LASU International Journal of Arts and Social Science Education
Vol. 2, No. 2, Nov. 2024

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

RELATIVE EFFICIENCY OF SENIOR SECONDARY SCHOOL CERTIFICATE EXAMINATION IN ECONOMICS IN NORTH-CENTRAL NIGERIA

[1]Rasaq SULYMAN (Ph. D), [2]Rashidat O. ABDULSALAM-NUHU (Ph. D)
[1]Arts and Science Education, Faculty of Education, Kwara State University, Malete, Ilorin.
rasaq.sulyman@kwasu.edu.ng +2348032444233 https://orcid.org/0009-0003-5950-9389
[2]Educational Foundations, Faculty of Education, University of Abuja, Nigeria.
rashidat.omotayo@uniabuja.edu +2348060361777 https://orcid.org/0000-0001-8724-7484

Abstract
The study aimed at investigating the relative efficiency of senior secondary school certificate examination in Economics in north-central Nigeria. The objectives of this study were to determine: (i)the IRT model fit for 2013 WASSCE Economics multiple-choice items; (ii) the IRT model fit for 2014 WASSCE Economics multiple-choice items; and (iii) the relative efficiency of 2013 and 2014 WASSCE multiple-choice Economics items. The study employed the descriptive survey design and the population comprised all 157,156 senior secondary school three (SS3) students in the North-central zone of Nigeria. The target population was all senior school three (SS3) students that offered Economics in their Senior School Certificate Examinations (SSCE). A sample of 1,120 students was drawn from 32 senior secondary schools in North-central, Nigeria. Multistage sampling procedures involving simple random, proportionate and purposive sampling techniques were used to select the sample. The research instruments were WASSCE 2013 and 2014 Economics paper 1, information indices of Akaike information criterion (AIC) and Bayesian information criterion (BIC) were employed to analyse research questions while t-test statistic was used to test for research hypothesis formulated. The study concluded that tests items of 2013 and 2014 WASSCE Economics were good and of the same qualities. This implied that Economics test items used for the assessment have the same model fit and can be used interchangeably. The study recommended that examination bodies should retain and sustain items with equal model fit statistics for future examinations.

Key words; Information function, Interchangeable, Model fit statistic, Relative efficiency

Introduction

Relative efficiency of the test is an important index in determination of the test qualities. It gives test experts an avenue to identify the trait of each of the test item to ensure dependability and reliability of the test. Efficiency of the two forms of test indicates the variability of the degree of information provided by the forms on the scoring scale. Relative efficiency is concerned with the test form that behaves well in providing greater amount of information at different point on the ability scoring scale. Test is a veritable tool used for gauging learners' progress and accomplishment. It is a measuring tool used to assess an individual's performance in a given task. It is used to evaluate and assess a student's expertise in a specific subject or an area within the subject. Sulyman (2021) stresses that different educational institutions use test to measure the proficiencies of students from different fields of study. It is used in a systematic practice to compare the individual's performance with a chosen standard of performance. It is capable of shaping an examinee's true ability so as to place learner on the ability level in order to avert dropout. Schools and higher institutions used test scores and grades for admission.

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

The incongruous evaluation of examinee's test scores may lead to a wrong decision where certain candidate from a particular academic programme would be excluded or limited by the ability of an examinee to practice the profession of his/her choice. The importance placed on the results of standardized testing make it necessary that, the test scores must provide a fair and equitable evaluation of the skills and abilities the test is purported to measure. The provision of a fair and equitable means of psychological or educational assessment is one of the major reasons for the existence of standardized testing. Abdulsalam-Nuhu (2018) emphasized that Test is expected to treat every examinee equally without bias. A test may become nonstandard when an unfair evaluation of an examinee's skills or abilities is done, which resulted into measurement error. The measurement error may occur due to physical testing environment, the administration of the test, the actual content of instruction and nature of psychometric characteristics of the test. Whether a test score is interpreted with regard to a norm, criterion or both, it must meet a specified or established standard. The norm may be established independently or by statistical analysis of a large number of participants.

It is important that the test scores obtained from the test are expected to reflect the most precise estimates of abilities and skills of the students. Doran, Moses and Eignor (2010) as cited in Sulyman (2021) underscored that examinees must be treated with fair-mindedness and impartiality on any ability test they took. A standardized test administered and scored must be treated in an orderly manner to ensure legal defensibility (Thissen & Wainer, 2011).

Statement of the Study

In recent time there is variability in the performance of the students at West African Senior Secondary Certificate Examination (WASSCE) relative to learner's ability as well as difference in examinee results grade of different year for the same purpose is of concerned. Researchers, educationists and general public are interested in relative efficiency of test and test items used for Senior School Certificate Examinations (SSCE) in Economics because of the emphasis placed on yardstick used for Universities admission. Relative efficiency is used to determine the quality of two different items while constructing test items. Therefore, the study aimed at relative efficiency of senior secondary schools Economics conducted by WASSCE in 2013 and 2014. This is to detect quality and effectiveness of item as well as the extent to which result of two tests can be used interchangeably at different times considering students' ability. So, the need to investigate and compare the quality of test used for the admission purposes arises which the present study aimed to address.

Purpose of the Study

The main purpose of this study is to determine the relative efficiency of Senior School Certificate Examination in Economics in North-central, Nigeria. Specifically, the study intends to find out;

1. the IRT model fit for 2013 WASSCE Economics multiple-choice items;
2. the IRT model fit for 2014 WASSCE Economics multiple-choice items;
3. the relative efficiency of 2013 and 2014 WASSCE multiple-choice economics items.

Research Questions

In order to guide this study, the following research questions were raised.

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

1. Which of the IRT model fitted 2013 WASSCE Economics multiple-choice items?
2. Which of the IRT model fitted 2014 WASSCE Economics multiple-choice items?
3. What are the relative efficiency of 2013 and 2014 WASSCE multiple-choice Economics items?

Research Hypothesis

Based on the research question raised, one research hypothesis was formulated to guide the study.

Ho_1: There is no significant difference between of 2013 and 2014 WASSCE multiple-choice Economics items.

Literature Review

De Ayala (2009) asserted that the comparison between two people should be independent, and then the two people and their observed reactions should be independent of the persons taking the test. He also clarified that relative efficiency of the test is the ratio of the test information for two test forms across the entire score scale whichvaries as a function of the ability. De Ayala (2009) as cited in Sulyman (2021) described the relative efficiency of one score (X) to another score (Y) as the level of their information. The scores X and Y may come from different instrument with certain dimension of idea in two different ways. The rate of the two curves is called relative efficiency. Samuel and James (2021) re-established that IRT models captured modeling measurement error for observed against expected values, as well as assessing the degree to which items might be measuring abilities differently between subgroups.

Parameter estimation model are essential in determination of information function. Items estimated with parameters model provide more information than those without parameter model. The dimension(s) underlying performance determine the number of parameters to be used for estimation. For instance, if responses to a test question is dichotomous and the difficulty level is only an item parameter that makes the item functioning, then the probability of the correct answer derived by test takers is the function of their ability and item difficulty, resulting in the Rasch/one parameter model. If items are allowed to vary in terms of their ability to discriminate examinees on the latent continuum, then the item discrimination parameter is added, which is known as the two-parameter model. The three-parameter model takes into consideration, the probability that low-ability examinees select the correct option based on partial information or guesses, but the four-parameter model considers the probability that high ability respondents select the wrong option due to carelessness or fatigue (Hoan, 2021).

One parameter model signifies the difficulty level of items; 2PL model indicates the discrimination level of items while the 3PL model adds a guessing parameter (the lower asymptote) to care for the effect of guessing. 4PL model points out carelessness or error due to anxiety. Carelessness or error due to anxiety may result in to estimation biases than what guessing does on the test. Rulison and Loken, (2009) reported that an examinee's ability is underestimated when examinee makes careless errors at the beginning of a test. Barton and Lord (1981) introduced a fourth parameter (the upper asymptote) in the four-parameter (4PL). The model cares for underestimation of the examinees performance due to errors as a result of heedlessness. Liao, Ho, Yen, & Cheng, (2012) as cited in Sulyman (2021) reported that 4PL model helped the examinee to recover from mistakes without reducing measurement precision and efficiency. They

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

added that the 4PL model is a more robust mechanism against divergent responses. Magis (2013) reported that 4PL model is a useful model to detect fitted data especially when there is careless or inattention patterns of the examinees. Ayanwale, Adeleke· and Abayomi (2018) said that Four-parameter model or Upper asymptote is proficient of caring for incorrect answer as a result of negligence on the part of the examinees.

Reckase (2009) as cited Sulyman (2021) suggested that in the determination ofthe number of dimensions underlying the tests, the responses of the examinees that answer the test needs to be subjected to parallel analysis. Lesdema& Valero-Mora, (2007) maintained that factor is retained when its eigenvalue is found greater than the mean of the eigenvalues from its randomly generated equivalent

Model Fit Assessment

The model that produced the best fit to the data is adjudged the model that fit the data. To achieve this fit, several measures apply. Oguoma, Metibemu and Okoye (2016); (Finch & French, 2015), prominent among the measures include Chi-square difference test and use of information indices. Information indices are simply measures of variance not explained by a model, with an added penalty for model complexity. Among the most popular of these indices are the Akaike information criterion (AIC; Akaike, 1973), the Bayesian information criterion (BIC; Schwarz, 1978), and the sample-size-adjusted BIC (SBIC; Enders & Tofihi, 2008). These information indices are computed using the -2loglikelihood chi-square value and is interpreted such that the model with the lower value exhibits a better fit to the data. In addition, the chi-square and likelihood ratio goodness of fit tests the null hypothesis that two nested models provide the same fit to a set of data. A statistically significant likelihood indicates a difference in the models under examination.

Methodology

The research design adopted for this study was a descriptive survey design. This is the most appropriate research design for this study. This is because the purpose of the study was to examine the relative efficiency of 2013 and 2014 Senior School Certificate Examination Economics in North-central, Nigeria. The study tends to source information from large respondents without any manipulation of any of the variables. The population for this study adopted all 157,156 public senior secondary schools three (SS3) students in the North-central Nigeria. The target population for the study was all Senior Secondary school 3 (SSS3) students who offfered Economics in all senior secondary schools in North-central, Nigeria in 2017/2018 academic session. Three states were randomly selected from the six states that constituted the North-central region of Nigeria.Kwara, Kogi, Niger and Federal Capital Territory Abujaconstituted the sample of the study from the north-central, Nigeria.

Multistage sampling procedure was adopted to select the sample for the study. At the first stage, simple random sampling technique was employed to select three states and Federal Capital Territory Abuja. A total sample 1,120 students were drawn from the three (3) selected states and Federal Capital Territory Abuja in north-central of Nigeria. At the second stage, five percent of the total number of the schools offering Economics in each of the three states and Abuja were used to select to 32 schools. At the third stage, 35 students that offered Economics were purposively chosen from each of the thirty-two 32 schools selected as sample. Purposive sampling technique was adopted because Economics as a subject is meant only for Humanities and Business students. SS3 students were

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

selected as respondents because they were in best position to respond to the tests since they were preparing for WASSCE which is the focus of this study.

The 2013 and 2014 WASSCE Economics papers 1 were adopted as instrument for this study. A total 100 multiple –choice items were used for the study. Economics Paper 1 of 2013 and 2014 contains 50 objective items each were used. The instruments were titled Economics Achievement Test (EAT 1) and (EAT 2) for 2013 and 2014 respectively. The same marks were allocated to both forms and the same number of questions was asked to answer. The instrument was assumed to be valid since it was prepared and administered by an examination body then, it was assumed to be reliable because it has been used for the certification of students in 2013 and 2014. Multidimensional IRT (*MIRT)*information indices were employed to analyse research questions while Chi-square difference was used to test for research hypothesis formulated.

Results

Research Question 1: Which of the IRT model fit 2013 WASSCE Economics multiplechoice items?

Table 1:

Model-data fit assessment of 2013 WASSCE Economics multiple-choice items

Factor	AIC	SABIC	BIC	-logLikelihood	X2	df	P
Comparing M1PL and M2PL models							
M1PL	64546.36	64917.17	65555.6	32072.18	461.997	43	0.000
M2PL	64170.37	64620.5	65395.51	31841.18			
Comparing M2PL and M3PL models							
M2PL	64170.37	64620.5	65395.51	31841.18	1795.158	50	0.000
M3PL	62475.21	63017.58	63951.41	30943.6			
Comparing M3PL and M4PL models							
M3PL	62475.21	63017.58	63951.41	30943.6	715.198	50	0.000
M4PL	61860.01	62494.63	63587.26	30586.01			

Table 1 indicates that when the fitness of M1PL and M2PL models to the Economics data set were compared, the result showed that the M2PL had AIC = 64170.37, SABIC = 64620.5, BIC = 65395.51 values that were lesser than the AIC = 64546.36, SABIC = 64917.17, BIC = 65555.6 values of the M1PL. In addition, the Likelihood ratio test that M2PL fitted the data better than M1PL with statistically significant (χ^2 (43) = 461.997, p < 0.05). These results showed that the M2PL model fitted the data better than the M1PL model. In search for a better model for the test data, the fitness of M2PL model to the Economic data was in turn compared to the fitness of M3PL model to the test data dimension model. The result showed that the M3PL model fitted the data better than the M2PL model (M3PL model's AIC = 62475.21, SABIC = 63017.58, BIC = 63951.41 values were respectively lesser than the M2PL model's AIC = 64170.37, SABIC = 64620.5, BIC = 65395.51; the Likelihood ratio test that M3PL model fitted the data better than M2PL model was statistically significant, (χ^2 (50) = 1795.158, p > 0.05)).

In the same vein, the model-data fit for M3PL model and M4PL model were compared and the results showed that M4PL model fitted the data better than the M3PL model (M4PL model's AIC = 61860.01, SABIC = 62494.63, BIC = 63587.26 values were lesser

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

than the M3PL model's AIC = 62475.21, SABIC = 63017.58, BIC = 63951.41 values respectively; Likelihood ratio test that M4PL model fitted the data better than M3PL was statistically significant, χ^2 (50) = 715.198, p < 0.05). These results showed that M4PL model clearly fitted the data better than M1PL, M2PL and M3PL models. These results showed that the Multidimensional 4-parameter logistic model fitted the 2013 WASSCE Economics test with 4 dimensions.

Research Question 2: Which of the IRT Model fitted the 2014 WASSCE Economics multiple choice items?

Table 2:
Model-data fit assessment of 2014 WASSCE Economics Multiple-choice items

Model	AIC	SABIC	BIC	-2logLik	X2	df	P
Comparing M1PL and M2PL model-data fitness							
M1PL	62514.13	62700.45	63021.26	-31156.06	3029.01	48	0.000
M2PL	59581.12	59855.99	60329.26	-29641.56			
Comparing M2PL and M3PL model-data fitness							
M2PL	59581.12	59855.99	60329.26	-29641.56	1467.621	50	0.000
M3PL	58213.5	58580.61	59212.69	-28907.75			
Comparing M3PL and M4PL model-data fitness							
M3PL	58213.5	58580.61	59212.69	-28907.75	488.245	50	0.000
M4PL	57825.25	58284.61	59075.5	-28663.62			

Table 2 indicates that when the fitness of M1PL and M2PL models to the Economics data set were compared, the result showed that the M2PL had AIC = 59581.12, SABIC = 59855.99 and BIC = 63021.26 values that were lesser than the AIC = 62514.13, SABIC = 62700.45 and BIC = 60329.26 values of the M1PL model. In addition, the Likelihood ratio test that M2PL fitted the data better than M1PL was statistically significant (χ^2 (48) = 3029.01, p < 0.05). These results showed that the M2PL model fitted the data better than the M1PL model. In search for a better model for the test data, the fitness of M2PL model to the Economic data was in turn compared to the fitness of M3PL model to the test data dimension model. The result showed that the M3PL model fitted the data better than the M2PL model (M3PL model's AIC = 58213.5, SABIC = 58580.61, BIC = 59212.69 values were respectively lesser than the M2PL model's AIC = 59581.12, SABIC = 59855.99, BIC = 60329.26; the Likelihood ratio test that M3PL model fitted the data better than M2PL model was statistically significant, (χ^2 (50) = 1467.621, p > 0.05)).
In the same vein, the model-data fit for M3PL model and M4PL model were compared and the results showed that M4PL model fitted the data better than the M3PL model (M4PL model's AIC = 57825.25, SABIC = 58284.61, BIC = 59075.5 values were lesser than the M3PL model's AIC = 58213.5, SABIC = 58580.61, BIC = 59212.69 values respectively; Likelihood ratio test that M4PL model fitted the data better than M3PL was statistically significant, χ^2 (50) = 488.245, p < 0.05).These results showed that M3PL model clearly fitted the data better than M1PL and M2PL models. More so, results of this study showed that the Multidimensional 4-parameter logistic model fitted the 2014 WASSCE Economics test. Hence, the probability of correct response P($\boldsymbol{\theta}$) was estimated using the M3PL and the result showed that all the items were locally independent.

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

Research question 3: What is the relative efficiency of 2013 and 2014 WASSCE Objective Economics items?

Table 3:
Relative Efficiency of 2013 and 2014 WASSCE Economics Test Items

ID	RE	ID	RE	ID	RE	ID	RE	ID	RE	ID	RE
1	490.8	2	6.1	3	0.4	4	38.1	5	0.6	6	10.0
9	2.0	10	0.6	11	2.4	12	1.7	13	1.2	14	0.6
17	0.0	18	0.9	19	4.7	20	0.1	21	41.7	22	14.3
25	8.1	26	19.9	27	138.3	28	12.2	29	1.3	30	26.9
33	81.2	34	7.4	35	0.0	36	1.4	37	0.4	38	1.2
41	0.2	42	8.7	43	0.0	44	0.0	45	0.1	46	0.0
49	1597.6	50	1.1	51	1.1	52	0.1	53	0.4	54	2.1
57	9.8	58	0.0	59	19.2	60	5.4	61	2.3	62	0.3
65	0.4	66	0.3	67	1.3	68	6.3	69	2.0	70	3.6
+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+
1097	37.9	1098	30.5	1099	31.5	1100	9.2	1101	443.7	1102	1.2
1105	67.9	1106	81.6	1107	0.5	1108	0.4	1109	0.0	1110	0.5
1113	0.0	1114	0.0	1115	0.1	1116	541.8	1117	614.5	1118	13.6

The table 3 shows the abridged relative efficiency of 2013 and 2014 WASSCE Economics test of all the 1,120 examinees in Economics Achievement Test (EAT) 2013 and 2014 that participated in the study. For examinee 1, there was 80% difference between 2013 and 2014 WASSCE Economics items. 2014 WASSCE Economics items were 80% more informative than 2013 for examinee 1. So, no matter the level of ability of the examinee 1, there was 80% difference between 2013 and 2014 WASSCE Economics items. For examinee 1103, 1105 there was 90% difference between 2013 and 2014 WASSCE Economics and that 2013 WASSCE Economics was 90% more informative than that of 2014 WASSCE Economics items. But, examinees 6, 9,17,35 among others have 0% difference between 2013 and 2014 WASSCE Economics. This implies that there was no difference in 2013 and 2014 WASSCE Economics items for examinee 6,9,17,35 among others. The implication of the study was that both 2013 and 2014 WASSCE Economics test items were less informative. But, 2013 WASSCE Economics test items provide relatively more information than the 2014 WASSCE Economics test items.

Research Hypothesis

Based on the research question raised, the following research hypotheses are formulated to guide the study.

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

Ho_1: **There is no significant difference between the 2013 and 2014 WASSCE Economics multiple-choice test.**

Table 4: Independent t-test Comparison of the Estimated SE of 2013 and 2014 WASSCE Economics test

YEAR	N	X	Std.Error Mean	T	Df	Sig.(2-tailed)
2013	1120	0.7278	0.01627	7.461	2235.353	0
2014	1120	0.5589	0.01572			

Table 4 presents the independent t-test of SE of 2013 and 2014 WASSCE Economics test. The table shows that the precision with which 2013 WASSCE Economics test measures Economics proficiency of examinees was better (Mean = 0.7, SE Mean = 0.02) than the precision with which 2014 WASSCE Economics test measures students' proficiency in Economics (mean= 0.5, SE mean 0.2) in North central geo-political zone of Nigeria. Independent t-test statistic shows that the difference observed in the tests' precision was statistically significant (t (2235.353) = 7.461, p-value < 0.05). The calculated significant level is 0.0 < 0.05; therefore, the null hypothesis was rejected. Thus, there is significant difference between the test information function of 2013 and 2014 WASSCE Economics multiple-choice test items. The implication of this study is that 2013 and 2014 WASSCE Economics multiple choice test items were statistically different in the North- central geopolitical zone of Nigeria.

Discussion

The results indicated that 4PL model fit the data better than M1PL, M2PL and M3PL models for both the 2023 and 2024 items. This result is in agreement with the study of Osgood, McMorris & Potenza (2002) reported that a 4PL model better fit to the data than the others. Loken & Rulison (2010) reported that the overall fitness is improved when using the 4Pl than other parameters. Loken & Rulison (2010) added that 4PM allows each item in upper asymptote error occurred due careless mistake of even a very high ability respondent Hoan (2021) affirmed that four-parameter model revealed that high ability respondents select the wrong option due to carelessness or fatigue.

However, the study was contrary to the study of Alagoz (2005), Moghadamzadah, Salehi & Khodaie (2011), Zieba (2013) and Ibrahim (2017). Alagoz (2005) reported information function on two parameters logistic model was higher than that of one parameter and the three parameters. Moghadamzadah, Salehi and Khodaie (2011) and Zieba (2013) reported that information function on two parameters model was higher than one parameter and the three parameters models. But there is no much difference in the 2PL and 1PL models.

These findings were concord with the study of Alaka and Obadara (2013), Adewuni, (2016), and Abdulsalam-Nuhu, (2018). Alaka and Obadara (2013) reported that there is poor students' academic performance in all geo-political zones of Nigeria. Adewuni (2016) also reported that the performance of students in NECO Senior School Certificate Examination objective test in Government for 2013 and 2014 were low. More so, Abdulsalam-Nuhu (2018) reported that the performance of examinees in practical Geography achievement test was on average.

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

Conclusion

The study concluded that the model fit for 2013 and 2014 WASSCE Economics items was M4PL model with the same multidimensional traits. However, the 2013 test items provided relatively higher information than that of 2014.The Economics items used for the assessment possessed the same dimensional traits. Thus, the test scores or grades could be used interchangeably for admission.

Recommendations

It was recommended that Examinations bodies like WAEC, NECO, NABTEB, JAMB and others could implement the re-use of the 2013 and 2014 WASSCE Economics items for testing. Seminar, training and conference on the importance of model fit statistics of items should be organized for the teachers and the lecturers in order to improve the quality of testing in schools. Future researchers may combine both the dichotomous and polytomous scoring items among examining bodies another in various disciplines at another years and may cover the entire country.

References

Abdulsalam-Nuhu,R. O. (2018). *Invariance of Item Parameters of Polytomous Geography Items in Senior School Certificate Examination in Nigeria*. A PHD Thesis submitted to University of Ilorin, Ilorin, Nigeria

Adewuni, D.A. (2016). *Hierarchical cluster analysis of the dimensionality of Nigerian senior school certificate examination objective tests in government* (2013 -2014). An unpublished Ph.D. thesis; University of Ilorin, Ilorin.

Adu, E.O and Ayeni, A.O (2004): *An appraisal of trend in performance of students' in Economics at the SSCE in Ibadan North L.G.A of Oyo state* (1994-1998). InElatunroti, D.F and Babarinde.K (Eds). Teachers' Mandate on Education and Social Development in Nigeria. Ibadan Stirling Hoden Publishes (Nig) Ltd. pp.129-141.

Alagoz, C. (2005). *Scoring Tests with Dichotomous and Polytomous Items*. Unpublished(M. ed)
Thesis Submitted to the Graduate Faculty of the University of Georgia.

Alaka, A.A & Obadara O.E (2013). *Scholastic performance of students at West African Senior Secondary Certificate Examination in Nigeria*. Journal of educational and social research 3(1) 275-285.

Baker, F., B. & Kim, S. H. (2010). Item response theory: Parameter estimation techniques.
New York, NY: Marcel Dekker.

Check, J. & Schutt, R.K. (2010). Research methods in Education. Sage Publication Inc. retrieved
from http:// www.ssagepub.com/upm-data/43589 pdf.

De Ayala, R. J (2009). *The theory and practice of item response theory*. New York: The Guilford Press.

Doran, N. J., Moses, T., & Eignor, D. (2010). *Principles and practices of test score equating* (ETS Research Report No. RR-10-29). Princeton, NJ: ETS.

Embretson, S. E., &Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum publishers.

Essen, Idaka&Metibemu (2017). *Item level diagnostics and model - data fit in item response*

http://www.lijassed.org
Print ISSN: 2992-4979
Online ISSN: 2992-4987

theory (IRT) using Bilog - mg v3.0 and Irtpro v3.0 programmes: Global Journal of Educational Research Vol 16, 2017: 87-94

Federal Republic of Nigeria (2014). *National Policy on Education*. Abuja: Nigeria Educational
Research Council NERC Press.

Ibrahim, A. (2017). An empirical analysis of one, two and three parametric logistic models of item response theory and differential item functioning in dichotomous and ordinal tests. *Journal of Educational Foundations* (JEF), ISSN: 1119-9210, 7(1), 123-136; 2017. Published by Faculty of Education, University of Jos, Jos, Plateau State, Nigeria.

Ibrahim A. K and Nukhet C. D. (2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. *Procedia- Social and Behavioral Sciences* 46 p.135 – 140

Lam, W.W (2012). *An Empirical Examination of the Impact of Item Parameters on IRT information functions in mixed format tests*.A Ph. D Dissertation Submitted to Graduate School of the University of Massachusetts. Amherst: Canada.

Moghadamzadeh, A. Salehi, K. & Khodaie, E. (2011). *A comparison of the Information functions of the Item and Test in one, two and three parametric model of the Item Response Theory (IRT)*. International Conference on Education and Educational Psychology: Published by *Elsevier limited. Social and BehaviouralSciences 29 (2011) 851359-1367*

Reckase, M.D. (2003). The Design of Test Information Functions to Meet Statistical Test Specifications. Paper presented at the Bilateral Workshop on Psychometrics Flanders–Chile, Leuven, Belgium, July12–13, 2003.

Sulyman R. (2021). *Analysis of items and test information function of Senior School Certificate Examination in Economics in North-central, Nigeria*. An unpublished Ph.D. thesis; University of Ilorin, Ilorin.

Thissen, D. (2000). *Reliability and measurement precision*. In H. Wainer, (Ed) computer adaptive testing: A primer (2nd ed. pp. 159-184), Westport, CT: Praeger Publishers. Foley, Brett P., "Improving IRT parameter estimates with small sample size: Evaluating the efficacy of a new data augmentation technique" (2010). Open Access theses and Dissertations from the College of Education and Human Sciences. Paper http://digitalcommons.unl.edu/cehsdiss75.

Thissen, D. J., & Steinberg, L. (2010). Using item response theory to disentangle construct at different level of generality. In S. Embretson (Ed.) measuring psychology construct: Advance in model-based approaches (pp.123-144) Washington, D.C: American Psychological Association.

Thissen, H. & Wainer, H. (2001). Overview of Test Scoring. In D. Thissen & H. Wainer (Eds), Test Scoring(pp.23-72) Hilldale, N.J: Lawrence Erlbaum Associates.

Thissen, D., & Wainer, H. (2011). Test Scoring. Mahwah, NJ: Lawrence Erlbaum.