DETERMINATION OF THE EQUIVALENCE OF WAEC AND NECO SSCE CHEMISTRY ITEMS USING LINEAR EQUATING APPROACHES OF CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

Babatimehin Temitope

Obafemi Awolowo University. NIGERIA

Abstract. The study determined the equivalence of the chemistry items using linear test equating approaches of CTT and IRT. The study adopted descriptive survey research design. The population for the study comprised 36,182 final year 2017/2018 senior secondary school students in Osun State. Multi-stage sampling procedure was used to select the sample for the study. Two instruments, Chemistry Achievement Test Type 1 (NECO), and Chemistry Achievement Test Type 2 (WAEC) were used to collect data. These were the adopted versions of June/July 2015 NECO and May/June WAEC 2015 Senior School Certificate Examination Chemistry (Objective) Paper 1. Data collected were analyzed using equated package of R-language. Results showed that linear test equating approaches of IRT and CTT produced different results. CTT equated scores of examinees' in NCI were higher (range from 5 – 56) than WCI (range from 4 – 49). Similarly, IRT equated scores of examinees' in NCI were

higher (range from 13 – 53) than WCI (range from 11 – 48). The study concluded that there is no statistical equivalence between WAEC and NECO Chemistry examinations in terms of difficulty. The WCI were more difficult than NCI.

Keywords: West African Examinations Council (WAEC), National Examinations Council (NECO), Senior School Certificate Examinations (SSCE), Chemistry

---

Introduction

Examination is one of the procedures used in the determination of how much of appropriate knowledge, skills, ability and attitude students have gained in the process of teaching and learning within a given time. It has therefore become a part of the educational system to assess the effect of teaching learning to distinguish people for certification and promotion.

In Nigeria, at the end of secondary school education, students are expected to sit for certification examinations such as Senior School Certificate Examination (SSCE) conducted by the West African Examinations Council (WAEC) and National Examinations Council (NECO) or National Business and Technical Certificate Education (NBTCE) conducted by National Business and Technical Examination Board (NABTEB). The purpose of these examinations is to measure the extent to which students have achieved the educational objectives of each subject. The certificate awarded by these examination bodies are officially recognized in Nigeria as equivalent. Of major importance is the grading of the certificates awarded by any of these examination bodies as it is a basic determinant of whether a candidate will be qualified for admission or otherwise into institutions of higher learning both in Nigeria and abroad. Candidates can also merge grades obtained from sittings in any two of the examinations.

Since WAEC and NECO have a similar mandate and use similar standardized tests to assess students' knowledge base in various subjects, it is believed that the test items, conditions of administration, procedure for scoring and interpretations are the same. In spite of this mandate, there are criticisms of different forms about the credibility of the examinations conducted by these bodies from major stake holders.

The criticisms include: non-equivalence in the quality of examination items, disparity in performance, mass leakage of examination papers, overcrowding in examination halls and examination malpractices among others. According to Peter (2012), the substandard nature of NECO made some Federal universities from 2002 to 2012 to have rejected NECO results. Ahmed (2014) stated that NECO questions from 2011 to 2014 were of higher standard than those of WAEC. Ojerinde & Faleye (2005) stated that there was no difference between NECO and WAEC, when they were compared. Of all criticisms levied against these examination bodies those that gave the researcher much concern was the non-equivalence of the items in terms of difficulty and disparity in performance of students.

Chemistry by nature deals with the study of the composition of a substance and the interaction between the properties and their composition. It has played a major role in science, technology and society, and it still does so today. There is hardly found anything in nature that chemistry does not have an influence or impact upon. It is no surprise the saying that without chemistry there will be no life. Many countries of the world have forged ahead in terms of development by appreciating the importance of chemistry in their national economy. Chemistry is required at a minimum of credit pass in O' level according to university admission requirements. For admission into the university to study courses in agricultural science, engineering, physical sciences, biological sciences, pharmacy, medicine and other related courses

Considering the importance and educational value of chemistry to the need of individual learner, economic and technological breakthrough of a nation and the effort of researchers to improve on its teaching and learning in Nigeria, it is essential that it is properly taught in schools using equivalent standards. Consequently, the procedure of item development, administration, scoring and interpretation of results of these examination bodies must be at equivalence. This way difference in performance will be an outcome of individual academic effort. At present, there are two popular statistical frameworks in educational measurement through which tests can be developed, validated, and finally used for assessing examinees performance. They are Classical Test Theory (CTT) and Item Response Theory (IRT).

Classical Test Theory (CTT) focuses on test level, its weak theoretical assumptions make it relatively easy to apply in many testing situations (Hambleton & Jones, 1993). It has a few weaknesses. Among these are: the person statistic is sample dependent, and the item statistics (item difficulty and item discrimination) are (examinee) sample dependent. These pose some theoretical difficulty in some measurement situations.

Item Response Theory on the other hand, focuses on item level (Xitao, 1988); it attempts to model the ability of a test taker and the probability of answering an item correctly based on the pattern of responses to all the items that constitute the test. Its ability parameter estimates are not test dependent and item parameter estimates are not group dependent. It overcomes the weaknesses of CTT with its ability to provide invariant item parameters. In order to adequately estimate the ability of the examinees from his/her response to a particular test item, the item parameters of the test should be taken into consideration. In Uysal & Kilmem (2016) opinion test developers attempt to prepare tests as similar as possible in terms of knowledge, skills and content, it is impossible to prevent changes in test difficulty and the differentiation of content and statistical qualities from one test format to another. Variations in difficulty and other statistical

characteristics of different tests scores can be measured and controlled using test equating.

Wendy (2002) described test equating as a statistical procedure for measuring and controlling for variations in the difficulty (and other statistical characteristics) of different tests so that scores from equated tests have comparable meaning. "Unless test scores are adjusted to take account of these differences, comparisons are not fair to all examinees tested" (Michaelides & Haertel, 2004). Test equating as defined by Agah (2013) is aimed at putting the scores obtained by students from different forms of a test on a common scale. Approaches like enhancement strategies such as moderation, self-assessment and test scores equating have been suggested as educational standards control mechanisms in Nigeria (Afemikhe, 2007).

There are situations that require testing of the same qualities in different individuals with different tests (security, justice and so on.) (Crocker & Algina, 1986). In Uysal & Kilmen (2016) opinion test developers attempt to prepare tests as similar as possible in terms of knowledge, skills and content, it is impossible to prevent changes in test difficulty and the differentiation of content and statistical qualities from one test format to another. As explained by Agah (2013) in conducting test score equating, numerous methods are available to the researcher. Some of these methods include; Mean Equating, Linear equating, Levine equally reliable linear equating, Levine unequally reliable linear equating, Tucker linear equating, Chained linear equating, Equipercentile equating (Frequency estimation equipercentile equating, Chained equipercentile equating), One parameter logistic (Rasch) model equating (Concurrent calibration, Fixed based procedure, Equating constant procedure, Major axis procedure), Two parameter logistic model equating (2pl concurrent calibration, 2pl partial credit model, 2pl generalized partial credit model), and Three parameter logistic model (separate and concurrent calibration). Among these methods of equating only the linear, mean and equipercentile equating methods apply to both CTT

and IRT, the others apply exclusively to IRT. Emphasis is on linear equating in this study because it is based on comparison of CTT and IRT and the linear equating method requires lesser sample size and simpler mathematical explorations than the mean and equipercentile equating methods. When two groups of examinees differ in ability levels, and when item parameters are estimated separately for each form, the units of the item parameters are not on the same scale because the examinees' mean ability levels and standard deviations are not equal (Agah, 2013). According to (Kolen, 1988; Cook & Eignor, 1991), under CTT and IRT, linear equating is achieved by setting the means and standard deviation on the two forms of the test equal. In this method, scores are converted so as to have the same mean and standard deviation as item parameter/scores on Eq. (1). This conversion is achieved by setting the standardized item parameter estimates/scores of Eq. (1) and Eq. (2) equal. Mathematically, this is expressed as

$$\frac{x_1 - \mu_1}{\sigma_1} = \frac{x_2 - \mu_2}{\sigma_2} \tag{1}$$

where, σ_1, σ_2 are the standard deviations of type 1 and type 2 respectively; μ_1, μ_2 are the means of type 1 and type 2 respectively; x_1, x_2 are scores on type 1 and type 2 respectively. Finding type 2 equivalent of type 1 requires making x_1 the subject of equation 2, this gives,

$$x_1 = \frac{\sigma_1}{\sigma_2} x_2 + \left[\mu_1 - \frac{\sigma_1}{\sigma_2} \mu_2\right] \tag{2}$$

The Eq. (2) (Kolen, 1988; Cook & Eignor, 1991), represents the model for placing type 2 on the same scale of type 1.

Standard comparison which is relative to equivalence in quality of assessment instrument as well as sincerity in scoring and reporting results is a major issue that needs to be tackled head on by WAEC and NECO. As it is

practically impossible to develop parallel test forms, it is important to make test scores comparable. The equivalence of test scores from two or more examinations can be achieved through test equating. Variations in difficulty and other statistical characteristics of different test scores can be measured and controlled using test equating. It is used in linking scores to achieve comparability in terms of difficulty and discrimination. Test equating is built mainly on the frameworks of CTT and IRT.

In a country and State that is aspiring high in science and technological development for sustainable economy, such as Nigeria and Osun State in particular, performance of students in core science subjects like Chemistry should be given a well-focused attention. Since there is no detailed information about the equivalence of the item parameters of the examinations conducted by these two examination bodies, it is imperative to establish the equivalence or otherwise of the chemistry items in terms of difficulty using linear test equating methods of CTT and IRT.

Many researches had been conducted on comparison of performance of students in WAEC and NECO (using only the analytical method of CTT) in different subjects and states of the federation with very few of such researches focused on students' performance in Chemistry in Osun State secondary schools, hence, this study

Purpose of study

The purpose of the study was to examine the equivalence of WAEC and NECO SSCE Chemistry items using linear test equating approaches of Classical Test Theory (CTT) and Item Response Theory (IRT) in Osun State, Nigeria. The specific objectives were to: (i) examine the difference in the item parameters of the NECO and WAEC chemistry examination items using CTT; (ii) examine the difference in the item parameters of the NECO and WAEC chemistry examination items using IRT; (iii) determine the comparability of the two

Chemistry examinations items in terms of examinee scores and item parameters using linear test equating method

Research questions

The study provided answers to these questions: (i) what are the item parameters of WAEC and NECO dichotomously scored chemistry items using Classical Test Theory approach; (ii) what are the item parameters of WAEC and NECO dichotomously scored chemistry items using Item Response Theory approach; (iii) What are the equated scores of chemistry items in the two examinations.

Methodology

The study adopted the survey research design. It's relevance to this study is based on the fact that responses were elicited to WAEC and NECO Chemistry Objective items from a large sample (considered representative of the whole group of Senior Secondary (SS) 111 Chemistry students) selected from the total population of 2017/2018 final year Senior Secondary (SS) 111 students in Osun State. The linear equating methods of the two test theories were employed to put examinee's scores on a common scale. The variables of the study are the Classical Test Theory (CTT), Item Response Theory models and Chemistry Achievement Test items.

The population for the study was 36,182 which consisted of 18,106 males and 18, 076 females Senior Secondary III (SS 3) who were enrolled by WAEC and qualified to sit for NECO Senior Secondary School Certificate Examination 2017/2018 academic session in Osun State Nigeria. This was achieved with the assistance of the office of the permanent secretary, Ministry of Education, Osun State through the letter of introduction from the HOD.

Multi-stage sampling procedure was employed. From each of the three senatorial districts, simple random sampling procedure was used to select five

LGAs making a total of 15. Two schools were also selected from each of the LGAs using simple random sampling technique. Thus, 30 schools were selected altogether. Intact classes were selected, meaning all 1,105 final year Senior Secondary (SS III) chemistry students in the 30 selected schools constituted the sample.

Two instruments titled Chemistry Achievement Test (Type 1 and Type 2) were used for data collection for the study. These were the adopted version of June/July 2015 NECO (Type 1) and May/June 2015 WAEC (Type 2) Senior School Certificate Examination Chemistry Objective Paper 1 only. The choice of June/July 2015 NECO and May/June 2015 WAEC items was based on adequate coverage of the topics in syllabus. Type 1 paper consisted of 60 items with each item having five options, lettered A-E while Type 2 consisted of 50 items with each item having four options, lettered A-D from which the test takers indicated the correct option. Correct response attracted a score of 1, while incorrect response attracted 0 based on the Senior Secondary School Chemistry curriculum.

The research instruments Type 1 and Type 2 representing NECO and WAEC examinations respectively, are standardized external examinations that is assumed to have been previously moderated and validated by the respective examination bodies. The researcher did thorough review of the test items on each instrument with their respective syllabus to confirm if the items were constructed within the dictate of the curriculum.

Data for this study were collected in two stages. Stage one, data collection from Chemistry Achievement Test Type 1 and stage two, data collection from Chemistry Achievement Test Type 2 with an interval of two weeks. Tests administrations were done under standard examination conditions. The researcher administered the instruments for the study on the students with help of research assistants and the Chemistry teacher \ teachers of each school through the permission of the principal to do so. The students were briefed about the

essence of the study and equally informed of the confidentiality of the information that would be provided. The research assistants who are graduates of different university instructed the students that the tests were to determine their level of preparation for their final examinations. They also assisted in the supervision of the whole process. Data were analyzed using equate package of R language.

Results

Research question 1: what are the item parameters of WAEC and NECO dichotomously scored chemistry items using Classical Test Theory approach?

Responses of Students to WAEC and NECO chemistry items were scored for each item. Based on students" performance on each of the items, an item analysis was carried out using mean and standard deviation, independent-sample Mann-Whitney U test to establish the difficulty index and discrimination power of each of the items.

Table 1. Descriptive statistics of the 50-item WAEC and 60-item NECO parameters using CTT

	Discrimination			**Difficulty**		
Items	**SD**	**Mean** ($\bar{X}$)	**Range**	**SD**	**Mean** ($\bar{X}$)	**Range**
WAEC(X)	0.10	0.29	0.14-0.50	0.14	0.39	0.11-0.69
NECO(Y)	0.19	0.28	- 0.08-0.53	0.17	0.41	0.06 - 0.75

Table 1 shows the descriptive statistics of the 50-item WAEC and 60-item NECO parameters using CTT. From the table it can be seen that discrimination indices for WAEC and NECO chemistry items ranged between 0.14 - 0.52 and -0.08 - 0.53 respectively.

The negative discrimination index observed in NECO items is an indication that the weak examinees' got difficult items right or vice versa. On the average the WAEC items had higher discrimination index ($\overline{X}$= 0.29, SD = 0.10) than the NECO items ($\overline{X}$ = 0.28, SD = 0.19).

To test whether the difference observed in the discrimination indices of the WAEC and NECO items was significant under CTT, Mann-Whitney U test was conducted. The results are presented in Table 2.

Table 2. Mann-Whitney U test showing the difference in the discrimination indices of WAEC and NECO items

U	Sig	Decision
0.769	0.442	Do not reject the null hypothesis

P > 0.05

Result, as presented in Table 2, shows that there was no significant difference in the distribution of the discrimination indices of WAEC and NECO Chemistry items (U= 0.769, p > 0.05). The implication of the result is that the WAEC and NECO Chemistry items discriminated almost equally among examinees under CTT.

Table 1 results show that WAEC chemistry items ($\overline{X}$ = 0.39, SD = 0.14) was higher in difficulty than NECO chemistry items ($\overline{X}$ = 0.41, SD = 0.17). In order to test whether the difference observed in the difficulty indices of the WAEC and NECO Chemistry items was statistically significant, Mann-Whitney U test was conducted. The results are presented is Table 3.

Table 3 shows that the difference in the difficulty indices of WAEC and NECO Chemistry items under CTT was not significant.

Table 3. Mann-Whitney U test showing the difference in the difficulty indices of WAEC and NECO items

U	Sig	Decision
0.703	0.482	Do not reject the null hypothesis

P > 0.05

Research question 2: what are the item parameters of WAEC and NECO dichotomously scored chemistry items using Item Response Theory approach?

Responses of students to WAEC and NECO chemistry items were scored for each item. Based on students" performance on each of the items, an item analysis was carried out using multi-dimensional IRT (MIRT) package, mean and standard deviation, independent-sample Mann-Whitney U test. The WAEC and NECO Chemistry items were calibrated using multidimensional 3-parameter logistic IRT model (M3PL). The choice of the model was based on the fact that M3PL fitted the data sets more than the M2PL and M1PL respectively. The WAEC Chemistry items had better discrimination index ($\bar{X}$ = 4.27, SD = 6.56) than the NECO Chemistry items ($\bar{X}$ = 7.54, SD = 13.71). See results in Table 4.

Table 4. Descriptive statistics of the 50-item WAEC and 60-item NECO parameters using IRT

	Discrimination			**Difficulty**		
Items	**SD**	**Mean** ($\bar{X}$)	**Range**	**SD**	**Mean** ($\bar{X}$)	**Range**
WAEC(X)	6.56	4.27	0.33-32.16	0.38	1.08	-28.36-15.66
NECO(Y)	7.54	13.71	0.02-65.55	4.94	0.70	-0.81-4.10

To test whether the difference observed in the discrimination parameters of the NECO and WAEC Chemistry items was significant, Mann-Whitney U test was further conducted. The results are presented in Table 5.

Table 5. Mann-Whitney U test showing the difference in the WAEC and NECO chemistry item discrimination indices under IRT

U	Sig	Decision
-0.925	0.355	Do not reject the null hypothesis

P > 0.05

This result indicated that the difference in the discrimination indices of WAEC and NECO items under IRT measurement framework was not significant.

Table 4 shows that difficulty indices ranged from -0.81 - 4.10 and -28.36 - 15.66 for WAEC and NECO Chemistry items respectively. On the average, the WAEC Chemistry items were of higher difficulty ($\overline{X}$= 1.08, SD = 0.83) than the NECO Chemistry items ($\overline{X}$ = 0.70, SD = 4.94). To test whether the difference observed in the difficulty parameters of the NECO and WAEC Chemistry items was significant, Mann-Whitney U test was further conducted. The results are presented in Table 6.

Table 6. Mann-Whitney U test showing the difference in the Difficulty Indices of WAEC and NECO chemistry items under IRT

U	Sig	Decision
1.099	0.272	Do not reject the null hypothesis

P > 0.05

Table 6 shows that the distribution of the difficulty indices of WAEC and NECO tests under IRT measurement framework as presented in Table 6 was not different significantly (U= 1.099, p > 0.05).

Research question 3: What are the equated scores of chemistry items in the two examinations?

Thus, to arrive at the equated scores of examinees in WAEC and NECO chemistry items, the test scores and the item parameters emanating from the tests were linked. Then, WAEC test scores and item parameters were transformed to the scale of NECO test using linear equating. According to Kolen & Brennan (2014), linear equating is represented by

$$m_y(x) = y = \frac{\sigma(Y)}{\sigma(X)}x + [\mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\ \mu(X)] \quad (3)$$

where $\frac{\sigma(Y)}{\sigma(X)}$ = Slope usually represented with A; $\mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\ \mu(X)$ = intercept, usually represented with B.

On substitution, Eq. (3) becomes

$$m_y(x) = y = Ax + B \quad (4)$$

Equated scores of WAEC and NECO chemistry items

The equated scores for the WAEC and NECO chemistry items were obtained under CTT and IRT frameworks respectively. The equated scores were obtained using equate package of R language and Environment for statistical computing. The results are presented as follow:

Table 7. Slope and intercept of WAEC and NECO chemistry items score equating function under CTT framework

Slope	Intercept
0.87401	-1.81417

Table 8. Slope and intercept of WAEC and NECO chemistry items score equating function under IRT framework

Slope	Intercept
0.84268	-0.48526

Tables 7 and 8 show the slope and intercept of the function used in placing the scores of NECO chemistry items on the scale of the WAEC chemistry items under CTT and IRT measurement frameworks respectively. Thus, the equating functions of the tests were obtained by substituting for slope and intercept in Eq. (4). On substitution for slope and intercept in Eq (iv) under CTT and IRT

CTT

$$m_{NECO}(WAEC) = NECO = 0.87401\,WAEC\ SCORE - 1.81417$$

IRT

$$m_{NECO}(WAEC) = NECO = 0.84268 WAEC\ SCORE - 0.48526$$

Thus, the equating function used in placing NECO chemistry items scores on the scale of WAEC chemistry items score for effective comparison of the test scores form the two tests scores obtained under CTT and IRT. The NECO and WAEC chemistry items scores equivalent using the equating function is presented in Table 9.

From Table 9 the equated scores of NECO (2015) chemistry items (range from 5-56) were higher than WAEC (2015) chemistry items (range from 4-49) under CTT while under IRT, equated scores on NECO (2015) chemistry items range from 13-53 and WAEC (2015) chemistry items range from 11- 48.

Table 9. Equated scores of the WAEC and NECO chemistry items using CTT and IRT scoring methods

CTT NECO score	WAEC Equivalent score	IRT NECO score	WAEC Equivalent score
5	4	13	11
6	4	14	11
7	4	15	11
8	4	16	11
9	4	17	12
10	5	18	13
11	6	19	14
12	7	20	15
13	8	21	16
14	9	22	17
15	10	23	18
16	11	24	19
17	12	25	20
18	13	26	21
19	14	27	22
20	15	28	23
21	16	29	24
22	17	30	25
23	18	31	26
24	19	32	27
25	20	33	28
26	21	34	29
27	22	35	30
28	23	36	31
29	24	37	32
30	25	38	33
31	26	39	34
32	27	40	35
33	28	41	36
34	29	42	37
35	30	43	38
36	31	44	39
37	32	45	40

38	33	46	41
39	34	47	42
40	35	48	43

Equated Scores of the WAEC and NECO Chemistry Items using CTT

42	37	50	45
43	38	51	46
44	39	52	47
45	40	53	48
46	41		
47	42		
48	43		
49	44		
50	45		
51	46		
52	47		
53	48		
54	49		
55	49		
56	49		

Discussion

Result showed that the difference in the discrimination and difficulty indices of WAEC and NECO Chemistry test items under CTT and IRT measurement frameworks were not significantly different from one another. Although, going by the means and standard deviations of the difficulty indices of WAEC and NECO Chemistry items which showed that WAEC chemistry items have better difficulty indices than NECO chemistry items, one may be tempted to say this result supported Abiri (2006) findings that difficulty indices of multiple choice test with fewer number of options say four (4) is better than with larger number of options. But the difference between the discrimination and difficulty indices of the two examinations is not significant under the CTT and IRT frame works. The implication of these results is that NECO and WAEC Chemistry tests' items discrimination and difficulty indices are comparable using IRT and

CTT frameworks. These results are in agreement with Valipour & Zoghi (2014) finding in their comparative study of CTT and IRT in estimating test item parameters in a linguistic test. Their result suggested that CTT and IRT parameters are comparable. In the same vein, Awopeju & Afolabi (2016) also concluded in their comparative analysis of CTT and IRT based item parameter estimates of NECO Senior Secondary School Certificate Mathematics examination that the two frameworks are comparable in estimating item characteristics of statistical and psychometric tests. As against Olutola (2015) claim that mean difficulty index of WAEC items is slightly higher than NECO items and with discriminating power higher than NECO.

A close look at the equated scores under the two frameworks showed that students have higher scores in NECO Chemistry items than WAEC Chemistry items. This is to say that the NECO Chemistry items are easier for the students than the WAEC chemistry items. More still, scores produced by IRT framework is higher than those produced using CTT in the Chemistry items. In the same vein, scores produced using linear equating methods of CTT and IRT indicated that WAEC Chemistry items are more difficult than NECO Chemistry items. This finding differs from the findings of Peterson et al. (1983) in their study of IRT versus conventional equating methods where CTT and IRT produced the same result. Although in the study linear, equipercentile and other IRT equating methods are compared the IRT equating methods investigated are the Tucker, Levine equally reliable and Levine unequally reliable models. The result also did not agree with the findings of Metibemu (2016) study on the comparison of CTT and IRT frameworks in the development and linear equating of physics achievement test where the linear equating methods of the two frameworks produced a similar result. The disagreement in findings may be as a result of the type of items used. This study compared two standardized tests while Metibemu (2016) study compared one standardized items and the other developed.

Conclusion and recommendation

The implication of this is that IRT equated scores are higher than CTT equated scores and examinees' scores in NECO chemistry items are higher than in WAEC items under the two frameworks. To ensure that test scores from different forms of test are on a common scale test equating should be employed.

REFERENCES

Abiri, J.O.O. (2006). *Elements of evaluation measurement and statistical techniques in education.* Ilorin: University of Ilorin.

Afemikhe, O.A. (2007). Assessment and educational standard improvement: reflections from Nigeria. *33rd Annual Conference of the International Association for Educational Assessment*, Baku, Azerbaijan. September $16^{th} - 21^{st}$.

Agah, J.J. (2013). *Relative efficiency of test scores equating methods in the comparison of students' continuous assessment measures: PhD thesis.* Nsukka: University of Nigeria.

Ahmed, M.F. (2014). Difficulty index of mathematics multiple-choice items of West African Examinations Council and National Examinations Council Senior Secondary School Certificate Examinations from 2006 – 2010. *Assoc. Tech. Ind. Papetiere, 13*, 25 – 31.

Awopeju, O.A. & Afolabi, E.R.I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *Eur. Scientific J., 12*(28), 263-284.

Cook, L.L. & Eignor, D R. (1991). IRT equating methods. *Educ. Measurement: Issues & Practice, 10*, 37-45.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* San Diego: Harcourt.

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

Kolen, M.J. (1988). Traditional equating methodology. *Educ. Measurement: Issues &Practice, 7*(4), 29-37.

Kolen, M.J. & Brennan, R.L. (2014). *Test equating, scaling and linking methods and practices*. New York: Springer.

Metibemu, M.A. (2016). *Comparison of classical test theory and item response theory frameworks in the development and equating of physics achievement tests in Ondo state, Nigeria*: *PhD thesis*. Ibadan: University of Ibadan.

Michaelides, M.P. & Haertel, E.H. (2004). *Sampling of common items; an unrecognized source of error in test equating*. Los Angeles: Center for the Study of Evaluation.

Ojerinde, O.O.D. & Faleye, B.A. (2005). Do they end at the same point? *J. Soc. Sci., 11,* 239-241.

Olutola, A.T. (2015). Empirical analysis of item difficult and discrimination indices of senior school certificate multiple choice biology tests in Nigeria. *41st Annual Conference of International Association of Educational Assessment*, 11th - 15th October.

Peter, K. (2012). A study of the attitude of some Nigerian science students towards NECO and WAEC. *J. Professional Science &Vocational Teachers Association of Nigeria, 12*(1), 15-18.

Peterson, N.S., Cook, L.L. & Stocking, M.L. (1983). IRT versus conventional equating methods: a comparative study of scale stability. *J. Educ. Stat, 8,* 137-156.

Uysal, I. & Kilmen, S. (2016). Comparison of item response theory test equating methods for mixed format tests. *Int. Online J. Educ. Sci., 6*(2) 1-11.

Valipour, V. & Zoghi, M. (2014). A comparative study of classical test and item response theory in estimating test item parameter in a linguistic test. *Indian J. Fundam. & Appl. Life Sci., 4*(54), 424-435.

Wendy, Y. (2002). Scaling and equating. A paper presented at the New York Technical Conference, October 12th.

Xitao, F. (1988). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educ. & Psych. Measurement, 58*, 357-381.

✉ Dr. Babatimehin Temitope
Department of Educational Foundations and Counselling
Obafemi Awolowo University,
Ile-Ife, Osun State, Nigeria
E-Mail: temitopebabatimehin@gmail.com