ASSEREN Journal of Education Vol. 9 No. 1, July, 2023 1-12

APPLICATION OF ITEM RESPONSE THEORY ANALYSIS ON NIGERIA AIR-FORCE SECONDARY SCHOOLS' JOINT PROMOTION EXAMINATION IN ECONOMICS

EZEKIEL OLUBAYO OJO, Ph.D.
&
YUSUF A. MUSTAPHA, Ph. D.
RESEARCH, MEASUREMENTS AND EVALUATION UNIT
EDUCATIONAL FOUNDATIONS DEPARTMENT, FACULTY OF EDUCATION,
UNIVERSITY OF JOS

Abstract
The study examined the application of item response theory analysis on Nigeria Air-force Secondary Schools' Joint Promotion Examination in economics. The study raised three research questions and formulated two hypotheses were tested at 0.05 level of significance to guide the study. The study employed descriptive survey research design. Survey design is a descriptive study which seeks to describe and document what exists, with the purpose of generalizing the results on the whole population. The population of the study was 1,035 SS II students' scripts for 2018 economics Joint Promotion Examination (JPE) from eleven (11) Airforce Secondary Schools in Nigeria. The study used purposive sampling techniques to sample two Airforce Secondary Schools from Lagos and Kano, sample size of 200 Students. The two study instruments were Joint Promotion Examination (JPE) question papers and Answers Scripts of Students for the year 2018 academic session. Content validity of instrument with Kendall's coefficient of concordance value of 0.75 and had Internal consistency reliability coefficient of the EAT was established with Kuder-Richardson formula (K-R20) was 0.83. The collected data were analyzed with x-caliber 4 software for item analysis. The results shows that 45(90%) of the 50 items fitted the three- parameter logistic model of the item response theory. The study recommends that organization should dropped question items 38 and 47 due to their biased nature, since the presence of DIF affects the validity of an item.

Key words: Difficulty, Dimensionality, Discrimination, Guessing Parameters, Information Functioning, Item Response Theory, Joint Promotion Examination.

Introduction

Education is the process of transmitting societal values and desirable attitude from one generation to another. It is an indispensable instrument for human progress, empowerment and development. For this reason, the Federal government of Nigeria adopted education as an instrument per excellence for effecting national development (Federal Republic of Nigeria, FRN, 2014). This implies that a nation that lacks sound educational culture and philosophy stands the risk of decay, whereas the one having it is bound to succeed. It is in the light of the foregoing that Nigeria developed and structured its educational system into Pre-Primary, Primary; Secondary and Tertiary education. Pre-primary education is the

education given to children between ages two to five years prior to their entering the primary school. Primary education is given to children from ages six to twelve years. These serve as the foundation for higher education (FRN, 2014).

Secondary education on the other hand, is given to students after completing primary education. Students spend six years at this level, made-up of three years in junior and three years in senior secondary school. At the Junior Secondary level, the subjects are divided into groups A, B and C such that the students are expected to offer a minimum of ten (10) and a maximum of thirteen (13) subjects. All the subjects in group A are compulsory and at least one subject each from groups B and C as contained in the Nation Policy on Education (Federal Republic of Nigeria, FRN, 2014). Group A comprises core subjects, group B consists of the pre-vocational elective courses and group C, the vocational elective courses. Economics is one of the elective subjects offered at the senior secondary school level as prescribed by the National Policy on Education (FRN, 2014). Economics is a social science subject which studies human behaviour as a relationship between ends and scarce means which have alternatives uses (Anyanwuoch, 2011).

Economics activity arises out of the conflict between the countless wants of human beings and the scarce means for satisfying them afforded by nature (Ewa & Agu, 2005). The problem of humanity has always been that its wants are too many while the means of satisfying them have been too few or limited. The economic life, with its institutions of business firms, employers, workers, marketers, international trade and the like, it is basically, what people have been doing and will do, to overcome that conflict. Recognizing this fact and doing something about it fall within the purview of economics.

The importance of economics as a subject cannot be overemphasized since its knowledge is desired by all citizens, either as producers or as consumers to enable them acquire the understanding of how economic systems work, thereby equipping individuals with solutions to the numerous socio-economic problems of life (Patrick & Audu, 2014). It also enables students to understand basic economic principles and concepts as well as the tools for sound economic analysis, contribute intelligently to discourse on economic reforms and development as it affects or would affect the generality of Nigerians to understand the structure and functioning.

Despite the laudable importance of economics to the students, society and the nation at large, it has been observed that students' performance in the subject in Nigeria is not encouraging (WAEC Chief Examiners' Report 2014, 2015, 2016 & 2017). There is a noticeable decline in students' performance in economics in Senior Secondary School Certificate Examinations (S.S.S.C.E). Also, performance of the Nigerian Air Force secondary school students in economics is poor. The report from the Directorate of Education (DOEDN) (2015 & 2016) show that the students performed poorly in the subject in their WAEC and NECO results. In 2015 and 2016 WAEC results, 40% of students passed at credit level while 60% failed. Also in 2016, only 45% passed at credit level while 55% failed the subject. It is a culture in the Air Force Secondary Schools that students in SSS II must pass a Joint Promotion Examination (JPE) before they can be promoted to Senior Secondary Class three (SS III). This means that any students who fail,

will not be promoted to SSS III, thereby depriving them the opportunity to register for the SSCE examinations.

The Nigerian Air Force Examination Committee (NAFEC) under the Directorate of Education (DOEDN) is responsible for setting and administering the Joint Promotional Examination (JPE) test questions for the SSS II students in their third term as a condition for promotion to SSS III class. This committee however, does not do analysis of the examination, and in many instances, the questions are repeated year-in year-out. To this end, the need arises to query the extent to which the items of the examination possess the required psychometric properties. It is therefore of utmost importance to carry out checks on the quality of the economics test that is given to the students. This is in a bid to see if this is one of the reasons why students do not do well in the Joint Promotion Examination but perform poorly in WAEC examinations. Some of the qualities desired of good test are validity, reliability, objectivity, good level of difficulty and discrimination and effectiveness of distracters in case of objective test items.

Airforce Secondary Schools are either single-sexed or mixed schools. In this school both the male and female students offer the subject, economics. Interestingly, it has been found that gender is a factor that affect students' performance in school subject. For example, female students have been found to perform better academically in the past in quantitative subjects like mathematics and economics (Tuntiwarodom & Potipiti, 2008); while their male counterparts were found to do better in quantitative subjects (Zulkfli, 2013). This controversy calls for the need to investigate whether the economics mock questions are affected students' gender.

The procedure for analysis of test can emanate from three available theories namely: Classical Test Theory (CTT), Item Response Theory (IRT) and Generalizability Theory (GT). Each of these theories has its strengths and accompanying weakness. The traditional approach to analyzing tests which embodies two concepts; item difficulty and item discrimination is referred to as the Classical Test Theory (CTT). The approach was first used in the twentieth century by Alfred Binet and Theophile Simon (Diu, Kunmi, Francis & Patrick, 2012). Classical Test Theory views the observed score (x) as a combination of the true score (T) and an error component (S) (Adedoyin, 2010). It is the observed score of a test-taker plus or minus some unobservable measurement error (Crocker & Algina, 2008). CTT is relatively simple and easy to interpret because it does not have complex theoretical model to relate an examinee's ability to succeed on a particular item. Instead, CTT collectively considers a pool of examinees and the examinee's ability to succeed on a particular item. The disadvantage of this measurement theory is that, the item difficulty could vary depending on the sample of test-takers. Also, Npkone (2001) posits that in CCT, the proportion of examinees in a sample that get an item correct change from a sample whose mean ability is high to one whose mean ability is low.

A new measurement theory which can overcome the disadvantages of CTT is the Item Response Theory (IRT). The IRT is a set of mathematical models designed to describe the fundamental relationship between examinees ability and performance on an item (Adedoyin & Mokobi, 2013). IRT model assumes that the performance of an examinee

can be completely predicted or explained from one or more abilities, it models the probability of a correct answer using three logistic functions. The one-parameter logistic (IPL) model attempts to address the probability of a correct answer by allowing each question to have an independent difficulty variable. The two-parameter logistic (2PL) model attempts to model each item's parameter which is called pseudo-guessing parameter that reflects the probability that an examinee will answer an item solely by guessing (Obinne, 2012).

Furthermore, test information function is another feature in the application of IRT in test development process. It is the estimates of the error associated with maximum likelihood ability estimation of testees. It means that more information is provided by a test a particular ability level. The smaller the error associated with the test, the higher the ability estimate. This will help in the estimation of the test (Baker, 2013). Item information function shows the contribution of a particular item to the assessment of ability. Item with high discriminating power (Ojerinde, Popoola, Ojo & Onyenho, 2013).

Item fit statistic is empirical evidence supporting the claim that a specific model is an adequate model for a data. It is index of fit of an item and a model. It is the degree of fit or misfit for each individual test item in a test in relation to unidimensionality, local independence, monotonicity, and no differential item functioning (Liu, 2014). It gives a test developer the condition under which an item can be retained or discarded during test item calibration. The implication of this to the present study is that for a model to be applied, in calibrating the test items, the item fir statistic will be determine to find out if the model suit the data.

Differential Item Functioning (DIF) is carried out in test development process to determine if examinees of equal ability but from different background have an unequal probability of answering an item correctly (Obinne & Amali, 2013). DIF is a condition when item in a test function differently for respondent with the same level of ability but from different background. For instance, one could investigate whether a test function differently for student in public and private secondary schools because findings from literature are inconsistent about the effect of school type based on DIF analysis in JSCE mathematics (Makobi & Adedoyin, 2014).

The IRT is a modern measurement theory that is used in developing achievement test. It is a family of statistical procedure for analyzing and describing test performance. It is specially designed to model the interaction between a testee and a item and that the testee must have a trait level sufficient enough to be able to answer the item correctly. The focus of IRT is on the pattern of responses rather than the total score of the students in a test as opposed to CTT, hence item difficulty (location), item discrimination and pseudo guessing are important component of test development using IRT (Chalmers, 2013). Also important are item calibration, item information function, test information function and differential item function.

In IRT, the item difficulty or location parameter (b) is the amount of latent trait a testee must possess to be able to answer an item in a test correctly. It is the probability that a testee

will be able to answer an item correctly if the student has the ability. Item with high values of (b) are difficult item in hard item with low ability examine having low probability of correctly responding to the item. On the other hand, item with low values of (b) are easy items whereby most examinees, including low-ability group having at least a moderate probability of answering the item correctly. Similarly, item discrimination parameter (a) in test development using item-responses theory indicate how well, test item differentiates between individuals of different latent trait level (Thata) and also indicates the different between high and low achievers in a test. Discriminating assumes that indicial with low ability have a much smaller chance of correctly responding to an item than person of higher ability (Zhang, 2010). Discrimination is important because item response theory requires that each person should be differentiated based on their latent trait and not the group to which they belong.

Also, IRT is anchored on three assumptions namely; unidimensionality, local independence and item characteristic curve. The unidimensionality assumes that an item measures one and only one area of content of knowledge or ability. Local independence assumes that the examinees' responses to questions are not statistically related to each other, while an item characteristic curve depicts the true relationship between the latent trait (ability) of an examiner and the responses to them (Ojerinde, 2013). Also, there is the need to establish the differential item functioning of the examination questions. Differential item functioning (DIF) is a technique under Item Test Theory (IRT) to identify examinees of equal ability but from different background that have unequal probability of answering an item correctly (Martiniello, 2009). Differential item functioning of the student's performance in economics JPE will be determined based on gender and school location.

IRT differs from CTT in that IRT focuses on performance of individual items rather than the whole test alone. IRT describes item performance at each level of student ability and it is model based. In view of the above, therefore construction and standardization of test items for JPE examination may require special attention if the examination is to be made valid and reliable. There is therefore the need to assess the psychometric properties of JPE economics examination using IRT 3-parameter model. This therefore formed the underlining basis for this study.

The Nigeria Air Force schools conduct a Joint Promotion Examination (JPE) as a centralized type of examination for senior secondary school students II to qualify them to get prompted to SSS III. The examination is conducted by the Nigeria Air Force Examination Committee (NAFEC) under the Directorate of Education (DOEDN) Abuja. The result from NAFEC in economics for the past three years, 2014-2016 show that the performance of students that passed at credit level and above in JPE were 80% and above, but the performance of students in SSCE WAEC and NECO in Air Force Secondary Schools was below average. For instance, in 2014, 2015 and 2016; 1210, 1200 and 1100 students respectively sat for the May/June examination in economics. Out of these numbers, in 2014-2016, 605 representing 50%, 540, representing 45% and 440 representing 40%, respectively could not score a credit pass in the subject.

Reports from DOEDN shows that the performance of SSS III students in economics in WAEC and NECO was poor compared with their JPE results especially that of Air Force Military School, Jos (WAEC result 2014 and 2015). Due to the yearly percentage increase in students' failure in SSCE economics results in 2014 and 2015, in Nigerian Air Force secondary schools, the NAF Headquarters directed her schools to take some remedial actions in the academics that will include all subjects where students had deficiencies; economics inclusive. The JPE was a conscious effort made by NAF Directorate of education in tackling the problems of poor performance in economics in the SSCE. This was to explore alternative and effective strategies for improving students' performance in economics.

The study was aimed at establishing the psychometric properties of the 2018 Air Force Secondary Schools Joint Promotion Examination in Nigeria using the IRT. Specifically, other objectives of the study include, to:

1. Establish content validity of the 2018 JPE Air Force Secondary School Economic Examination Questions
2. Determine discrimination, difficulty and guessing parameters.
3. Find out the test information functioning and the dimensionality of the test.

RESEARCH QUESTIONS

The following research questions guided the study:

1. How many of the 2018 JPE Airforce Secondary School Economics multiple-choice items fit the 3-PL model of item response theory.
2. What is the reliability of 2018 JPE Airforce Secondary School Economics Examination questions
3. What are the discrimination parameters of the 2018 Economics JPE for the Airforce Secondary Schools?

HYPOTHESES

The following hypotheses were tested at 0.05 level of significance:

1. There is no significant difference between the fit of 3-parameter logistic model and the items response data for 2018 Economics JPE multiple-choice items.
2. The items of the 2018 NAF JPE for Economics do not function differentially due to gender of students.

Methodology

The study employed descriptive survey research design which seeks to describe and document what exists, with the purpose of generalizing the results on the whole population. The population of the study was 1,035 SS II students' scripts for 2018 economics Joint Promotion Examination (JPE) from eleven (11) Airforce Secondary Schools in Nigeria. The study used purposive sampling techniques to sample two Airforce Secondary Schools from Lagos and Kano, sample size of 200 Students. The two study instruments were Joint Promotion Examination (JPE) question papers and Answers Scripts of Students for the year 2018 academic session.

The instrument consists of two Sections A and B. Section A consist of the data of the

students such as students' gender and name of school, school type. Section B consist of items that measure the students' achievement in economics which were of two parts: Parts 1 constitutes 50 four-option multiple-choice objective test items and Part 2 constitutes 2 essay items. Content validity of instrument were obtained from four experts judgments from Economics Education Unit, Research Measurement and Evaluation Units from Faculty of Education, University of Jos with Kendall's coefficient of concordance value of 0.75 based on experts scrutinized the instruments relevance, clarity, simplicity and ambiguity of the items. Internal consistency reliability coefficient of the EAT was established with Kuder-Richardson formula (K-R20) was 0.83. The descriptive statistics of mean and standard deviation were used to answer research questions 1 to 5 while inferential statistics was used to test hypotheses 1 to 3 the data were analyzed with x-caliber 4 software for item analysis.

RESULTS AND DISCUSSION

Research Question One

How many of the 2018 JPE Airforce Secondary School Economics multiple-choice items fit the 3-PL model of item response theory?

Table 1: Item Fit Statistics for Multiple-choice

Item ID	x^2	p	Item ID	x^2	p
1	9.96	0.87	26	23.18	0.62
2	22.55	0.54	27	9.33	0.84
3	28.55	0.24	28	11.81	**0.03**
4	29.70	**0.03**	29	16.21	0.08
5	15.08	0.24	30	20.51	0.53
6	18.16	0.21	31	6.09	0.89
7	28.46	0.09	32	18.21	0.21
8	17.17	0.64	33	13.08	0.41
9	78.03	0.53	34	11.92	0.53
10	12.68	0.07	35	3.11	0.97
11	33.2	0.08	36	2.74	0.08
12	18.3	0.10	37	18.32	0.20
13	10.4	0.71	38	9.48	**0.01**
14	18.11	0.09	39	7.21	0.89
15	16.2	**0.04**	40	8.12	0.44
16	17.0	0.51	41	4.46	0.81
17	11.02	0.06	42	8.21	0.92
18	13.8	0.08	43	6.91	0.98
19	12.01	0.07	44	21.01	0.07
20	10.21	0.09	45	16.12	0.08
21	9.06	**0.01**	46	11.05	0.60
22	28.01	0.42	47	9.38	0.31
		0.03	48	16.41	0.24
24	5.34	0.89	49	21.04	**0.02**
25	8.04	0.91	50	30.08	0.74

Table 1 revealed that items 4, 23, 28, 38 and 49 did not fit the 3-parameter logistic model of item response theory since the p-values are less than 0.05. It means that 45(90%) of the 50 items fitted the 3-parameter logistic model of the item response theory. The three-parameter model fit the data with a Chi-square value of 860.945, degree of freedom (df) of 588, a probability of 0.06 and -2 logistic likelihood of 2557. Since the P value is greater than 0.05, it implies that the data fits the 3 parameter IRT model.

Research Question Two

What is the reliability of 2018 JPE Airforce Secondary School Economics Examination questions

Table 2: Reliability Statistics of 2018 JPE Air Force Economics examination

Kuder Richardson	Number of Items
0.993	50

Secondly, the instrument had Kuder-Richardson (k-20) reliability coefficient of 0.99 which is high enough for the study. The finding is in line with Barko (2011) investigated of the psychometric properties of junior school certificate in business studies examination in Plateau State in 2011 who finding had reliability coefficient of 0.74 which was moderate (see Table 2).

Research Question Three

What are the discrimination parameters of the 2018 Economics JPE for the Airforce Secondary Schools?

Table 3: Discrimination parameters for Economics multiple choice items

Item ID	"a"	Item ID	"a"
1	0.702	26	0.596
2	0.636	27	0.715
3	0.639	28	0.821
4	0.562	29	0.818
5	0.501	30	0.781
6	0.687	31	0.741
7	0.728	32	0.660
8	0.585	33	0.639
9	Not calibrated	34	0.669
10	0.664	35	0.778
11	0.786	36	0.832
12	0.716	37	0.757
13	0.583	38	0.770
14	0.891	39	0.874
15	0.856	40	0.927
16	0.814	41	0.879
17	0.837	42	0.716
18	0.739	43	0.810
19	0.703	44	0.400
20	0.677	45	0.760
21	0.576	46	0.745
22	0.817	47	0.753
23	0.821	48	0.747
24	0.609	49	0.768
25	0.573	50	0.772

Table 3 shows items that were calibrated are discriminating well, since none of the items has less than 0.50 thresholds. A large value indicates that the item is more strongly differentiating examinees which are typically considered better and indicate that the item differentiates examinees well according to their ability. The finding is in agreement with Adewoye (2015) findings that determine the reliability coefficient, the difficulty, the discrimination, the distracter indices at Education Resource Centre in Federal Capital Territory, Abuja. The average discrimination index of the test was 0.21 respectively. From the discrimination index, it was found that three items were very good, five items are reasonably good by possibly subject to improvement, 5 are marginal items needing improvement and the remaining 11 are poor items that needed to be rejected or improved by a revision.

Testing of Hypotheses

Hypothesis One

There is no significant difference between the fit of 3-parameter logistic model and the items response data of 2018 Economics JPE multiple choice items?

Table 4: Summary Statistics of Model Fit of 2018 JPE Airforce Secondary School Economics JPE Multiple Choice Items

Test	Items	Chi square	df	*P*	-2LL
Full Test	49	860.945	588	0.060	2557

The model fit was determined using three parameter model of item - response theory. The result on Table 4 revealed that the three-parameter model fit the data with a Chi-square value of 860.945, degree of freedom (DF) of 588, a probability of 0.06 and -2 logistic likelihood of 2557. Since the P value is greater than 0.05, it implies that the data fits the 3 parameter IRT model.

Hypothesis Two

There are no significant differential items functioning of the JPE in economics due to gender

Table 5: Differential Item Functioning of Multiple-Choice Items due to Gender

ITEM ID	WALD TEST	P	BIAS AGAINST	ITEM ID	WALD TEST	P	BIAS AGAINST
1	1.2273	0.9214		26	0.96	0.9838	
2	1.1304	0.9524		27	0.6333	0.8135	
3	1.3333	0.8911		28	0.1395	0.2804	
4	2.0625	0.7499		29	0.225	0.4276	
5	1.45	0.8616		30	13.1429	0.5259	
6	0.4848	0.7028		31	2.5	0.701	
7	5.6	0.5641		32	4.4444	0.589	
8	1.0417	0.984		33	15.3333	0.5287	
9	-	-		34	1.7222	0.8044	
10	0.8148	0.9174		35	1.8824	0.7777	
11	5.125	0.5746		36	1.1304	0.9524	
12	2.0625	0.7506		37	0.5312	0.7416	
13	3.0833	0.6511		38	8	0.003	FEMALE
14	0.6897	0.849		39	2.2667	0.7267	
15	18.8	0.5328		40	0.96	0.9838	
16	18.8	0.5328		41	2.4138	0.7095	
17	0.125	0.2709		42	1.5789	0.8329	
18	2.7692	0.6769		43	1.2273	0.9214	
19	0.4848	0.7023		44	2.0625	0.7522	
20	1.8824	0.7769		45	1.45	0.8616	
21	0.96	0.9838		46	0.3243	0.5439	
22	0.1667	0.3268		47	0.4848	0.001	FEMALE
23	2.7692	0.6769		48	6	0.5583	
24	1.0417	0.984		49	32	0.5592	
25	1.0417	0.984		50	2.5	0.701	

Fourthly, Table 5 the Differential Item Functioning (DIF) for gender was determined with Wald test statistical analysis technique using Item Response Theory Patient Report outcome (IRTPRO) shows that items 38 and 47 were biased against female with *p* values of 0.003 and 0.001 respectively. All the other 47 calibrated items were not biased in terms of gender. This finding is in agreement with Hungi (2013) differential item functioning (DIF) analyses was carried out on the Mathematics test to determine sex bias and province bias findings shows that two question items had some sex by class interval DIF. Further analyses of the ICCs of these two items revealed that item 17 was slightly biased in favour of girls and item 18 was slightly biased in favour of boys. Since the expected score for persons with equal ability level did not differ across the gender groups, these two items did not show any serious DIF problems.

Conclusion

In conclusion, application of item response theory analysis on Nigeria Air-Force secondary schools' joint promotion examination in economics revealed that the data fits the 3 parameter IRT model and psychometric properties of junior school certificate in business studies examination in Plateau State is moderately reliable. The discrimination index shows that three items were very good, five items are reasonably good by possibly subject to improvement, 5 are marginal items needing improvement and the remaining 11

are poor items that needed to be rejected or improved by a revision. Item Characteristic Curve (ICC) of these two items revealed that item 17 was slightly biased in favour of girls and item 18 was slightly biased in favour of boys. Since the expected score for persons with equal ability level did not differ across the gender groups, these two items did not show any serious Differential Item Functioning (DIF) problems.

Recommendations

Based on the findings of the study the following suggestions were made:

1. Teachers should use the content validity of the JPE Air Force Secondary School Economic Examination Questions for help the students overcome the difficult areas.
2. Government Ministry of Education should use discrimination for other subjects.
3. The organization should drop question items 38 and 47 due to their biased nature, since the presence of DIF affects the validity of an item.

References

Adedoyin, O. O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theory. *International Journal of Educational Science*. Retrieved November 30, 2012, from http://www.uniBotswana./journal/education/science.

Adedoyin, O. O. & Mokobi, T. (2013). Using item response theory in examining the quality of junior certificate mathematics multiple choice examination test items. *International Journal of Asians Social Science*, *3*(4), 992-1011.

Baker, T. (2008). *Computer adaptive testing in higher education; The validity and reliability approach*. Retrieved from: t.1barker@hertz.zc.uk.

Crocker, L. & Algina, J. (2008). Introduction to classical and modern test theory. Fourth worth: Harcourt brace jovanocich.

Makabi, T.O. & Adedoyin, E.F. (2014). Identifying location biased items in the 2010 Botswana Junior Secondary Certificate Examination Mathematics Paper One using the item response characteristic curve. *International Review of Social Science and Humanities, 7*(2), 63-82.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. Educational Assessment, 14(3-4), 160-179, DOI: 10.1080/10627190903422906

Nkpone, H.L. (2001). Application of latent trait models in the development and standardization of phages achievement test for senior secondary students. Unpublished doctoral dissertation university of Nigeria, Nsukka.

Obinne, A. D. E. (2011). A psychometric analysis of two major examinations in Nigeria: standard error of measurement. International Journal of Education Science, 3(2),137-144.

Obinne, A. D. E. (2012). Using IRT in determining test item prone to guessing. *World Journal of Education. 2*(1), 91-95.

Obinne, A.D.E. & Amali, A.O. (2014). Differential item functioning: The implication for educational testing in Nigeria. *International Review of Social Science and Humanities, 7*(1), 52-65.

Ojerinde, D., Popoola, K., Ojo, F. & Onyenechu, P. (2012). Introduction to item response

theory: parameter models, estimation and application. Abuja, Nig. Mavelose Mike Press.

Tuntiwarodom, L. & Potipiti, T. (2008). Determinants of happiness and academic performance of economic students. *Chulalonakom Journal of Economics*, *20*(3), 183-200.

West African Examination Council (WAEC). Chief examiners' Report 2014, 2015, 2016 and 2017)

Zhang, Y. (2015). Multiple ways to detect differential item functioning in SAS. Retrieved from http://www.support.bas.com/papers/2900-2015.

Zulkifli, I. (2013). Happiness and students' performance in quantitative subjects: a preliminary study. Malaysia, Pak Publishing Group. Retrieved from: file:///c:usuers/user/ICEFMO-295-302economics-pdf.