis in factor analysis is the data adequacy which is very germane and precursor to factor analysis. Item analysis is the process of examining students responses to individual test items in order to assess the quality of those items and the test on a whole (Mehta, 2011). In this study, Kaiser-Meyer-Olkin and Bartlett's test of Sphericity were used to establish data suitability. Kaiser-Meyer-Olkin was used to assess the stability of the NECO Biology multiple-choice test for the period of three consecutive years, the Bartlett's test of Sphericity, which is a correlation matrix, and determinant score for detecting the appropriateness of the data set for functioning factor analysis (Pett, Lackey, & Sullivan, 2003). The KMO values ranges form 0 to 1. As a rule of thumb according   , interpreting the statistic:

- ✓ KMO values between 0.8 and 1 indicate the sampling is adequate.
- ✓ KMO values less than 0.6 indicate the sampling is not adequate and that remedial action should be taken. Some authors put this value at 0.5, so use your own judgment for values between 0.5 and 0.6.
- ✓ KMO Values close to zero means that there are large partial correlations compared to the sum of correlations. In other words, there are widespread correlations which are a large problem for factor analysis.

Item response theory is a modern mathematical theory that allow prediction of students' test performance from the analysis of individual item based on trait and the parameters that make constituted the items in a test (Hambleton and Swaminathan,1985). Item response theory is scientific procedure that could be employed to analyse responses to test items with a view to improving measurement accuracy and reliability of score interpretation. As part of the method of improving validity of multiple-choice test, unidensionality assumption is centred around finding a commonly dominant factor that is clearly exist among other items in a test (Hambleton & Jones, 1993). A test is said to be unidimensional if the test shows evidence that a single factor is underlying the construct that is being measured (Lumseden, 2007).

The theory incorporates measurements assumptions about examinee, item and test performance and how this performance relates to knowledge as measured by individual items on a test (Ojerinde, Popoola, Ojo, & Onyeneho, 2012). There are four (1PL, 2PL, 3PL and 4PL) IRT models which are considered appropriate for dichotomous item response data (e.g., true and false questions or multiple-choice questions) to assess student academic performance in large-scale assessments and make informed decisions in the educational system (Baker, 1992, Hambleton, Swaminathan and Rogers, 1991; Harwell, Baker and Zwarts, 1998; and Lord, 1990). The 1PL, 2PL and 3PL IRT models preceded the 4PL model. In the 1PL and 2PL models, the probability of passing ranges between 0 and 1 as the examinee's ability (denoted as θ) is rated between $-\infty$ and ∞. In multiple-choice tests, however, the actual probability of making a correct choice does not approach 0 even for low-ability students. Birnbaum (1968) introduced a lower asymptote to model the situations in which examinees either make a random guess or answer on the basis of their knowledge. In the same manner, Barton and Lord (1981) introduced an upper asymptote parameter, expressed by the lowercase d, into the 3PL model, resulting in the 4PL model:

$$\textbf{3PL: } P(\theta) = c + (1-c)\frac{1}{1 + e^{[(-7.702a(\theta - b)]}} \tag{1}$$

$$\textbf{4PL: } P(\theta) = c + (d-c)\frac{1}{1 + e^{[(-7.702a(\theta - b)]}} \tag{2}$$
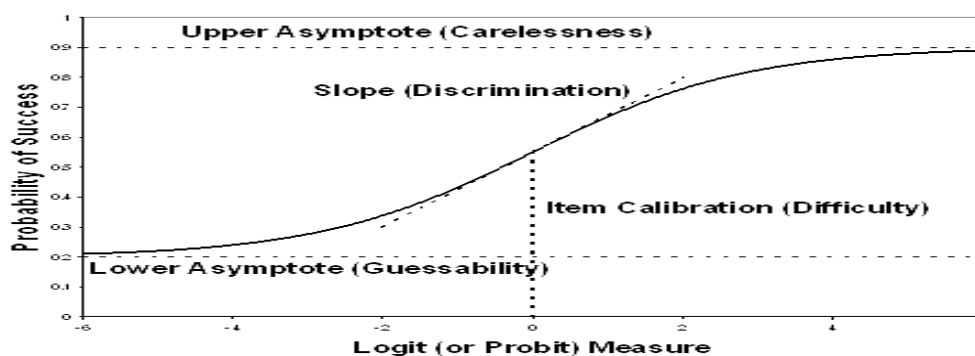
Figure 1: 4-PL IRT Item Characteristic Curve

In Figure 1, the lower asymptote c (the guessing parameter) represents the probability that an examinee with extremely low ability will correctly answer an item with difficulty b. The parameter a is called a discrimination parameter, which allows for discrimination among the examinees for each item (Harvey & Hammer, 1999). Similarly, according to Lord's 4-PL model (Barton & Lord, 1981) incorporates an upper asymptote parameter for item-specific

"carelessness". Here is a "carelessness" model, written in log-odds format, with d_i as the upper asymptote.

Statement of the Problem

In this study, it is of essence to establish if the unidimensional assumption of the IRT holds in the NECO Biology multiple-choice test items across the three years under review as evidence of validity of the examination. Also, validation of test instrument is germane to developing valid and reliable items that will be capable of differentiating between brilliant and poor students, as well being fair to all category of students. Most researches focused on factors that tracable to students' attitude, understanding subject contents, teachers factors, school factors, parental factors are usually associated with students performance in WAEC, NECO but attention needed to be directed towards examining the parameter estimates of the test items for validity seek. To further provide empirical evidence on the credibility of the items used by the examination body like NECO. Hence, the need to trace the trend of stability of multiple-choice items in NECO such as Biology Paper III is of essence. It is therefore against this background, the researcher consider it necessary to comparatively study one of the subjects (Biology) of NECO over a period of three years, to further ascertain its credibility, consistency, validity and reliability of the items.. The choice of the subject was congenital out of the fact that empirical studies are limited with respect to comparatively establishing the validation of the Biology multiple-choice test items.

Purpose of the study

The purpose of this study was to investigate the extent to the National Examination Council (NECO) Biology multiple-choice tests are valid across three years consecutively in Nigeria. Specifically, the study;

1. compare the validity of NECO Biology multiple-choice tests across the three years (2015 to 2017);
2. compare the reliability coefficients of NECO Biology Multiple-choice examinations across the three years (2015 to 2017), using Item response theory approach; and

3. compare the item parameters (a,b, and c) of NECO Biology Multiple-choice examinations across the three years (2015 to 2017) Southwestern Nigeria.

Research Questions

The following research questions were postulated for the study.

1. How comparable is the validity of NECO Biology multiple-choice tests across the three years (2015 to 2017)?

2. How comparable are the reliability coefficients of NECO Biology Multiple choice examinations across the three years (2015 to 2017)?

3. Are the item parameters (a,b, and c) of NECO Biology Multiple-choice examinations across the three years (2015 to 2017) comparable?

Research Hypotheses

The following research hypotheses were generated for the study.

1. The scores of male and female students are not comparably significant in the 2015 Biology Multiple- choice across the States in the Southwestern Nigeria.

2. The scores of male and female students are not comparably significant in the 2016 Biology Multiple- choice across the States in the Southwestern Nigeria.

4. The scores of male and female students are not comparably significant in the 2017 Biology Multiple-choice across the States in the Southwestern Nigeria.

Methodology

The study adopted ex-post facto research design. Secondary data was used for the study, hence ex-post design was considered appropriate. The population for the study comprised students who sat for NECO Biology paper III Female=188462 (51.8%) students for NECO Biology paper III from 2015 to 2017 (details in Table I). A proportionate sample of 10% was taken across the states from 2015 to 2017. This consisted of 36,383 (Malefrom 2015 to 2017 in Southwestern Nigeria. It comprised 363,807 (Male=175345 (48.2%) and =17641 (48.5%) and Female=18742 (51.5%) students (see Table 2). The instrument for the study consisted of 60 NECO Biology paper III used in 2015 to 2017 and the responses to those questions were obtained from the NECO base with permission from the NECO headquarters in Minna, Niger State, Nigeria. The data collected was analysed using Exploratory Factor Analysis (EFA) for validity, Cronbach's Alpha and Spearman Brown Split-Half to assess the reliability, while t-test was use for comparing students' scores.

Table 1:

Distribution of Students who Sat Biology Paper III from 2015 to 2017

| Year	Sex	Southwestern States						Total
		Ekiti	**Lagos**	**Ogun**	**Ondo**	**Osun**	**Oyo**	
2015	Male	3593	18016	6331	5500	8212	16616	58268
	Female	4304	18980	6489	6147	8697	17830	62447
Total		7897	36996	12820	11647	16909	34446	120715
2016	Male	2752	19279	6985	6709	9118	15447	60290
	Female	3111	20776	7302	7078	9804	16114	64185
Total		5863	40055	14287	13787	18922	31561	124475
2017	Male	2955	17790	6923	6869	7869	14381	56787
	Female	3505	19788	7215	7623	8330	15369	61830
Total		6460	37578	14138	14492	16199	29750	118617
Male		9300	55085	20239	19078	25199	46444	175345
Female		10920	59544	21006	20848	26831	49313	188462
Grand Total		20220	114629	41245	39926	52030	95757	363807
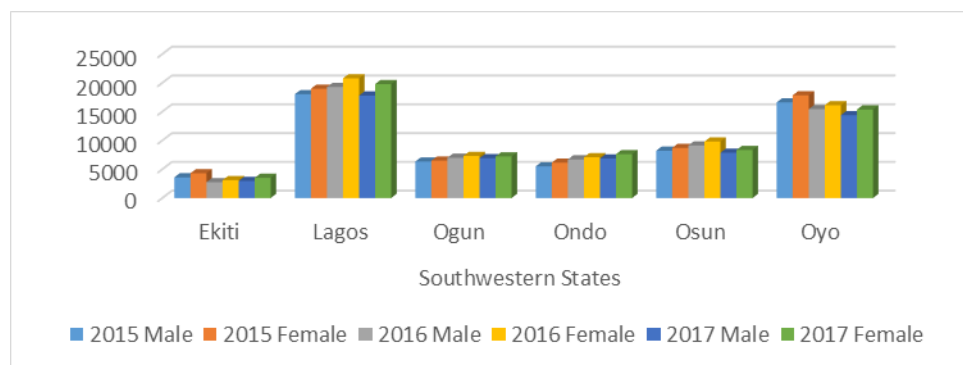
Figure2. Enrolment of Students in NECO Biology Paper III by Sex across Southwestern Nigeria

Table 1 and Figure 2 shows enrolment distribution of students who for the Biology Paper III in Southwestern State in Nigeria from 2015 to 2017. It can be seen that Lagos State had the highest number of male and female students who registered for the subject consistently with a total number of one hundred and fourteen thousand, six hundred and twenty nine {(114,629 (male=55085, female=59,544)}, followed by Oyo State with ninety five thousand seven hundred and fifty seven {(95,757 (male=46,444, female= 49,313)}. Osun State with fifty two thousand and thirty {(52,030 (male=25,199, female=26,831)}; Ogun State with forty one thousand two hundred and forty five {(41,245 (male=20,239, female=21,006)}; Ondo State with thirty nine thousand nine hundred and twenty six {(39,926 (male=19,078, female=20848) and Ekiti State with twenty thousand two hundred and twenty {(20,220 (male=9,300, female=10,920)}. Table 2 shows the selection for the study.

Table 2:

Sample Selection across the Southwestern States from 2015 to 2017

Year	Sex	Southwestern States						Total
		Ekiti	**Lagos**	**Ogun**	**Ondo**	**Osun**	**Oyo**	
2015	Male	342	1739	625	619	844	1693	5862
	Female	448	1961	657	546	847	1752	6211
Total								
		790	3700	1282	1165	1691	3445	12073
2016	Male	268	1930	709	655	903	1571	6036
	Female	318	2076	720	724	989	1585	6412
Total								
		586	4006	1429	1379	1892	3156	12448
2017	Male	307	1759	699	726	804	1448	5743
	Female	339	1999	715	723	816	1527	6119
Total								
		646	3758	1414	1449	1620	2975	11862
Male		917	5428	2033	2000	2551	4712	17641
Female		1105	6036	2092	1993	2652	4864	18742
Grand Total								
		2022	11464	4125	3993	5203	9576	36383

Research Questions

Research Question One: How comparable is the validity of NECO Biology multiple-choice tests across the three years (2015 to 2017)?

To answer this question, the 60 Biology Multiple-choice item of paper III for 2015, 2016 and 2017 were scored dichotomously (1 for correct option and 0 for incorrect option). The data was subjected to Bartlett's test of Sphericity and Keiser-Meyer-Olkin (KMO). For the 2015 Biology Multiple-choice, the KMO of 0.963 (>0.05) and significant Barlett's test of Sphericity ($p < 0.05$) were recorded. For the 2016 Biology Multiple-choice, KMO of 0.967 (>0.05) and significant Barlett's test of Sphericity ($p < 0.05$) were recorded and 2017 Biology Multiple choice, KMO of 0.985 (>0.05) and significant Barlett's test of Sphericity ($p < 0.05$) were recorded were recorded. All the recorded values of KMO for sample adequacy were well above minimum value of 0.06 suggested by Awang, 2010; 2012 and Hoque et al., 2016; 2017. These showed that the data was appropriate for exploratory factor analysis. The results from the exploratory factor analysis are presented in Tables 3, 4, and 5, as well as the scree plots in Figures 1, 2, and 3 for establishing the validity of the NECO Biology Multiple-choice items across the three years under review.

Table 3:

Total Variance Explained for 2015 NECO Biology Multiple-Choice Items

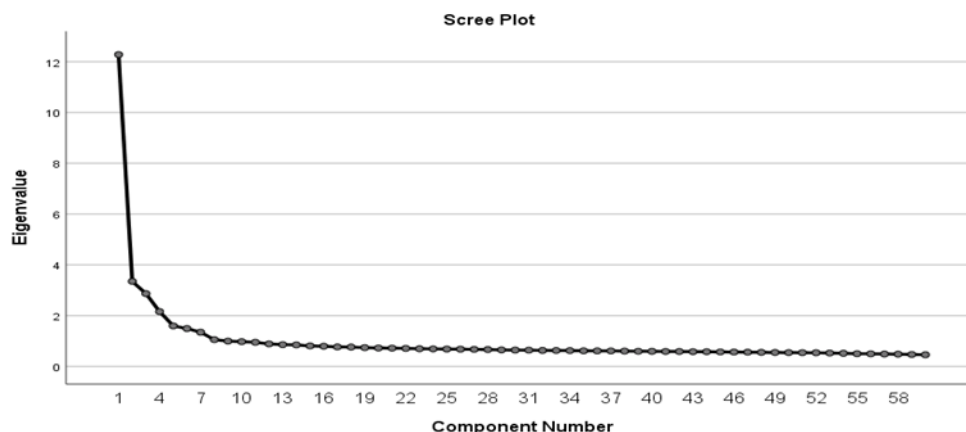| Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	12.280	20.466	20.466
2	3.351	5.585	26.051
3	2.869	4.781	30.832
4	2.160	3.600	34.432
5	1.596	2.660	37.092
6	1.493	2.488	39.581
7	1.347	2.245	41.825
8	1.056	1.759	43.585

Ajeigbe,T.O.

Figure 3. Scree Plot NECO Biology Paper III for 2015

Table 3 showed that 8 factors had eigenvalues greater than one. The 8 factors with eigenvalues greater than one accounted for 43.59% of the total variance. Also, the eigenvalue of the first factor showed 12.820 which accounted for 20.47% variance, while the second factor showed 3.351, accounted for 5.59% variance. This implied that the first factor was clearly greater than the second factor as evidence that a single construct is being measured (that is Biology items), as well as indication that the 2015 NECO Biology Multiple-choice items exhibited good construct validity. Also, the results from the rotated component matrix revealed factor loading ranges from 0.611 to 0.721, which was above the minimum level of 0.6 recommended by Awang (2010; 2012) & Hoque et al. (2016, 2017) for ascertaining the contribution to the construct being measured. This was further supported by the scree plot in Figure 3 revealing a single dominant factor. It can be seen that the 2015 NECO Biology Multiple-choice examinations are comparable in terms of the validity of the items.

Table 4:

Total Variance Explained for 2016 NECO Biology Multiple-Choice Items

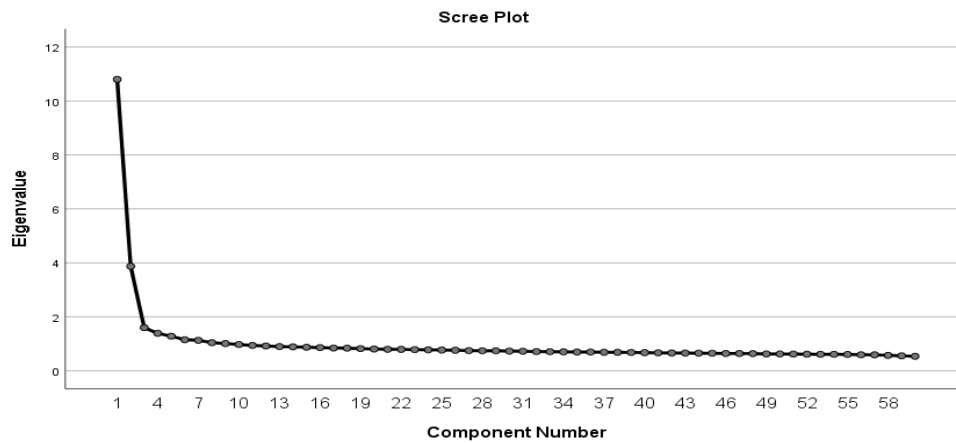| Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	10.799	17.999	17.999
2	3.873	6.454	24.453
3	1.600	2.667	27.121
4	1.390	2.316	29.437
5	1.284	2.140	31.577
6	1.153	1.922	33.499
7	1.132	1.886	35.385
8	1.044	1.739	37.124
9	1.012	1.686	38.811

Figure 4. Scree Plot NECO Biology Paper III for 2016

Table 4 showed that 9 factors had eigenvalues greater than one. The 9 factors with eigenvalues greater than one accounted for 38.81% of the total variance. Also, the eigenvalue of the first factor showed 10.799 which accounted for 18% variance, while the second factor showed 3.873 which accounted for 6.45% variance. This implied that the first factor was clearly greater than the second factor as evidence that a single construct is being measured (that is Biology items), as well as indication that the 2016 NECO Biology Multiple-choice items exhibited good construct validity. This can further be inferred from the scree plot in Figure 4 revealing a single dominant factor. It can be clearly seen that the 2016 NECO Biology Multiple-choice examinations are comparable in terms of the validity of the items.

Table 5:

Total Variance Explained for 2015 NECO Biology Multiple-Choice Items

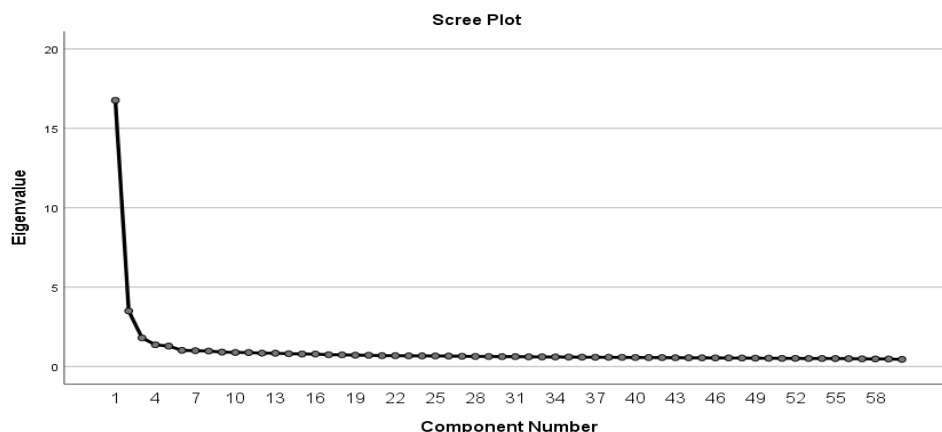| Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	16.762	27.937	27.937
2	3.496	5.826	33.763
3	1.802	3.003	36.766
4	1.367	2.278	39.044
5	1.285	2.141	41.185
6	1.013	1.689	42.874

Figure 5. Scree Plot NECO Biology Paper III for 2017

Table 5 showed that 6 factors had eigenvalues greater than one. The 6 factors with eigenvalues greater than one accounted for 42.87% of the total variance. Also, the eigenvalue of the first factor showed 16.762 which accounted for 27.94% variance, while the second factor showed 3.496 which accounted for 5.83% variance. This implied that the first factor was clearly greater than the second factor as evidence that a single construct is being measured (that is Biology items), as well as indication that the 2017 NECO Biology Multiple-choice items exhibited good construct validity. This can further be inferred from the scree plot in Figure 5 revealing a single dominant factor. From the forgoing, it can be clearly seen that the 2017 NECO Biology Multiple-choice examinations are comparable in terms of the validity of the items.

Research Question Two: How comparable are the reliability coefficients of NECO Biology Multiple-choice examinations across the three years (2015 to 2017)?

Table 6:

Reliability Coefficients of NECO Biology Multiple-choice Examination across three Years

Year	Reliability Methods			
	Split-Half Reliability	%	Cronbach's Alpha	%
2015	0.94	94	0.92	92
2016	0.91	91	0.91	91
2017	0.95	95	0.93	93

The results from Table 6 showed the reliability coefficient of NECO Biology Multiple-choice examination for three consecutive years under Spearman Brown Split-Half and Cronbach's Alpha. For 2015, 2016 and 2017 NECO Biology Multiple-choice examination, the reliability coefficients using Split-Half recorded 0.94 (94%), 0.91 (91%), 0.95 (95%) respectively. In the same manner, Cronbach's Alpha yielded reliability coefficient of 0.92 (92%), 0.91 (91%) and 0.93 (93%) for the three years respectively. These results implied that the reliability coefficients of the NECO Multiple-choice examination are comparable and consistently very high. The disparities are insignificant. The

coefficients recorded are very highly reliable as recommended by Cohen, Manion and Morrison, (2011); Polit and Becks, 2017.

Hypothesis Testing

Hypothesis One: There is no significant different in the comparability of item difficulty, discrimination and guessing parameters of NECO Biology Multiple choice examinations across the three years (2015 to 2017).

To test this hypothesis, a preliminary calibration was carried out to establish item difficulty, discrimination and guessing indices using computer software program "ltm"(Latent Trait Models) under the IRT framework of Birnbaum's three parameter model (Rizopoulos, 2006) under the Item Response Theory approach of Rstudio package:

```
tpm(boas, type = c("latent.trait", "rasch"), constraint = NULL,
    max.guessing = 1, IRT.param = TRUE, start.val = NULL,
    na.action = NULL, control = list())
```

Thereafter, the results were subjected to descriptive (mean) and inferential (One-way Analysis of Variance) statistics under statistical packages for social science (SPSS) version 26. The results are presented in Tables 7 and 8 respectively.

Table 7:

Mean Analysis of Item Difficulty, Discrimination and Guessing Parameters across the three Years.

Years	Parameters	Items	Mean	SD	Min	Max
2015	b	60	-0.635	2.274	-11.050	4.871
2016	b	60	1.046	10.141	-8.435	73.937
2017	b	60	0.059	4.812	-12.962	29.032
2015	a	60	1.388	0.796	-.2380	3.781
2016	a	60	1.832	3.948	-.2790	27.298
2017	a	60	1.409	1.864	-7.6140	5.072
2015	c	60	0.117	0.129	0.000	0.382
2016	c	60	0.087	0.092	0.000	0.278
2017	c	60	0.121	0.196	0.000	0.809

Table 8:

Analysis of Variance Showing Comparability of NECO Biology Multiple-choice Examination for Three Years (2015 to 2017)

Parameters	Sources	Sum of Squares	df	Mean Square	F	Sig.
Difficulty	Between Groups	85.619	2	42.810	0.979	0.378
	Within Groups	7739.134	177	43.724		
	Total	7824.753	179			
Discrimination	Between Groups	7.530	2	3.765	0.573	0.565
	Within Groups	1162.220	177	6.566		
	Total	1169.750	179			
Guessing	Between Groups	.042	2	.021	0.984	0.376
	Within Groups	3.738	177	.021		
	Total	3.780	179			

Tables 7 and 8 showed the descriptive analysis and results of one-way analysis of variance obtained. For item difficulty, there was no statistical significant difference ($F_{(2,177)}= 0.979$; p>0.05) across the three years; for discrimination, there was no statistical significant difference ($F_{(2,177)}= 0.573$; p>0.05) and finally for guessing parameter, there was no statistical significant difference ($F_{(2,177)}= 0.984$; p>0.05) across the three years. This so because the p-value is greater than the 0.05 level of significant, hence, the non hypothesis of no significant difference is hereby accepted. Therefore arising the these results, it can be concluded that the item difficulty, discrimination and guessing parameter of the NECO Biology Multiple-choice items are reliably comparable over the three years considered in the study.

Discussion

The findings from the first objective of the study revealed that in terms of sample adequacy of the Biology multiple-choice tests across the three years recorded values of KMO higher than the minimum value of 0.06 suggested by Awang, 2010; 2012 and Hoque et al., 2016; 2017. Based on the finding as well, it was found out there was a distinct factor underlying the construct of measurement in the 60 Biology multiple-choice test items in years reviewed in this study. The factors retained in each year whose eigenvalues are greater than one contributed well to the variability of the date, this may be associated with the high KMO recorded. Also, the first factor recorded variance greater than 20% across the years, which attest to the fact that the NECO Biology multiple-choice test is unidimensional in nature as proposed by Lumseden, (2007). The variance explained in this study was 31.16% which exceeded the requirement of this criterion, demonstrating a unidimensional trait of the dataThis is tangential to This is an indication that the items of the Biology multiple-choice tests item measured what they were purported to measure.

The findings revealed very high reliability coefficients in NECO Biology multiple-choice tests across the three years that were the focus of this study. The reliability coefficients from the Spearman Brown split-half method and Cronbach's Alpha consistently very high (0.91 – 0.95). The reliability of the NECO Biology multiple-choice test items across the three years conformed with the standard level of very highly reliable range of >0.90 proposed by Cohen, Manion and Morrison, (2011).This is also consistent with the findings reported from the study carried out Cohen etal (2011), the basis for adjudging good items. This result is also in accordance with the guidelines stipulated by Kerlinger and Howard (2000). This may be because of the fact that the examination body (NECO) knowing fully well that the standard of high stake examination cannot be compromised and the need to use items well standardized and unbiased items in assessing students from different parts of the Country of germane.

The last objective of the study revealed that the parameter estimates are stable, which is pointer to the fact that the quality of the test items are secured. This is line with the submission Margaret (2014) that quality is a systematic process of ensuring if the products or services being measured meet with standard specification. ' scores are consistent over years and similar performance was recorded across the states in the Southwestern Nigeria.

Conclusion

The aim of the study was to compare the validity and reliability of the National Examination Council (NECO) Biology multiple-choice tests items across the six states in the Southwestern

Ajeigbe,T.O.

Nigeria for three consecutive years. To this end, the study concluded that the NECO Biology multiple-choice test items had construct validity, unbiased and reliably measured learning outcomes and the parameter estimates (a, b and c) were consistently stable among the examinees.

References

[1]   Awang, Z. (2010). "Research Methodology for Business and Social Sciences", Kelantan: UniversitiTeknologi MARA.

[2]   Awang, Z. (2012). "Structural equation modeling using AMOS graphic", Penerbit UniversitiTeknologi MARA.

[3]   Baker, F.B. & Kim, S. H. (2004). "Item response theory: Parameter estimation techniques" (2nd ed.), New York, Marcel Dekker.

[4]   Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. Research Bulletin 81-20. Princeton, NJ: Educational Testing Service.

[5]   Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores (pp. 397-479). London, UK: Addison-Wesley

[6]   Cohen, L, Manion, L. & Morrison, K (2011). Research methods in Education. New York: Library of Congress Cataloguing-in-publication Data.

[7]   Downing, S. M. (2003), Validity: on the meaningful interpretation of assessment data.*Medical Education, 3*7: 830–837.

[8]   Dziuban, Charles D. and Shirkey, Edwin C. (1974) When is a correlation matrix appropriate for factor analysis? Some decision rules. Psychological Bulletin, 81 (6) 358 - 361.

[9]   Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory:Principles and applications*. Boston: Kluwer Nijhoff.

[10]   Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991).Fundamentals of itemresponse theory.Sage, Newbury Park, CA.

[11]   Hambleton, R. K., & 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), 3847.

[12]   Harwell, M.R., Baker, F., &Zwarts, M. (1998). Item parameter estimation via marginal maximum likelihood and EM algorithm: A didactic. *Journal of Educational Statistics, 13,* 243-271.

[13]   Hoque, A. S. M. M., & Awang, Z. (2016). The Exploratory Factor Analysis (EFA) of Entrepreneurial Marketing Scale - Development and Validation. Tourism Conference 20-22 APRIL 2016 (p. 22).

[14]   Hoque, A. S. M. M.; Awang, Z.; & Siddiqui, B. A. (2017b). Technopreneurial Intention among University Students of Business Courses in Malaysia: A Structural Equation Modeling. International Journal of Entrepreneurship and Small & Medium Enterprise (IJESME), 4, 1-16.

[15]   Kerlinger, F. N. & Howard, B. L. (2000). Foundations of behavioural research (4th Ed.). USA: Earl Mcpeek.

[16]   Lumseden, S. E. (2007). A priori considerations in choosing an item response model. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 57-70). Vancouver, Canada: Educational Research Institute of British Columbia.

[17]   Lord, F. M. (1980). Applications of item response theory to practical testing problems.Erlbaum, Hillside, NJ.

[18]   Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

[19]   National Examination Council (NECO) (2007). National Examination Council Retrieved 1/3/2007 at info@neconigeria.com; service@neconigeria.org

[20]   Ojerinde, D, Popoola, K., Ojo, F., Onyeneho, P. (2012): Introduction to Item Response Theory: Parameter Models, Estimation and Application, Lagos, Nigeria, Goshen,Print Media Limited.

[21]   Pett, M.J., Lackey, N.R., Sullivan, J.J., Making sense of factor analysis: The use of factor analysis for instrument development in health care research, SAGE, Thousand Oaks, CA, 2003.

[22]   Polit, D. S., Beck, C. T. (2017) *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. Wolters Kluwer/Lippincott Williams & Wilkins Philadephia, 784 p.

[23]   Rizopoulos, D. (2006) **ltm**: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, **17(5)**, 1–25. URL http://www.jstatsoft.org/v17/i05/

[24]   Tabachnick, B.G. and Fidell, L.S., Using multivariate statistics (6th ed.), Pearson, 2013.

[25]   Van Barneveld, C. (2007). The Effect of Examinee Motivation on Test Construction withinan IRT Framework.Applied *Psychological Measurement, 31*:31-46.

[26]   West African Examinations Council (WAEC) (2007). History of WAEC. Retrieved 1/3/2007 at http// www: waecnigeria.org. history.htm.

Profile on Taiwo Oluwafemi AJEIGBE

Area Specialisation: Psychometrics/Educational Testing

Number of Publication: 29

Academic and Professional Qualifications

I am a holder B.Sc.Ed (Economics) from Lagos State University, Ojo, (1997-2000), M.A.Ed in Tests and Measurement (2005-2008) and Ph.D. in Tests and Measurement (2008-2012) from Obafemi Awolowo University, Ile-Ife. He joined the services of Obafemi Awolowo University in the Department of Educational Foundations and Counselling on 3rd October, 2006 I was employed as Graduate Assistant in the services of Obafemi Awolowo University. After obtaining a Master of Arts in Education (Tests and Measurement) degree in 2008, he was regraded to the position of Assistant Lecturer on 27th August, 2008. I was later promoted to Lecturer II on 1st October, 2011. I became Lecturer I after the completion of my Ph.D.Ed. (Tests and Measurement) on 28th November, 2012 and promoted Senior Lecturer on 1st October, 2016 and my appointment was confirmed to retirement age on 2nd May, 2013. I am a member of Association of Educational Researchers and Evaluators of Nigeria (ASSEREN)

Current Research

My researches in the area of Tests and Measurement include Comparison of psychometric characteristics of multiple-choice test items, Item analysis and Differential Item Functioning, Item Parameter Models, Dichotomous and Polytomous scoring in Item Response Theory and Item Bias, Comparative analysis of scores, I have presented papers in conferences both locally and internationally. I have supervised over 30 undergraduate projects, 3 Master in Education (Professional), 8 Masters of Arts in Education (M.A.Ed.), 1 Master of Philosophy in Education (M.Phil.Ed.) and currently supervising 1 Ph.D student.

Publications

I have 29 publications to my credit with good spread in local, national, and international journals of repute. He is first author in 14 of the papers and sole author in 6 papers. Also 12(41%) of these publications are in international journals and 64% of his publications focused in Tests and Measurement. The analysis of the paper shows that Dr. Ajeigbe has demonstrated competence in applying research tools to investigate problems in both Testing and Measurement. I have made valuable contributions through my scholarly publications in the areas of Tests and Measurement.

Ajeigbe,T.O.

Services and Administrative Skills (S/AD)

I have always available to render official services whenever I am called upon. I have been adding values to the Department and my University at large. I have saddled with many responsibilities in the Department, Faculty and University, all of which I have been discharging well. At the Departmental level, I served as Part-Adviser to students of Education/Economics between years 2007 to 2012, member of Departmental Postgraduate Committee from year 2012 till date and Secretary, Departmental Postgraduate Committee from 2007 to 2013 and Departmental Examination Officer from 2017 to 2019, Coordinator, Departmental Postgraduate Committee from April, 2019 till date. At the Faculty level, I served effectively as a Member of the University Examination Coordinating and Vetting Committee. I am also being a member, Faculty Board of Education and University Examiner (Postgraduate), Department of Educational Foundations and Counselling. Outside the University, I have been rendering useful services. I served as the Registration Area Centre Officer (RACO) during the 2011 general elections under Independent National Electoral Commission. I also served as Collation Officer (CO) during Kwara State Senatorial re-run election and Presidential election in 2018 and 2019 respectively. I served as Examination Monitoring Officer the SSCE conducted by National Examinations Council (NECO) from 2012 to 2019. In the same manner, I served as resource person in several workshops organised by the Universal Basic Education Board of Osun and Ondo States during capacity building for primary and secondary school teachers.