Journal of Educational Research on Children, Parents & Teachers, Volume 4, Number 1, 2023, 68-79
ISSN: 2664-3812, https://ercptjournal.org/

DIMENSIONALITY OF NECO AND WAEC PRACTICAL PHYSICS; ASSESSINGTEST QUALITY USING STATISTICAL FIT OF PARTIAL CREDIT MODEL

Adonu, I. Ifeanyi[1]; Eze, A.E[2]; Odo, J.A[3]; Ihediohamma, C.A[4]; Ejinkonye, F.O[5*]
[1,2,4,5]Department of EDUPSY (Measurement and Eval. Unit), Federal College of Education Eha-Amufu, Enugu State.
[3]Department of General Studies, Federal College of Education Eha-Amufu, Enugu State.
***Corresponding Author**: Ejinkeonye, F.O; ejinkonye.felicitas@fceehamufu.edu.ng*

Abstract

The aim of the study was to assess the dimensionality of WAEC and NECO practical physics tests using partial credit model. Two research questions and two hypotheses were formulated to guide the study. An instrumentation research design was adopted. A sample of 670 SS3 students were drawn using a multistage sampling procedure. WAEC and NECO for years 2020 and 2021 were used for data collection. The data was subjected to analysis using infit, outfit statistics and response residuals. The study found out that majority of the items on the practical tests had fit statistics that were within the accepted/recommended range that depicts unidimensionality and in effect valid items. The hypothesis indicated consistently that there is no significant difference in the fit statistics of WAEC and NECO practical physics tests. Based on the findings of the study it was recommended that test makers and stakeholders in the education industry are to always utilize IRT model and precisely partial credit analysis for development and analysis of polytomously scored items to enthrone objectivity fully into assessment.

***Keywords**: Dimensionality, Infit, Outfit statistics, PCM, WAEC, NECO*.

Introduction

The West African Examination Council [WAEC] and the National Examination Council [NECO] are two examining bodies in Nigeria. The major responsibility of the two examination bodies is the summative assessment of secondary school students in their school subjects like physics and other subjects. This assessment is done both at ordinary and advanced levels. WAEC was established in 1950 following a report submitted to British council of state which was adopted by four West African governments- Nigeria, Ghana, Sierra Leone and Gambia. [WAEC 2002]. And NECO was established in 1999 with the motive for Nigeria to have independent national examination body that has the same standard with WAEC and has headquarters at Minna [NECO, 2001]. NECO conducts entrance examination into unity secondary schools and ordinary level school certificate examinations.

The two examination bodies are ultimate for secondary schools in Nigeria in test development and it would be tantamount to over emphasizing the obvious to say that they are required to employ the global best practices in test development. At the core of

secondary school science subjects is physics and a minimum of credit pass is required for entry into tertiary institutions for science and technology based courses. Development and psychometric quality analysis of WAEC and NECO practical physics questions has since inception been anchored on classical test theory. Ndalichako and Rogers(1997), Obinne (2008), Adonu (2014), firmly supported the same view when they stated that almost every educational instrument in Nigeria if not all still utilize classical test theory. It is quite unfortunate according to them that the classical test theory (CTT) they rely on has such limitations as circular dependency (item statistics depend on population and population parameters depend on items), weak theoretical assumptions etc. Consequently, CTT cannot be used to provide solutions to such measurement problems as test score equating, computerized adaptive testing identification of biased items etc.

There is a modern measurement theory that would obliterate measurement problems and enthrone objectivity for psychometric quality and analyses of items by WAEC and NECO. This is the item response theory – (IRT). The assessment of dimensionality of test items is one of the most important ways to verify the validity and reliability of such items (Umobong and Udeme, 2017). This study categorically stated that the evaluation of dimensionality of the test is a requisite stage in getting evidence for validity of interpretation concerning total score. This is more so when the test development and analyses are executed using the IRT format.

The assessment of psychomotor/practical physics skills is polytomously scored rather than dichotomously scored and as such partial credit scoring is inevitable in practical physics assessment situations. The usual motive for partial credit scoring as stated in Masters (1982) is the hope that it will lead to a more precise estimate of persons ability than simple dichotomous score. Since practical physics is polytomously scored in partial credit format, the IRT analysis for practical physics is the partial credit model. This is an adaptation of Rasch measurement model (one parameter logistic model).

As stated before dimensionality verifies the validity or otherwise of an instrument. Unidimensionality connotes that a single latent variable fully explains task performance and this implies that the item is valid. Conversely, multidimensionality means more than one latent variable explaining task performance. An item that is multidimensional (when the responses are dominated by more than one major factor) is not valid.

In the recent past, studies have advanced arguments with respect to dimensionality assessment and understanding of test item structure as a crucial step in accessing the examinees abilities (Zhang, 2007, Jang and Roussos, 2007). Also, Tate (2003) posited that developing and evaluating large scale tests require the assessment of dimensionality as this avails us empirical backing for the content and cognitive aspects of the test validity.

Through a thorough examination of dimensionality research have been able to associate interpretation with statistical observables for a better understanding of testee by item interactions. Hence investigating the dimensionality of a test helps to provide

evidence for different aspect of validity issues. Zhang (2007) noted that assessing the dimensionality of items forms the very basis of statistical analysis of data.

Apart from the role of dimensionality assessment in understanding the statistical observations of testee by item interactions. Numerous researches have advanced the following plausible reasons why the dimensionality assessment of instruments is inevitable:

- dimensionality assessment detects the presence of biased items through the understanding of multidimensional test structure that could potentially arise from factors that are irrelevant to the underlying construct (Tate, 2003).
- it also helps to identify and eliminate potential threats to validity through the examination of items with different item functioning arising from multidimensional test structure attributed to construct irrelevant factors. Tate(2003) categorically stated that the study of dimensionality of items which exposes the reason why such items are biased would help the test maker to avoid such bias in future test construction.
- also Sfone and Yeh as summarised in Umobong and Udeme (2017) indicated that dimensionality assessment which is the investigation of internal structure of a test avails us the opportunity of identifying the domain of knowledge being measured by the item, explains the probable multidimensionality of the test score and helps identify construct variance that are not relevant.
- Adonu (2014) posited that total scores from a unidimensional test provides us with technically valid items with clear meaning. This is because the irrelevant construct which would have made the items multidimensional would have been jettisoned and it would now not be ambiguous what the scores represent having assessed the dimensionality.

Dimensionality assessment can only be achieved through IRT. But despite the advantages of IRT over CTT in test analyses, both public and school based examination in Nigeria continue to utilize CTT in test development. (Obinne, 2008; Adonu, 2014; Umobong and Udeme (2017). IRT produces item statistics independent of testee sample and person statistics that does not depend on items administered. This is known as *invariance property* of item/person statistics in IRT and has been illustrated theoretically (Hambleton, Swaminathan and Rogers, 1991). This *invariance property* helps us to overcome those measurement problems not possible in CTT such as computerized adaptive testing and test score equating.

By and large, IRT is a mathematical model that attempt to model the relationship between examinee's ability and the probability of his responding correctly to a test item. The item parameters in IRT neither depends on population nor does the population parameters depend on the items. But for this crucial property of IRT the statistical rigours and complexity can hardly be justified (Fan,1998). The partial credit model is an adaptation of Rasch one parameter logistic model of IRT. IRT model is predicated on the

fact that each examinee possess a latent ability or a value of the construct gauged by the item and each of the test item that indicates a certain value of the construct.

All IRT model including the partial credit model (PCM) is characterised by specific objectivity and uni dimensionality. According to Mellember (1994) specific objectivity implies that comparison of two item difficulty are assumed to be independent of any group of subject studied and does not depend on any set of item administered. And according to Carlson (1993) Uni dimensionality implies that a single latent variable (construct) fully explains task performance. PCM being adaptation of Rasch (one parameter) model has some assumptions such as equal item discrimination, local independence and low (constant) susceptibility to guessing. The probability that a testee responds correctly to an item is a logistic function of the difference between testees ability and the difficulty level of the item. Based on the above characteristics and assumptions of PCM; each testees pattern of responses to items is determined by the testees ability and item difficulty. This produces a response pattern that typifies Rasch measurement scale. The study would among other thing explain the statistical fit and misfit among the data obtained.

Item bi serials used in validity assessment in CTT has the disadvantage of being sample dependent. In IRT the validity/ dimensionality of a test is assessed with respect to statistical fit of each item to the model utilized. According to Korashy (1995), the analysis of fit is a check on the validity. When the fit statistics of an item is susceptible, then its valid and if a set of item fit the model, it is evidence that they refer to unidimensional ability. Bryce (1981) noted that a large positive fit statistics implies no fitting while a low fit statistics nearer one (1) implies a better fit and this enables the test developer to identify and delete misfitting/ bad items. For PCM in specific terms the infit and outfit mean square statistics has goodness of fit when the infit and outfit of the items has the range 0.7 to 1.5 (Opsomer, Jenson, Nusser, Drignei and Amemiya 2002; Bond and Fox, 2013). It was noted by Ostini and Nering (2006) that to asses the model fit to PCM we use the residual based measure, that fit can be classified in terms of generality of application and that fit can be assessed in terms of the fit of specific group of items from a test if specific hypothesis about the fit is to be tested. Finally, response residuals could be summed over respondents to obtain an item fit measure and accumulation is done with squared standardised residual divided by total number of respondents to obtain the mean chi square (fit statistics), (Masters and Wright; (1997); Ostini and Nering; 2006)

Lian and idris (2006) studied the algebraic solving the ability of form four students in using linear equation. Forty (40) form four students in a Malaysian secondary school formed the sample for the study. Qualitative and quantitative approach were utilized to asses the students algebraic solving ability. Partial credit analysis was utilized for the data analysis using winstep software programme to estimate validity, reliability index and difficulty index. The results of the study showed that the fit statistics (infit and outfit) fall between 0.7 and 1.3 and thus showed sufficient validity, the item reliability index of 0.91.

On the whole, the result provided evidence for the significance of their model in assessing algebraic solving ability.

Wallace, Prather and Duncan (2012) investigated the study of general education astronomy students understanding of cosmology iii. Students responses to their instrument were analysed using partial credit model to assess the reliabilities of the four section of the instrument. A sample of 4359 students responses to the instrument in four semesters of 2 academic session of University of Arizona was used. The study estimated item difficulty and reliability of their school/teacher instrument. The result of the study showed appropriate item difficulty, reliability of the instrument and subsequently provided insight to conceptual knowledge and abilities of the students.

The two examples of studies presented above are few instances where PCM have been used for analysis. But a thorough search through the other studies so far showed that it is mostly the dichotomously scored items of our subject that their qualities are being studied. This is the concern with which this study assessed the dimensionality of a polytomously scored practical physics in our NECO and WAEC examinations to ascertain the robustness with respect to whether our practical tests asses what it is meant to measure.

In WAEC and NECO physics, dichotomously scored paper (objective) has about one-third of the total score in the subject (ie 50 marks). Polytomously scored papers-practical and essay have 50 and 60 marks respectively, together giving slightly above two third of the total score in physics in both WAEC and NECO. In an extensive and diligent sojourn through literature, it was observed that less than 3% of studies in test development and/or analyses have ever attempted polytomously scored aspect of their various subject matter. This is an obvious lacuna, and in part this study has been able to address it. It is also absolutely necessary at this stage in researches to fully explore the properties of our tests using the item response theory approach given the advantages of the IRT approach earlier mentioned.

Purpose of the Study

The objective of the study was to investigate the dimensionality of WAEC and NECO practical physics tests using partial credit model

The specific objectives of the study were:

1. investigate the validity and proportion of fit of practical physics test produced by NECO to PCM
2. investigate the validity and proportion of fit of practical physics test produced by WAEC to PCM
3. compare the fit statistics of NECO and WAEC practical physics tests

Research Questions

The following research questions guided this study

1. How valid is NECO practical physics test and what is the proportion of fit of NECO practical physics test to partial credit model
2. How valid is WAEC practical physics test and what is the proportion of fit of WAEC practical physics test to partial credit model.

Hypotheses

H_{O1}. There is no significant difference ($p<.05$) in the fit statistics of WAEC 2020 and NECO 2020 practical physics tests

H_{O2}. There is no significant difference ($p<.05$) in the fit statistics of WAEC 2021 and NECO 2021 practical physics tests

Research Method

Instrumental research design was used for the study. Instrumentation research according to international centre for educational evaluation (1982) is a study aimed at introduction of new or modified content, procedure in technology or instruments of education practice. This study aimed at validation and verification of efficacy of practical physics tests from the examination bodies.

The population of the study was 13,050 students ie all the SS III physics students that registered for 2021/2022 physics senior secondary school certificate examination of WAEC and NECO in 282 public secondary schools in Enugu state. A sample size of 670 students about 5% of the population was used for the study. Sample was composed using multi stage procedure. The simple random sampling was used to select three out of six education zone, stratified the schools into local government areas, and from various strata purposive sampling were used to select two schools where we have highest number of SSIII physics students in each local government area.

The instruments utilized for data collection were the NECO 2020/2021 and WAEC 2020/2021 practical physics questions. The practical work and psychomotor demands of each questions plus the essay questions asked on each question were collapsed into eight items. This gave a total of twenty four items for each examination question. The validation and reliability of the instrument were not done. This is because the instruments are from standard testing body and as such they would have validated and verified the reliability. More so validation of this instrument is the major thrust of the study.

With the help of physics teachers and trained research assistants, the researchers administered the instrument to the sampled school students in life laboratory sessions. The aim of the study was explicitly explained to the students and were requested to perform the practical with every seriousness and report same. They took the practical

questions in turns and at the end the sampled students responded to every question. Their reports of the practicals were collected, scored and collated. Their achievement scores were used for the analysis.

The data collected were analysed using winstep 3.80.1 computer programme of partial credit model analysis. The research questions were answered using item response theory descriptive statistics estimation procedure of infit and outfit statistics. Hence the validity is appropriate or considered unidimensional when the infit and outfit statistics is within the range 0.7-1.5. (Opsomer, et al, 2002; Bond and Fox, 2013). Also to test the hypothesis, independent t-test was carried out using SPSS at 0.05 level of significance. Ostini and Nering (2006) posited that to test hypothesis about fit, the response residual is summed over respondents (for specific group of items from a test) to obtain an item fit measure and accumulation is done with squared standardized residual over the total respondents to obtain mean fit statistics.

Results

Research Question One: How valid is NECO practical physics tests and what is the proportion of fit of NECO practical physics tests to partial credit model?

Table 2: Validity of test items of practical physics test by NECO (fit statistics) for years 2020 and 2021

Item	Infit 2020	Outfit 2020	Infit 2021	Outfit 2021
1	1.03	0.84	1.04	1.03
2	1.09	1.01	1.05	1.06
3	1.00	0.99	1.48	0.45
4	1.12	1.29	0.86	0.78
5	0.93	0.91	1.01	1.04
6	0.81	0.79	1.11	1.13
7	1.02	1.10	1.13	1.03
8	1.00	0.91	1.33	1.86
9	0.94	0.78	1.08	1.17
10	0.98	1.00	1.01	1.02
11	0.74	0.71	1.09	1.10
12	0.83	0.85	0.89	0.83
13	0.93	0.96	1.24	2.38
14	0.89	0.84	0.96	0.93
15	0.97	0.90	0.96	0.85
16	1.32	1.29	1.03	1.17
17	1.52	2.50	0.97	0.83
18	1.51	1.74	0.88	0.93
19	1.41	1.51	0.88	0.85
20	1.08	1.10	0.81	0.76
21	1.00	1.02	0.93	0.94
22	1.06	1.01	0.86	0.84
23	0.84	0.79	0.96	0.90
24	0.92	0.90	0.91	0.79
Mean	1.04	1.07	1.02	1.07
S.D	0.20	0.38	0.15	0.26

Journal of Educational Research on Children, Parents & Teachers, Volume 4, Number 1, 2023, 68-79
ISSN: 2664-3812, https://ercptjournal.org/

Table 1 indicates the result of fit statistics of NECO practical physics for years 2020/2021 June/July examination using PCM. The result indicated that the test items for 2020 had infit statistics range of 0.83 to1.53 and outfit statistics range of 0.84 to 2.50. Item 17 has infit and outfit of 1.52-2.50 and item 18 has outfit of 1.74. For these two items, their infit/outfit were beyond the accepted range. The fit statistics of NECO 2020 therefore are perfectly valid and thus unidimensional except for item 17and 18. 22 out of 24 had a good fit to PCM. This means that 91.57 % or 0.92 of the items of NECO 2020 had a good fit to PCM.

Also, table 1 showed the result of fit statistics of NECO 2021 to have infit statistics range of 0.88 to 1.48 and outfit statistics range of 0.76 to 2.38. Items 8 and 13 had outfit statistics of 1.86 and 2.38. These two items have their outfit beyond the accepted range. Thus, the fit statistics of NECO 2021 are valid and hence unidimensional except for items 8 and 13. So 22 out of 24 had a good fit to PCM.

Research Question Two: How valid is WAEC practical physics test and what is the proportion of fit of WAEC practical tests and what is the proportion of fit of WAEC Practical physics tests to pcm.

Table 2: Validity of test items of practical physics tests by WAEC (fit statistics) for years 2020 and 2021.

Item	Infit 2020	Outfit 2020	Infit 2021	Outfit 2021
1	1.03	0.84	1.04	1.03
2	1.09	1.01	1.05	1.06
3	1.00	0.99	1.48	0.45
4	1.12	1.29	0.86	0.78
5	0.93	0.91	1.01	1.04
6	0.81	0.79	1.11	1.13
7	1.02	1.10	1.13	1.03
8	1.00	0.91	1.33	1.86
9	0.94	0.78	1.08	1.17
10	0.98	1.00	1.01	1.02
11	0.74	0.71	1.09	1.10
12	0.83	0.85	0.89	0.83
13	0.93	0.96	1.24	2.38
14	0.89	0.84	0.96	0.93
15	0.97	0.90	0.96	0.85
16	1.32	1.29	1.03	1.17
17	1.52	2.50	0.97	0.83
18	1.51	1.74	0.88	0.93
19	1.41	1.51	0.88	0.85
20	1.08	1.10	0.81	0.76
21	1.00	1.02	0.93	0.94
22	1.06	1.01	0.86	0.84
23	0.84	0.79	0.96	0.90
24	0.92	0.90	0.91	0.79
Mean	1.04	1.07	1.02	1.07
S.D	0.20	0.38	0.15	0.26

In table 2 the result of fit statistics for WAEC practical physics test item using pcm are presented.

In the year 2020 the infit statistics ranged from 0.75 to 3.07. Items 13 and 17 had their outfit of 3.08 and 2.22 respectively. Apart from these two items, the result showed that the remaining 22 items are highly valid and hence unidimensional. 22 out of 24 items had a good fit to PCM and by implication 0.92 of the items had perfect fit to PCM.

Also, Table two showed the result of fit statistics of WAEC 2021 practical physics items to have infit range of 0.76 to 1.45 and outfit range of 0.72 to 2.14. Items 5 and 17 with outfit of 2.14 and 1.52 had their outfit outside the accepted range. It therefore implies that 22 out of 24 items of WAEC 2021 hard their fit statistics within the range that makes them valid and consequently unidimensional. Also, 0.92 of WAEC 2021 practical physics item fit the partial credit model.

Ho_1: There is no significant differences ($p<0.05$) between the fit statistics of NECO 2020 and WAEC2020

Table 3: PCM fit statistics for NECO 2020 and WAEC2020

Variable	n	$\bar{X}$	SD	df	t	sig	Decision
NECO2020	50	18.79	12.41	49	.78	.46	NS
WAEC2020	50	21.33	20.58				

$\propto = 0.05$ NS

The result of table 3 showed t value of 0.76 and a probability of 0.46 associated this probability is more than the associated ∝level of 0.05. as a result, we fail to reject the null hypothesis. This therefore indicates that “there is no significant difference between the fix statistics of NECO2020 and WAEC2020.

Ho_2: There is no significant difference ($P<.05$) between the fit statistics of NECO 2021 and WAEC 2021.

Table 4: t-test analysis of the difference ($P<.05$) between the fit statistics of NECO 2021 and WAEC 2021

Variable	n	$\bar{X}$	SD	df	t	sig	decision
NECO2021	50	21.27	9.69	49	1.37	.18	NS
WAEC2021	50	17.95	14.36				

$\propto=0.05$ NS

The result in table 4 indicated t value of 1.37 and a probability of 0.18 associated. This probability is greater than associated ∝ level 0.05. Consequently, we fail to reject the null hypothesis. This implies "there is no significant difference between the fit statistics of NECO 2021 and WAEC 2021.

Discussion of the Findings

Research question one was aimed at examining the fit statistics of NECO practical physics tests for 2020 and 2021 June/July certificate examination the result indicated that 92% of both 2020 and 2021 and NECO practical physics questions had the items fit the requirement of partial credit model. For the item to fit the model implies that the items are valid and thus 92% of the items are valid in accessing students' ability in physics hence 0.92 proportion of the practical items were unidimensional while about 0.08 proportion had misfitting items. Overall, therefore the test/instrument met the assumption of unidimensionality similar to the finding in this first research question is the finding of Umobong and Udeme (2017), and Green and Frantom (2002) who in the same vein reported that irrespective of few misfitting items, the instrument measured a unidimensional construct.

The aim of research question two was to access the fit statistics of WAEC practical physics test for 2020 and 221 June/July certificate examination. The result for both years showed the 22 items in each had a good fit to the partial credit model and two items had either their infit or outfit outside the accepted range (0.7-1.3). Hence 0.9 proportion of the items were unidimensional. The instrument therefore met the assumption of unidimensionality since only two items had a bad fit out of twenty four. This is also in concord with Umobong and Udeme (2017), Green and Fanton (2002) and Adonu (2014) that discovered few misfitting items and regardless of that adjudged the instrument unidimensional and in consequence valid.

Two hypotheses were tested and the results obtained indicated that there were no significant difference between WAEC2020 and NECO2020; WAEC2021 and NECO2021fit statistic of practical physics items consistent similarity in the quality of practical physics tests in the two examination bodies. This result is similarly in agreement with the findings of Adonu (2014) that reported no significant difference in the validity of WAEC and NECO practical physics test as indicated in the partial credit model framework.

Conclusion

The analysis of fit statistics of WAEC and NECO practical physics tests anchored on partial credit model framework has revealed that almost all items on the test measured unidimensional constructs. This means that most of the items were valid. Only a very few items (2 out of 24) were not unidimensional.

Recommendations

1. All examination bodies should subject their polytomously scored test items (practical and Essays) to partial credit analysis. This will enable them to come up with only valid items meant to measure unidimensional traits. Without this approach the item-population circular dependency syndrome of the classical test theory can hardly be surmounted.
2. Test developers, teachers, examination bodies (NECO,WAEC,NABTEB), etc and infact the entire psychometric community should be groomed in item development and analysis using the partial credit model framework. This will occasion the compelling advantages of item response approach to the measurement community.
3. Due to the relative complexity of IRT analysis, the stakeholders in education should train and retrain the test makers including teachers of all categories in the use of IRT analytical software for test development. This will annihilate the teachers' phobia of jettisoning CTT for IRT in test development and usage.

References

Adonu, I.I (2014). Psychometric analysis of WAEC and NECO practical Physics tests using partial credit model. (An Unpublished Ph.D thesis). University of Nigeria.

Bond, T.G and Fox, C.M (2013). Applying the Rasch model: Fundamental Measurement in Human Sciences. 3rd ed. Mahwah,N.J. Lawrence Farbaum Associate.

Bryce, T.K(1991). Rasch Fitting. British Educational Research Journal, 7(2)137-153.

International Centre for Educational Evaluation (1982). A Conference on Priority in Educational Research in Nigeria. Institute of Education, University of Ibadan

Carlson J.E (1993). Dimensionality of NAEP Instruments that Incorporate Polytomously Scored Items. A Paper Presented at Annual Meeting of American Educational Research Association. Atlanta.

Fan, X . (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item / Person Characteristics. Educational and Psychological Measurement, 53 (3): 357-382

Green, KE and Frantom CG (2002). Survey development and validation with the Rasch model. Paper presented at International Conference on Questionnaire Development, evaluation and testing. Chalseton. Mahwah, New Jersey: Lawrence

Hambleton R.K; Swaminathan, H and Rogers, H. J. (1991).Fundamentals of Item Response Theory. Newbury park CA:Sage.

Jang EE and Roussos,L(2007). An Investigation into Dimensionality of TOEFL using Conditional Covariances based on Parametric Approach. Journal of Educational

Measurement,44:1-22.

Korashy, A F. (1995). Applying the Rasch Model to Selection of Items for Metal Ability. test. *Educational and Psychological measurements 55(5): 753-763*

Lian, H and Idris, N (2006). Assessing Algebraic Solving Ability of Form Four Students. *International Electronic Journal of Mathematics Education*, (1), 55-76. Also available at www.iejme.com.

Masters G.N (1982). A Rasch Model for Partial Credit Scoring, Psychometrica,47(2):150 Zhang J. (2007). Conditional Covariance Theory and Detection for Polytomous Items. *Psychometrica*, 72(69-91).

Masters G.N and Wright B.D (1997). The Partial Credit Model in W.J Vander Linder and R.K . Hambleton (Eds). *Handbook on Modern Item Response Theory* :101-121.

Mellembergh G.J(1994). A Unidimensional Latent Trait Model for Continuous Item Responses. *Multivariate Behavioral Research 29*:223-236.

NECO (2001). Facts about National Examination Council. Nigeria

Obinne, A.E (2008). Comparison of psychometric properties of WAEC and NECO test items under IRT. (Unpublished Ph.D thesis), University of Nigeria.

Opsome J.D; Jensen H.H; Nusser S M; Dregnei D Amemiya, Y. (2002): Statistical Considerations For The USPDA Food Insecurity Idex. Www.Card.Iastate.Edu. Downloaded 10 feb 2021

Ostini R and Nering M.L (2006). Polytomous Item Response Theory Models. Quantitative Application in Social Sciences. International Educational and Professional Publishers. London New Delhi.

Tate R.L (2003). A Comparison of Selected Empirical Methods for Assessing the Structure of Responses to test items. *Applied Psychological Measurements, 27*,159-203.

Umobong, M.E. & Udeme, E.T. (2017). Dimensionality of National Examination Council`s Biology Examinations. African Journal of Theory and Practice of Educational Assessment 5: 15- 30.

WAEC (2002). West African Examination Council. West African Book Publishers Limited Lagos