AJEMATES
African Journal of Educational Management, Teaching and Entrepreneurship Studies
VOL.12 (1), May-August, 2024 https://ajemates.org

COMPARABILITY OF STANDARD AND FAIRNESS IN SCHOOL BASED ASSESSMENT PRACTICE IN SECONDARY SCHOOLS IN EBONYI STATE, NIGERIA

Ugo Julius Nwali
Department of Science Education
Faculty of Education
Ebonyi State University, Abakaliki
ugojulius@gmail.com

ABSTRACT

The purpose of the study was to determine comparability of standard and Fairness in School based Assessment scores using one parameter logistics model on mathematics achievement test. The design of the Study was causal comparative design. The Area of the Study were the three Educational Zones in Ebonyi State; Ebonyi North Educational Zone, Ebonyi Central Educational Zone and Ebonyi South Educational Zone. The population of the Study comprised all Senior secondary School two (SSS2) Students of 2023/2024 Academic Session in those Three Educational Zone And a sample of 307 Students were drawn using Yaro Yarman's Sampling Formula. The instrument used for Data collection were senior secondary school based assessment test form for the three Educational Zone in Ebonyi State. Three Research Questions and Three hypotheses guided the study. The research Questions were answered using Item parameter indices (Difficulty index) and Descriptive statistics (Mean, Standard deviation), IRT differential Item functioning for one parameter. Three hypotheses were tested at 0.05 level of significance using ANOVA, independent T-test. The Data collected were analyzed using BILOG-MG and SPSS. Major Findings of the Study showed that: There was a significant difference in the Difficulty index of the three forms of mathematics Achievement test (MAT) administered to the students. There was a significant difference in the mean ability of the students in the three Educational Zone due to variation in the test item difficulty. There was no significant difference in the Item functioning between Genders in the Educational Zones in Ebonyi State. Based on these Findings, it was recommended among other things, that Development of State or Federal Item pool or Bank where each school in various states and Educational Zones in Nigeria can withdraw reliable and valid assessment instrument. This will enhance uniformity and Fairness in School based Assessment.

Keywords: Comparability, Standard, Fairness and Assessment.

Introduction

School based assessment is an assessment embedded in teaching and learning process. It involves the teacher from the beginning to the end; from planning the assessment programme, to identifying and developing appropriate

assessment tasks right through to making the assessment judgements.

Schools based assessment in Nigeria started in 1977 when Nigeria government first introduced continuous assessment as contained in the national policy on education. The government taken notice of the gave consequences of a malpractice ridden exam and certification made haste to form various commission of enquiries into the problem in which one of the outcomes of commissions work was the introduction of 6-3-3-4 system of evaluation called continuous assessment or school based assessment. The school based assessment is a not a finished and neatly packaged product. There are indeed a few unresolved issues in the practice (Ojerinde, 2002). One of such issues is the problem of comparability of standards (Nworgu, 2003; Okoye, 2007). This problem arises because of the difference in the quality of tests and other assessment instruments in different schools (Nworgu, 2003). Ugo (2012) Lamented that it is difficult to compare a score of 90% given to a learner in school A with another 90% given to another learner since both tests may not be equal in difficulty in items of the test form.

Onjewu (2007) in her presentation on assuring fairness in the continuous assessment component of school based assessment practice in Nigeria observed that there are internet problems with the derivation of Continuous Assessment Scores (CAS) and the Situation deserves the focus of academics. Onjewu (2007) further argued that the entire practice of Continuous Assessment is surrounded by Laxity, and this includes laxity in timing, laxity in terms of the mode that the CA exercise takes and laxity in relation to the content of CA. The School Based Assessment Scores (SBAS) perse, has generated a lot of controversies between schools and examination bodies like West African Examination Council (WAEC) and National Examination Council (NECO). Schools are accusing examination bodies of not using the Continuous Assessment Scores forwarded to them, while the examination bodies are in turn accusing the school of inflating the CAS sent by them. This accusation exist on the ground that the learners who scored high in continuous assessment test (CAT) failed at the end of the Terminal Assessment Test (TAT).

At the secondary level of education in Nigeria, CAS is incorporated into the terminal assessment score and there is growing concern from students, Teachers, parents and some stakeholders on how effectively this is being done (Afemikhe, 2007), Since there is no uniform items in CAT that scores are generated for CAS. The terminal assessments particularly those conducted by West African Examination Council (WAEC) and National Examination Council (NECO) have uniform items for each student across schools in Nigeria, this enhances comparability of raw scores. But in the continuous assessment, the comparison of raw scores from various schools in each state is difficulty and no reasonable measures have been taken towards this direction.

Currently, examination bodies like NECO incorporate CAS into TAS through the use of T-score transformation. The T-score transformation approach used by

NECO and WAEC for the standardizing of CAS does not consider how scores are obtained, especially in areas of items in the test forms. The issues of comparability of standards which deals with uniformity and quality of assessment instruments, as well as honesty and integrity in reporting of assessment result among others has been a problem right from the introduction of school based Assessment. Based on the problem under study enhancement strategies as educational control mechanism in Nigeria (Afemikhe, 2007).

Assessment moderation and self-assessment has been practice especially at the tertiary and not at the secondary and primary level of education in Nigeria, whereas test score equating is not practiced at all. It become necessity to carry out study on test score equating in Nigeria using Continuous Assessment Scores (i.e school based assessment scores) generated from continuous assessment test, considering the test items.

The assessment result, in which test scores are paramount has a consequential effect on students' future. Therefore, the need for the standardization and placing of test scores on common scale to enhance fairness and comparability. Varied approaches can be used ion ensuring that test scores are standardized and placed on a common scale, and one of such approach which has not been properly explored in Nigeria is test score equating.

Wendy (2002) describes test equating a statistical procedure for measuring and controlling variations in the difficulty of different tests so that scores from equated tests have comparable meaning.

The process of equating enables the tests users to interchange multiple from of a test (Angoff and Cowel, 1985). Test equating is an empirical procedure used in establishing the relationship between the raw-scores of two or more test forms.

This study is therefore designed to examine the comparability of standard in continuous assessment score using one-parameter Logistic model on Mathematics achievement test in Ebonyi State, since schools in Ebonyi State incorporate their CAS in TAS, with varied test items or continuous assessment test (CAT) items.

Statement of the Problem

It is a known fact that schools in Nigeria submit the 30% scores generated from continuous assessment test (i.e school based assessment test) to external examination bodies for incorporation into terminal assessment scores. The non-uniformity of test items on CAT across the state, at which examination bodies like WAEC/NECO that make use of the 30% generated from CAT have tried to address the issue but failed through application of T-score transformation method in standardizing the CAS from various schools in all states across the country. This T-score method is not quite appropriate because it does not guarantee accurate comparison of students' achievement, since item-per-person is not considered.

School based assessment score (i.e Continuous Assessment Score) are generated by teachers without considering the item quality and person-parameter. In as

much as it seems desirable to use one common standard in judging the quality of the human capital being formed by different schools within and across states in Nigeria, it is however impossible to do so given diversity of the teachers, school environments, students, tests, among others in the case of continuous assessment. It is on this basis that the researcher decided to request for score equating using one parameter logistic model, to tackle the problem of comparability of standard and fairness in school based assessment practice in secondary schools in Ebonyi State, Nigeria.

Purpose of the Study

The main purpose of the study is to determine comparability of standard and fairness in continuous assessment scores using one parameter logistic model on mathematics achievement test. Specifically, the study sought to:

(1) Determine the item parameter estimates of the test forms of internal mathematics mock achievement test in three Education Zone of Ebonyi State.

(2) Estimate the item parameter consistency values (differential item functioning – DIF) for male and female using one parameter logistic model.

(3) Determine the mean ability of students in internal mathematics mock achievement test among the students in three Education Zone of Ebonyi State.

Research Questions

The following research questions were raised for the study.

1. What are the item parameter estimates of internal mathematics mock achievement test in three Education Zone of Ebonyi State?
2. What are the item parameter consistency values (different item functioning – DIF) for male and female students in the three Education Zone of Ebonyi State, using one parameter logistic model?
3. What are the mean ability of students in internal mathematics mock achievement test among the students in three Education Zone of Ebonyi State?

Research Hypotheses

To guide the study, the following null hypotheses were formulated and tested at 0.05 level of significance.

H_{01}: There is no significant difference in the item parameter estimates of the test forms in the three Education Zone of Ebonyi State, using one parameter logistic model.

H_{02}: The test items parameter do not significantly function differentially between male and female students in the internal mathematics mock achievement test in the three Education Zone of Ebonyi State.

H_{03}: There are no significant differences in the mean ability estimates of students in the three Education Zone of Ebonyi State.

Research Method

Design of the Study

The design used in this study is casual-comparative research design. according to Akpan and Joseph (2005) causal-comparative design aimed at the discovery of possible causes for a behaviour pattern by comparing subjects in whom this pattern is presented with similar known subjects in whom it is not present or present to a lesser degree. The design was appropriate for this study because the researcher was interested in identifying the level of difficulty index in the test forms in CAT meant for generation of CAS in various secondary schools in Ebonyi State.

Population of the Study

The population of the study consisted of all senior secondary school two (SSS II) students in government owned schools in Ebonyi State. There are 172 government owned secondary schools in Ebonyi State, comprised of 53 secondary schools in Ebonyi South Education Zone, 56 secondary schools in Ebonyi North Education Zone and 63 secondary schools in Ebonyi Central Education Zone (Ebonyi State Secondary Education Board, 2023). The number of SS II students in ebonyi state was 46324, which is made up from 14,00 student's from Ebonyi South Education Zone, 15,203 students from Ebonyi North Education Zone and 17,121 students from Ebonyi Central Education One, that formed the population of the study. The study concentrated on senior secondary school two students, because they are always subjected to internal mock exam for CAS generation, since the students have taught almost all aspects of WAEC, NECO syllabus in mathematics and they are equally preparing for external exam.

Sample and Sampling Technique

The sample was derived from 46,324 students that made up the population of the study. The sample size for this study was determined by using Taro Yarman's sampling formula and put as follows:

$$n = \frac{N}{1+N(e)^2}$$

Where n = Sample size

e = Error limit or level of significant = 0.05

N = Population

I = Constant

$$n = \frac{46324}{1+46324(0.05)^2} = \frac{46324}{116.81} = 306.6$$

n = 307 students

Therefore, the sample size for this study is 307 (2023/2024) SS 2 students in Ebonyi State. Systematic sampling technique was used in this study which involved randomly selecting schools and students for the test administration at a given internal of 151 to ensure equal representation, derived from the quotient of the total population 46324 and the sample size of 307.

Therefore, out of the 307 sample of the students in the three Education Zones in Ebonyi State. 102 students sample were drawn systematically from South Zone (C), 103 from Central Zone (B) and 102 from North Zone (A).

Instrument for Data Collection

The instrument for data collection were senior secondary school two school based mathematics achievement test form for generation of 30% score to be forwarded to examination bodies like WAEC/NECO for incorporation into TAS. The researcher randomly picked one mathematics mock achievement test form from one of the schools in each Education Zone and administer it to the sampled students. It is made up of 50 multiple choice test items. The instruments for data collection solely based on multiple choice tests commonly known as internal mathematics mock achievement test developed in different schools in Ebonyi Education Zones.

Method of Data Collection

The researcher visited the three Education Zones and administered the test forms to sampled students in each zone and collect the answer script, score it with the help of some mathematics teachers in each zone. The 307 scripts administered were validity retrieved.

Method of Data Analysis

The data collected were analyzed using BILOG-MG and SPSS software. The research questions were answered using item parameter indices, descriptive statistics (mean, standard deviation) IRT Differential Item Functioning for one parameter logistics model. The hypotheses were tested at 0.05 level of significant using ANOVA, independent t-test.

Decision rule

In this study, p-values ranged from 0 to 1, from zero to 0.4 indicate too difficult item and 0.6 to 1 is less difficult, then 0.5 is moderately difficult. According to Anastasi and Urbina (1997) "the closer the difficulty of an item approaches 1 or 0 the less differential information about the test takers. #it contributes. Conversely, the closer the difficulty level approaches 0.5 the more differential the item can make. In differential item functioning (DIF), when DIF value was less than 1.0 implies no DIF, if it was 1.0 to 1.5, it implies weak DIF and when DIF-value was greater than 1.0, it implies strong DIF.

In this study, the referenced group is advantaged group (male) and focal group is the disadvantaged (unfavoured) group.

In ANOVA and t-test testing of hypotheses, if p-value is greater than 0.05 accept null hypotheses if not rejected.

Presentation of Results

Analysis of the research questions

Research Question 1: What are the item parameter estimates of the test forms of mathematics mock achievement test in the three Education Zone in Ebonyi State?

Table 1: Item Difficulty Index of the test Forms of Mathematics Mock Achievement Test in the three Education Zone in Ebonyi State.

Zone (A) Ebonyi North				Zone (B) Ebonyi Central				Zone (C) Ebonyi North			
Item	P	Item	P	Item	P	Item	P	Item	P	Item	P
1	0.6	26	0.6	1	0.2	26	0.3	1	0.8	26	0.8
2	0.3	27	0.5	2	0.5	27	0.5	2	0.5	27	0.2
3	0.8	28	0.7	3	0.3	28	0.1	3	0.4	28	0.4
4	0.5	29	0.8	4	0.5	29	0.6	4	0.6	29	0.7
5	0.6	30	0.3	5	0.8	30	0.2	5	0.2	30	0.3
6	0.7	31	0.7	6	0.4	31	0.3	6	0.3	31	0.1
7	0.8	32	0.9	7	0.1	32	0.5	7	0.1	32	0.8
8	0.6	33	0.6	8	0.6	33	0.7	8	0.7	33	0.2
9	0.4	34	0.4	9	0.2	34	0.8	9	0.2	34	0.4
10	0.6	35	0.7	10	0.4	35	0.3	10	0.4	35	0.6
11	0.2	36	0.8	11	0.7	36	0.4	11	0.5	36	0.1
12	0.7	37	0.6	12	0.3	37	0.6	12	0.6	37	0.4
13	0.9	38	0.9	13	0.5	38	0.3	13	0.8	38	0.2
14	0.7	39	0.5	14	0.4	39	0.1	14	0.1	39	0.6
15	0.5	40	0.8	15	0.2	40	0.4	15	0.4	40	0.3
16	0.6	41	0.7	16	0.8	41	0.5	16	0.6	41	0.7
17	0.3	42	0.4	17	0.3	42	0.8	17	0.2	42	0.5
18	0.8	43	0.6	18	0.6	43	0.2	18	0.5	43	0.4
19	0.6	44	0.8	19	0.1	44	0.4	19	0.3	44	0.2
20	0.4	45	0.6	20	0.5	45	0.2	20	0.6	45	0.7
21	0.7	46	0.5	21	0.8	46	0.5	21	0.5	46	0.1
22	0.1	47	0.8	22	0.3	47	0.2	22	0.1	47	0.4
23	0.6	48	0.6	23	0.4	48	0.4	23	0.3	48	0.5
24	0.2	49	0.7	24	0.2	49	0.6	24	0.7	49	0.3
25	0.8	50	0.5	25	0.7	50	0.1	25	0.3	50	0.2

P = Difficulty Index; **Source:** Field Work (2024)

Table 1 shows that the difficulty values (p.parameter) for Zone A (Ebonyi North) are, 34-items very easy with value range from 0.6 to 0.9, 6-item moderate with value 0.5 and 1.0. For Zone B (Ebonyi Central) are 11-items very easy or less difficult, 8-item moderate and 31-items too or very difficult. For Zone C (Ebonyi South) are 15-item less difficult, 6-items moderate and 29-items too difficult. It can be seen from the table that zone A uses easy items to generate their CAS whereas zone B and C uses too difficult items to generate their CAS.

Research Question 2: What are the item parameter consistency values (differential item functioning – DIF) for male and female students in the three Education Zone in Ebonyi State, using one parameter logistic model?

Table 2: Summary of Item Parameter Consistency using Threshold Difference Values for Sex.

Item	Group (M-F)		Item	Group (M-F)	
	DIF value	S.E		DIF value	S.E
1	0.43	0.13	26	-0.15	0.06
2	-0.25	0.10	27	0.17	0.12
3	0.68	0.16	28	-0.25	0.10
4	1.20	0.83	29	0.66	0.16
5	0.22	0.08	30	0.85	0.18
6	0.88	0.17	31	1.21	0.95
7	0.91	0.18	32	0.92	0.24
8	0.29	0.01	33	0.55	0.14
9	0.34	0.12	34	0.42	0.13
10	0.75	0.18	35	1.32	0.01
11	1.20	0.83	36	1.18	0.62
12	1.32	1.01	37	0.88	0.12
13	1.21	0.95	38	0.26	0.08
14	0.15	0.07	39	0.75	0.19
15	0.72	0.16	40	0.63	0.15
16	-0.54	0.04	41	0.99	0.26
17	0.77	0.19	42	-0.45	0.02
18	0.63	0.15	43	0.68	0.18
19	0.48	0.14	44	0.48	0.14
20	0.92	0.24	45	0.79	0.21
21	0.37	0.11	46	0.34	0.15
22	0.58	0.15	47	0.28	0.07
23	0.79	0.20	48	0.88	0.17
24	0.43	0.13	49	0.64	0.14
25	0.34	0.12	50	1.23	0.71

M = Male, F = Female, DIF – differential item functioning value; **Source:** Field Work (2024)

The research was answered using the adjusted threshold (difficulty) values based on sex. The result on table 2, reveals that DIF based on sex range from -0.54 to +1.32. In this DIF analysis, male students represent M were used as reference group, while female students represented by F were focal group. Item 4, 11, 12, 13, 31, 35, 36 and 50 had a weak DIF as their values were greater than 1.0 but less than 1.5 42 items had no DIF, having DIF values less than 1.0. This implies that there is consistency values in gender consideration of test items in the three Education Zone in Ebonyi State (i.e No DIF in gender consideration).

Research Question 3: What are the mean ability of students in internal mathematics mock achievement test among the students in three Education Zone in Ebonyi State?

Table 3: Summary of the Mean Ability Estimates of Students in the Three Education Zones in Internal Mathematics Mock Achievement Test

Zones	N	$\bar{X}$	SD
A	102	0.36	0.12
B	103	2.41	1.05
C	102	1.18	0.29

Source: Field Work (2024)

In the table 3, the mean ability estimates of students in Zone A is 0.36, Zone B is 2.41 and Zone C is 1.18. It indicates slight difference in the mean ability estimates of students in Zone A, B and C due to difference in the difficulty index of the items.

Hypotheses testing

H_{01}: There is no significant difference in the item parameter estimates of the three test forms of internal mathematics mock achievement test used in the three education zone in ebonyi state using one parameter logistic model.

Table 4: Summary of Analysis of Variance of Item Parameter Indices.

Difficulty index	Source of variance	SS	Df	MS	F	Sig	Decision
P-value	Between group	6.631	2	3.316	0.571	0.0361	S
	Within group	25.125	147	0.198			
	Total	31.756	149				

Source: Field Work (2024)

The result in table 4 reveals that the f-values obtained for p-value when test forms from Zone A, B and C were compared is 16.76. Since the p-value (0.0361) is less than 0.05 level of significant. It means that there is significant difference in the difficulty level of the three test forms used in the three Education Zone in Ebonyi State.

H_{02}: The item parameter (difficulty index) do not unction differentially between male and female students in the three Education Zone in Ebonyi State.

Table 5: Summary of Independent t-test Analysis of Item Parameter Threshold Differences in Values for Sex (Gender)

Sex	N	$\bar{X}$	SD	Df	F	Sig	Decision
Male	160	1.97	0.83	3.05	0.126	0.745	NS
Female	147	1.82	0.79				

Mean difference 0.15; **Source:** Field Work (2024)

In the table 5, the value of t-calculated is 0.126 and p-value was 0.745. The p-value is greater than 0.05 (i.e P > 0.05), the null hypothesis was accepted. Therefore, there was no significant difference in the item functioning between sex in the three test forms used in the three Education Zone in Ebonyi State.

H_{03}: There is no significant difference in the mean ability estimates of students in three Education Zone in the three test forms administered using one parameter logistic model.

Table 6: Summary Analysis of Variance of Students' Performance when Scores are Obtained Separately.

Source of variance	SS	Df	MS	F	Sig	Decision
Between group	0.04	2	0.01	0.450	0.020	S
Within group	36.10	304	0.03			
Total	36.124	306				

Source: Field Work (2024)

The table 6, indicate that the t-value is 0.450 and p-value was 0.020. The p-value was less than 0.050. Therefore, the null hypothesis was rejected, which implies that there is significant difference in the mean ability of the students, due to item difficulty level variation in the three test forms from the three Education Zone in Ebonyi State.

Discussion of the Finding

The use of one parameter logistic model (difficulty index) in the calibration of 50-items internal mathematic mock achievement test in the three Education Zone in Ebonyi State, yielded that parameter and ability estimate results that are very important to test developers and users as well as researchers, the item consistency were all based on findings made about them.

Items parameter estimates:

The result from research question 1 in table 1 reveals that Zone A has 34-items very easy with value range from 0.6 to 0.9, 6-item moderate with value 0.5 and 10-items very difficult, value range from 0.1 to 0.4 for Zone B (Ebonyi Central) has 11-items very easy, 8-items moderate and 31-items too difficult for Zone C (Ebonyi South) has 15-items less difficult, 6-items moderate and 29-items too difficult. It can be seen from the table that Zone A uses easy items to generate their CAS whereas Zone B and C uses too difficult items to generate their CAS. It then standard that students in Zone B and C is not fairly treated in terms of CAS to be incorporated in the TAS. Corresponding hypothesis tested with ANOVA at 0.05 level of significant in table 4 indicated that, there is a significant difference in the item parameter estimates of the three test forms used.

The finding is contrary to Nkpone (2001) assertion that, item calibrated using IRT procedures can be used to estimate item

parameter and its consistency across content area. The study focused on determining the estimate of item parameter using 1-parameter and 2-parameter logistic model consistency across content area while this present study result indicate inconsistency on items across the three Education Zones in Ebonyi State.

Estimation of the item parameter consistency values for male and female

The result from research question 2 in table 2 reveals that DIF-value on the performance between sex in the three education Zones were low, which implies that performance corresponding hypothesis 2 tested with f-test analysis of items parameter threshold difference values for sex revealed that f-value (0.26) is less than the critical value at df (305) which is 0.745 at 0.05 level of significant. This result revealed that there was no significant difference in the item functioning between sex in the Education Zones in Ebonyi State, which implies that null hypothesis is accepted.

The finding is contrary to Obinne (2008), who carried study onk DIF effect on WAEC and NECO biology examination items based on sex in Nigeria for 2 years, the result revealed that some items favoured girls while some favoured boys.

Estimation of mean ability

The result from research question 3 in table 3 revealed that the mean ability estimated students in Zone A, B and C were different and corresponding with hypothesis 3 with ANOVA using difficult index table 6 at 0.05 level of significant indicated statistically significantly difference in the mean ability of the students due to variation in test item difficulty in the Education Zones.

The finding is consistent with Forsthy (1998) assertion that, if the item in a test have been calibrated using IRT procedures, then presumably a subject of items could be used to estimate on examinees achieving level and a school achieved level.

Findings

1. There was a significant difference in the item parameter estimates of the three test forms administered to the students.
2. There was no significant difference in item functioning between sex in the Education Zones in Ebonyi State.
3. There was a statistical significant difference in the mean ability of the students due to variation in test items in Education Zones in Ebonyi State.

Conclusion

Based on the result, the following conclusions were drawn.

1. The items used on the test forms differ among the examinees in terms of difficulty level in each Zone, which depict unfair treatment.
2. The item parameter estimates show that the items of the internal mathematics mock achievement test in the three education zones in ebonyi state did not function differentially among the examinees by gender (sex).

3. The mean ability level of the students in each zone differ due to variation in test form difficulty level.

Recommendations

It is therefore recommended that:

1. Test score equating methods should be used by teachers, state examination board and external examination bodies like WAEC and NECO in incorporating CAS to TAS, since there is variation in difficulty level of test items used to generate CAS in Ebonyi State and Nigeria in general.
2. Score assigned to students' responses for cognitive based mock examination should be reported in par-by-item response pattern by the examiner. This will permit better IRT analysis to be performed.
3. Development of state or federal item pool or bank where each state in various geopolitical zone in Nigeria can withdraw and construct reliable and valid assessment instrument for their use. This will ensure uniformity of standard.
4. The differential item functioning effect of items in instrument for measurement of latent trait should always be determined by examiner to ensure examinees are not unfairly treated.
5. Examination bodies and teachers should be trained to become experts in evaluation applying IRT method.

REFERENCES

Afemikhe, O. A. (2007). Assessment and educational and standard improvement, reflection from Nigeria. *A paper presented at the 33rd Annual Conference of the International Association for Educational Assessment held at Balm Azarbaijan September 16th – 21st 2007.*

Akpan, U. & Joseph, E. (2005). *Foundations of educational research.* Ikot Ekpone: Joe Graph Publications, p. 67.

Anastasi, A & Urbma, S. (1997). *Psychological testing.* New jersy: prentice Hall.

Forsthy, K. (998). A comparison of IRT equating and beta 4 equating. *Journal of Educational Measurement, 44*(1), 74-85.

Nkpone, P. (2001). *Application of latent trait models in the development and standardization of physical achievement test for senior secondary school in Nigeria*. Enugu: Ojioguru Graphic Publication, p. 119.

Nworgu, B. G. (2003). *Educational measurement and evaluation: theory and practice.* Nsukka: University Trust Publishers.

Obinne, B. (2008). *Psychometric properties of items of the biology examination conducted by NECO and WAEC,*

applying IRT. Enugu: Marybound Publishers, p.54.

Ojerinde, D. (2002). *Public examinations in Nigeria*. India: India Melrose Books and Publishing Limited.

Okoye, R. O. (2007). Continuous assessment at the tertiary education level: Theory vs practice. *Sokoto Educational Review, 5,* 23-45.

Ugo, N. (2012). Comparison of item and person statistics based on item response theory versus classical test theory. *Journal of Qualitative Education, 4*(1), 26-135.

Wendy, Y. (2002). Scaling and equating: *A paper presented at the New York technical conference, October 12, 2002*. Retrieved April, 18, 2024 from http://www.erst.