International Journal of Evaluation and Research in Education (IJERE)
Vol.5, No.4, December2016, pp. 261~270
ISSN: 2252-8822

Evaluation of Northwest University, Kano Post-UTME Test Items Using Item Response Theory

Ado Abdu Bichi, Hadiza Hafiz, Samira Abdullahi Bello
Department of Arts and Social Sciences Education, Northwest University, Kano-Nigeria

Article Info

Article history:

Received Aug 21, 2016
Revised Okt 21, 2016
Accepted Nov 27, 2016

Keywords:

Economics
Item analysis
Item response theory
Nigerian universities
Post-UTME

ABSTRACT

High-stakes testing is used for the purposes of providing results that have important consequences.Validity is the cornerstone upon which all measurement systems are built. This study applied the Item Response Theory principles to analyse Northwest University Kano Post-UTME Economics test items. The developed fifty (50) economics test items was administered to a sample of 600 students. The data obtained was analysed using XCALIBRE 4 and SPSS 20v softwares to determine items parameters base on IRT models. Indicate that, the test measure single trait by satisfying the condition of unidimensionality. Similarly, the goodness of fit test revealed that, the two parameter IRT model was more suitable since no misfit item was observed and the test reliability was 0.86. The mean examinee ability was 0.07 (SD =0.94).The mean item difficulty was -0.63(SD=2.54) and mean item discrimination was 0.28 (SD=0.04). 16 (33%) items were identified as "problematic" based on difficulty indices, 35(71%) also failed to meet the set standards on the basis of discrimination parameters. it can be concluded that, using the IRT approach, the NWU Post-UTME items are not stable as far as item difficulty and discrimination indices are concerned. It is recommended that, the Post-UTME items should be made to pass through all process of standardisation and validation; test development and content experts should be involve in developing and validating the test items in order to obtain valid and reliable results which will lead to valid inferences.

Corresponding Author:

Ado Abdu Bichi,
Faculty of Education,
Northwest University, City Campus, Kofar Nassarawa, PMB 3220 Kano Nigeria.
Email: adospecial@gmail.com

1. INTRODUCTION

Assessment of students learning is an indepensable part of educational process, the major aim of assessment is to measure students' achievement in order to make a variety of decisions based on learners' performance example, to know the present level of students' learning and the extent to which they are ready for next learning experiences [1]. The nature and the quality of information gathered from the achievement test can control the educational development efforts and direct the instruction [2]. Test is used for a number of purposes, which include; improving instructional planning, motivating learners to improve their performances, licencing, certification, selection e.t.c. In addition, tests are used in certifying students as having attained specific levels of achievement [1],[3]. Information from the Achievement tests helps to know the extent at which students' progresses beyond the minimum basics, the extent to which students achieved the learning goals of the course, and then whether the learners are ready for the next learning experience [4].

University is regarded as the single and most important industry for the production of high-level manpower in Nigeria. To support these principles the stakeholders in Nigerian university education sector tend to guard jealously the integrity of the university and the quality of graduates produced [5]. However, in recent years the integrity attached to Nigerian universities seems to have faded away. This is evident in the

way and manner in which the stakeholders in the university education sector maintain the constant criticisms of the admission procedures and quality of graduates produced by Nigerian universities today.

Until recently, admission into Nigerian universities is exclusively through the Unified Tertiary Matriculation Examination (UTME) conducted by the Joint Admissions and Matriculation Board (JAMB). However, due to the universities' loss of confidence in the UTME, resulting from lack of correlation between candidates' UTME scores and their performances in university examinations, the universities have introduced the Post-Unified Tertiary Matriculation Examination (Post-UTME).

The controversies surrounding the introduction and sustenance of Post-UTME by Nigerian universities, though well intentioned, is giving a lot of concern to stakeholders, as it makes the process of securing admission into universities cumbersome and expensive. Yet, the need to ensure that competent candidates are admitted into universities cannot be compromised, as witnessed prior to the introduction of Post-UTME.

1.1. Post-Unified Tertiary Matriculation Examination (Post-UTME)

The obvious weaknesses observed in the process of admitting entrants into Nigerian universities through UTME and the incessant decline in the expected roles of universities in the Nigeria, couple with the several calls for an alternative method of admission into Nation's universities. Example, [6] noted that for many years Nigerian universities were able to admit only 7 to 10% of the applicants. This low number of students being offered admission annually resulted in both applicants and their parents being very desperate in their bid to be the few offered admission in these universities. This ushered in different types of examination malpractices in the conduct of UTME.

These and many other concerns by stake holders resulted in the federal government of Nigeria under the then, leadership of President Olusegun Aremu Obasanjo resolve to grant power to universities through the then Education minister Mrs. Chinwe Abaji to conduct screening tests (Post-UTME) for admission into their various undergraduate programmes in 2005. Under this policy it became mandatory for all universities in the country to organise a screen test for prospective candidates after passing their UTME and before offering them a place into their programmes. Moreover, candidates who scored the required cut-off marks in UTME are shortlisted by JAMB and sent to the universities of their choices. Thereafter, the universities would then screen the candidates using oral interviews, aptitude test, or even another examination [7].

Although Post-UTME was introduced into Nigerian university to help ameliorate lots of problems in the system, several problems were identified by stakeholders in the country's university education. Several studies have confirmed that, the contents of UTME and Post-UTME items are not the same and that, the Post-UTME items are difficult than the UTME. Similarly, studies have also confirmed the failure of Post-UTME to predict the future academic success of undergraduate students of Nigerian universities [7]-[12].

Moreover, little attention has been paid, by researchers in universities to the in-depth analysis of the items contained in the Post-UTME. It is not surprising that some of the items being used for taking decisions on students are not good. A comprehensive study of the process in which the Post-UTME items were constructed as well as their psychometric characteristics may suggest ways of improving students' performance in Post-UTME.

1.2. Purpose of The Study

The main Purpose of this study is to evaluate the quality of the Nigerian universities' Post-UTME economics test items. It is organised to meet the following specific research objectives:

i To Examine the model fit of the Northwest University (NWU) Post-UTME Economics test items to IRT models
ii Test for violations of essential IRT basic assumptions of unidimensionality and local independence
iii Identify the distribution of item discrimination values, difficulty, and pseudo-guessing parameters for the NWU Post-UTME Economics test items.

1.3. Research Questions

i To successfully attain the set objectives of this study, the following research questions will be addressed. Thus;
ii Do the NWU Post-UTME Economics test items fit into the IRT models ?
iii Do the NWU Post-UTME Economics test items satisfy the essential IRT basic assumption of unidimensionality and local item independence ?
iv What are the item parameters (discrimination values, difficulty, and pseudo-guessing parameters) of the NWU Post-UTME Economics test items ?

2. ITEM RESPONSE THEORY

According to [13] Item response theory (IRT) is a set of latent variable techniques especially designed to model the interaction between a subject's "ability" and the item level stimuli (difficulty, guessing, etc.). The focus is on the pattern of responses rather than on composite or total score variables and linear regression theory. The IRT framework emphasizes how responses can be thought of in probabilistic terms. In IRT the item responses are considered the outcome (dependent) variables, and the examinee's ability and the items' characteristics are the latent predictor (independent) variables [14].

The characteristics of Item Response Models, as summarised by [15] are, first, an IRT model must specify the relationship between the observed response and underlying unobservable construct. Secondly, the model must provide a way to estimate scores on the ability. Third, the examinee's scores will be the basis for estimation of the underlying construct. Finally, an IRT model assumes that the performance of an examinee can be entirely predicted or explained by one or more abilities.

2.1. Assumptions of Item Response Theory

In using IRT Model, it is very important to assess the extent to which the IRT model assumptions are valid for the given data [16]. The most significant assumptions common to all IRT models is unidimensionality, other assumption relates to IRT is Local independence. The test data can only be valid for latent trait model estimation only if these assumptions are met.

2.1.1. Unidimensionality

Unidimensionality states that the items in a test measure single unidimensional ability or trait and that the items form a unidimensional scale of measurement [17]. The Item response models that assume a single latent ability is referred to as unidimensional. This assumption means that the items measure only one area of ability or knowledge. This assumption is empirically assessed by investigating whether dominant factor exists among all the test items [18].

2.1.2. Item Local Independence

The assumption of local independence means that, the probability of an examinee getting item correctly is not affected by the answer given to other items in the test. For example, if the responses to one item structurally constrain the possible answers to other items, then the items are not locally independent. If these assumptions are met, an IRT model can be successfully employed [19].

2.2. Item Response Theory Models

[20]Summarised the models when he said "IRT models differ depending on whether the relationship between item performance and knowledge is considered a one-, two- or three-parameter logistic function.

2.2.1. The one-parameter logistic model

The 1-parameter model explains the relationship between the ability and probability of a correct response on the item in terms of the item difficulty. An item's difficulty parameter (b) is the point on the ability scale corresponding to the location on the item characteristic curve (ICC) where the probability of a correct response is 0.5 [18].

$$P(\theta) = e^{a(\theta)}/1 - e^{a(\theta)} \quad (1)$$

where:
P (θ) = ability of a student
a(θ) = difficulty level of item
e =2.73= discrimination index
b(θ) = 1 in this model.

2.2.2. The two –parameter logistic model

The 2-Parameter model makes use of the b parameter (item difficulty) just as in the 1PLM, and addition add an element that indicates how wills an item separates students into different ability levels this parameter is called item discrimination (a). The item discrimination (a) parameter used in the 2PLM is equal to the slope of the item characteristics curve when it is at its steepest [18].

$$P(\theta) = \frac{1}{1+ e^{-1.7a(\theta-b)}} \quad (2)$$

2.2.3. The three-parameter logistic model

The 3PL model builds upon the two-parameter model by adding pseudo-chance-level parameter c. The c parameter is the value of the lower asymptote of the item characteristic curve and is indicative of the probability that an examinee with a very low ability score would answer an item correctly.

$$P(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}} \quad (3)$$

2.3. IRT-Item Characteristic Curves (ICC)

An item characteristic curve plots the probability that an examinee will respond correctly to an item solely as a function of the test's latent trait [21]. The values on the X-axis of an ICC represent the latent trait, usually ranging from -3 to +3. The Y-axis represents the probability of an examinee's success. As the latent trait increases, the probability of the examinee responding correctly will increase but with diminishing returns.The probability of a correct response is determined by the item's difficulty and the examinee's ability. This probability can be seen as illustrated using item Characteristic curve (ICC) in Figure 1.

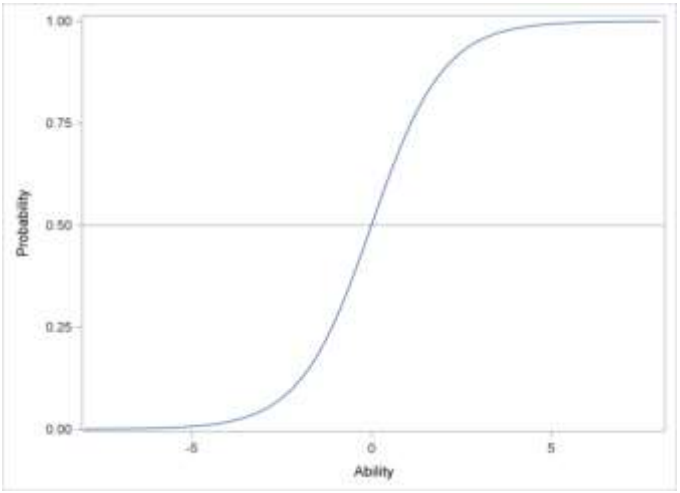

Figure 1. Item Characteristics Curve [22]

From this ICC we can observe that as the examinee's ability increases, the probability of a correct response increases; this is what you would expect in practice**.**The earlier discussion and equation suggests that the probability of endorsing an item correctly or a correct response is 0.5 for any examinee whose ability is equal to the value of the item difficulty.

2.4. Item Response Theory Item Analysis Statistics

Test item analysis involved statistics that help in analysing the effectiveness of the items and improving test items. These statistics can provide useful information to determine the validity and accuracy of an item in describing learners or examinees ability from their response to each of the item in a test. The common Item Response Theory item analysis statistics are Item discrimination (a-values), Item difficulty (b-values), Pseudo Guessing (c-values) and (d) Reliability. This paper will therefore cover the two major statistics of Item difficulty (b-values) and discrimination (a-values). The values of the IRT item parameters will be interpreted base on the XCALIBRE recommended acceptable ranges of item parameters $a \geq 0.30$; $-3.0 \leq b \leq 3.0$ [23].

3. RESEARCH METHOD

3.1. Design of The Study

This study is a quantitative and survey design will be adopted to collect the relevant data for the study. The IRT models guide the study. The IRT is used in order to overcome the limitations of the Classical Test Theory (CTT) [24].

3.2. Participants

The population of this study comprises the entire senior secondary schools three (SS III) students in Nigerian. 550 SS III students, age (16-18) selected using a stratified random sampling technique participated in the study. The SS III students were selected because they have through their SS1, SS II and III been taught all the relevant topics in Economics and are been prepared to write their final examination in the same area.

3.3. Instrument for Data Collection

The Northwest University, Kano 2014/2015 and 2015/2016 Post-UTME (NWU Post-UTME) questions which were designed and constructed for assessing the suitability of the prospective students to be admitted for 100level undergraduate programmes of Northwest University, Kano was used. The NWU Post-UTME uses multiple-choice items format with four answer choices/options (A-D). The NWU Post-UTME items were used for 2014/2015 and 2015/2016 admission exercises respectively and may likely be replicated in the coming admission exercises.

3.4. Data Collection Procedure

The NWU Post-UTME 50 multiple-choice items were administered to the sample after receiving specific instruction for the test by the researchers with the help of research assistants and the teachers in the samples schools. The test items were dichotomously scored and the students' responses and scores from the test are used for the analysis.

3.5. Method of Data Analysis

Data collected were scored dichotomously (right = 1, wrong = 0), the data file was prepared using Microsoft Excel 2010. The IRT item analysis and detection of poor items were carried out. The two psychometric properties of the items were determined. Two specialised Softwares (i.e. XCALIBRE 4.2 and SPSS 20v) were used in order to analyse the Economics Test items in this study. The SPSS 20V was used to assess the most important assumption common to all IRT models (i.e unidimensionality), Principal component factor analysis was carried out, and the eigenvalues were checked. To estimate students' abilities and item difficulty and discrimination for the tests, as well as the goodness of fit of the items according to IRT, XCALIBRE 4.2 software was used for the analysis.

4. PRESENTATION OF THE RESEARCH FINDINGS

The result of this study as explained in the method of data analysis above is presented in the form of IRT analysis. Similarly, all results were presented under each research question. Table 1 presents the Post-UTME Economics test itemssummary statistics. The total number of items in the test is forty nine (49) and the number of students who sat for the test is 600 as presented in the second column of the table. The overall reliability which is called internal consistency reliability coefficient of the test as measured by the Cronbach's Alpha is 0.86; this shows that the test is reliable since the coefficientis high, greater than the [25] recommended anacceptable value of 0.70. Similarly the items mean score is 28.82 with the standard deviation of 3.59. The mean item difficulty (b-values) is -0.63; the mean item discrimination of the test (a-values) is 0.28 and the mean student's ability (θ) is 0.074 as presented.

Table 1. Summary Statisticsfor all Calibrated Items

Items *(n)*	Examinees *(n)*	Reliability *(Alpha)*	Mean *Scores*	Means *b*	Mean *a*	Mean θ
49	600	0.86	28.82 (3.59)	-0.63 (2.54)	0.28 (0.04)	0.074(0.94)

Research Question 1: Do the NWU Post-UTME Economics test items fit into a 1-parameter logistic (1PL)/Rasch, 2-parameter logistic (2PL), and 3-parameter logistic (3PL) IRT models ? The results for items fit in the XCALIBRE output for 1PL, 2PL and 3PL analysis of test data are given in the form of a graph which displays of the fit between the IRF and the observed proportions of correct responses across examinee's ability levels. Thep-values associated with two popular statistical tests for identifying item fit, Chi-square and standardized residual (z) with aprobability of less than 0.05 ($p < .05$) is signalling items misfit. As maintained by [26] given the sensitivity of the Chi-square test to sample size, its p-values for this test may be ignored and use the p-values associated with the standardized residual, z, instead. Table 2 provides the number of items identified as misfitting the given IRT models at the [Alpha] =.05 level.

Table 2. Summary of items fitting each model

IRT Model	1PL	2PL	3PL
Number of Items fitting the model	47	49	38
Non-fitting Items	30 & 45	0	11, 29, 39 & 44
% of items fitting the model	95.9%	100%	91.8%

From the summary of standardised residual (z Resid) result fit test above, two items (Item30 and 45) and four items (11, 29, 39 and 44). This means that4.1% and 8.2% of the total items in the test were statistically significant and did not fit the 1PL and 3PL models respectively. However all the 49 items were not statistically significant and fitted the 2PL model 0.05 level of significance. Thus 2PL as the model with the most compatibility to the test data, where the entire 49 test items fitted in was found suitable and therefore used to estimate the item statistics based on the IRT model in this study. Research Question 2 : Do the NWU Post-UTME Economics test items satisfy the essential IRT basic assumption of unidimensionality and local item independence ?

The results of the factor analysis produce eighteen (18) items with eigenvalues greater than one. These 18 factors explain 80.22% of the variance. The first eigenvalue was 5.84 higher than the next eigenvalue (i.e 3.72, 3.10, e.t.c).The first factor explained 11.67% of the variance; the second factor explained 7.44% of the remaining variance. The remaining variances were explained by other 31 factor. Hence, there is one dominating factor in the factor structure of the item set. Since there is adominating factor that explained 11.67% of the variance the assumption of unidimensionality is established. The result of the eigenvalue test produced the scree plot to determine whether the dimensionality could be inferred. Looking at Figure 2, the eigenvalue of the first factor was larger compared to the second factor, and the eigenvalue of the remaining factors are all about the same.

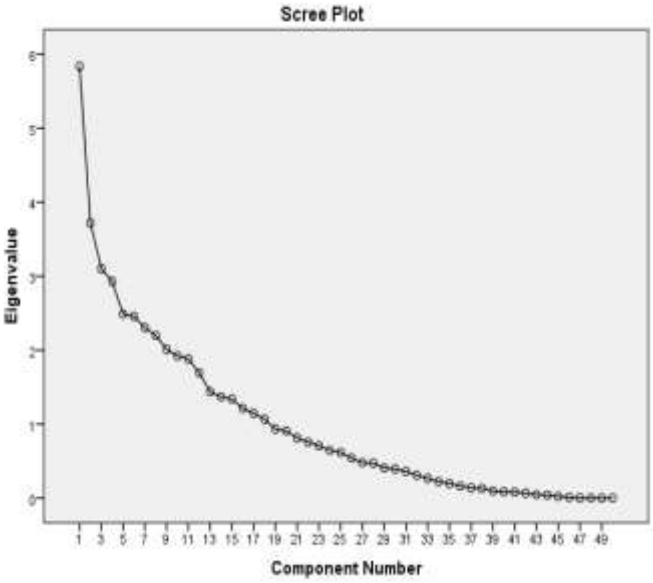

Figure 2.Scree plot for eigenvalues of Post-UTME EAT

Research Question 3: What are the item parameters (discrimination values, difficulty, and pseudo-guessing parameters) of the NWU Post-UTME Economics test items ? The item parameters of Post-UTME Economics Achievement test generated using IRT framework [i.e item difficulty (threshold or b) and item discrimination (slope or a)] are presented inTable 3.

Figure 3 displays the Test Information Function for all forty calibrated items. This present the amount of information the test is providing at each level of ability or theta (θ). The maximum information provided by the test was 2.022 at an ability or theta level (θ) of -1.600.

Table 3.Item Parameters for Dichotomously Scored Post-UTME EAT Items

Item	Item Difficulty (b)	Item Discrimination (a)	Item	Item Difficulty (b)	Item Discrimination (a)
1	-4.00	0.42	26	-4.00	0.32
2	1.09	0.29	27	-2.33	0.27
3	-2.12	0.34	28	-4.00	0.30
4	-1.63	0.28	29	-1.99	0.25
5	2.49	0.24	30	4.00	0.25
6	4.00	0.28	31	-0.85	0.27
7	1.85	0.27	32	1.67	0.26
8	1.48	0.30	33	-4.00	0.32
9	-3.19	0.28	34	3.08	0.23
10	2.53	0.27	35	-4.00	0.31
11	-0.51	0.25	36	-2.45	0.28
12	-1.73	0.26	37	0.91	0.24
13	-4.00	0.31	38	-0.28	0.29
14	0.89	0.25	39	-2.78	0.27
15	-2.67	0.28	40	-1.63	0.27
16	-1.33	0.25	41	1.71	0.21
17	1.37	0.23	42	-3.83	0.29
18	4.00	0.30	43	0.61	0.28
19	***	***	44	2.87	0.32
20	-4.00	0.28	45	1.44	0.31
21	-3.22	0.31	46	-0.51	0.25
22	-3.25	0.25	47	-4.00	0.31
23	-1.64	0.23	38	-0.82	0.28
24	2.18	0.28	49	1.34	0.24
25	1.29	0.30	50	-0.85	0.27

*****items of special interest are in bold

*******Item 19 was automaticallyremoved by XCLIBRE because it has NO VARIANCE

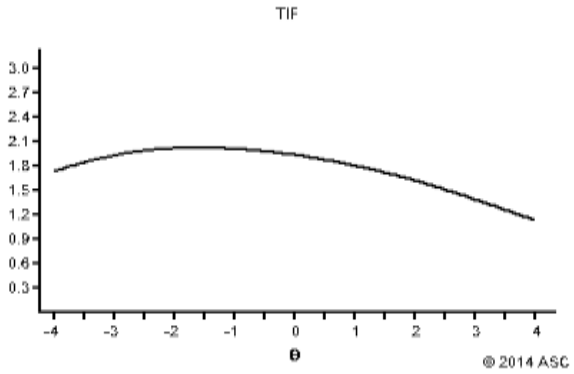

Figure 3. Test Information Function

5. DISCUSSION OF FINDINGS

The focus of this study is to evaluate the quality of the NWU Post-UTME economics test items using Item Response Theory (IRT) modeling.

5.1. Research Questions 1

The first question relates to IRT models assumption verification. The result of the factor analysis produced 18 items with eigenvalues greater than one. These 18 factors explain 80.22% of the variance. The first eigenvalue was 5.84 higher than the next eigenvalues.The first factor explained 11.67% of the variance; the second factor explained 7.44% of the remaining variance. The rest of the variance was explained by other 31 factor. Hence, there is one dominating factor in the factor structure of the item set. Since there is adominatingfactor that explained 11.70% of the variance the assumption of unidimensionality is established. The result of the eigenvalue test produced the scree plot to determine whether the dimensionality could be inferred. By this the test measures a unidimensional construct.

5.2. Research Questions 2

To test the model data fit, standardised residual (z Resid) fit test was used. All the 49 items were not statistically significant and fitted the 2PL model 0.05 level of significance. Thus 2PL as the model with the

most compatibility to the test data, where the entire 49 test items fitted in was found suitable and therefore used to estimate the item statistics based on the IRT model in this study.

5.3. Research Questions 3

The second question is dealing with the determination of the Post-UTME economics test item parameters. The parameters were determined and presented in Table 3 with items of special interest been highlighted in bold. Similarly, Figure 3 represents the Test Information Function for all 49calibrated items. The maximum information provided by the test was 2.022 at an ability or theta level (θ) of -1.600.

The results are interpreted base on the XCALIBRE recommended acceptable ranges of item parameters $a \geq 0.30$; $-3.0 \leq b \leq 3.0$ [23]. The findings reveal on the basis of *b* (difficulty) that 16 (33%) of the items were problematic. The Item Characteristic Curves of the test items show different behaviour (some were difficult while others wereeasier). Generally, on the basis of difficulty, 33 (67%) of the Items were of acceptable difficulty level. Example Figure 4 , presents the ICC obtained for item 1, shows that the value obtained on the ability scale (difficulty parameter estimate) 0.5 probability of examinees getting the item right is low (-4.00) this means the item is very easy for the students. However, Figure 5, obtained for items 6 shows that the value obtained on the ability scale 0.5 probability of examinees getting the item right is high (4.00) this means the item is difficult for the students. These difficult items should be rejected, modified or eliminated from the test completely.This finding is consistent with the findings of [26]-[28] whose findings revealed that, themajority of the items were acceptable as far as difficulty of the item**.**

The findings on the basis of item discrimination indices, the results indicates that 35 (71%) items failed to differentiate between students of different abilities having possessed poor or marginal discriminating ability,however 14 (29%) of the items were of moderate and good discriminating ability. Similarly, these 35 items which present a marginal and poordiscriminating abilities of$\leq$0.30, cannot differentiate substantially between low and higher achieving students, the items therefore need to be reviewed or should be rejected.

The poor performance of the identified test items and the students could have been due to poor understanding of difficult topics, ambiguity in wordings of the questions or even inappropriate key; it may also be due to personal variations in students' intelligence level [29]. This finding disagree with the that of many studies example;[27]-[30] whose findings revealed majority of the test items used in their studies (i.e more than 50%) were within acceptable level of item difficulty and discrimination.

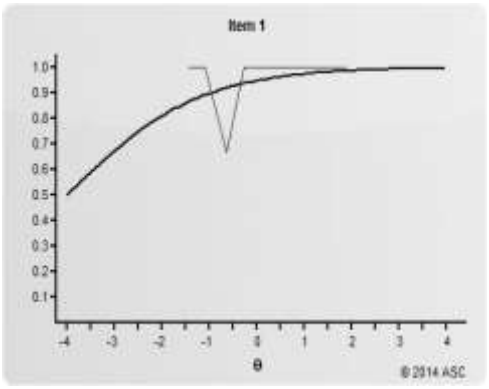

Figure 4. ICC of Item 1

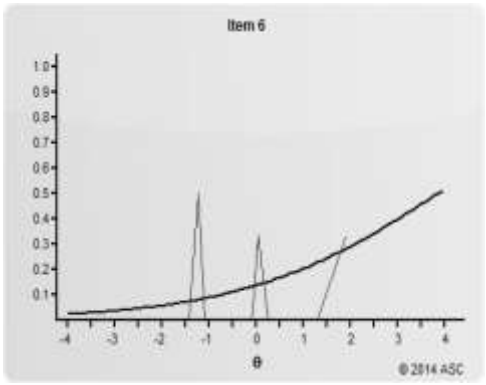

Figure 5. ICC of Item 6

6. CONCLUSION AND RECOMMENDATIONS

Findings of this studyindicates that the items of the qualifying examination are not stable as far as item discrimination and item difficulty indices are concerned using the IRT Frameworks.Item analysis results generated may be influenced by many other factors which include examinees having apoor understanding of difficult topics, ambiguity in wordings of the questions or even inappropriate key, instructional procedure applied, it may also be due to personal variations in students' intelligence level. Similarly, the results of this study shows the need to improve the Post-UTME test items of Nigerian Universities especially in developing valid and reliable items, through the mandatory involment of expert in test, measurement and evaluation in the process.

It is recommended that IRT should be maintained in development and analysis of test items, because of its position in the investigation of reliability and in minimizing measurement errors.In compliance with the

standards and current practice in development and validation of test items, the 'problematic' items identified in this study, having failed to satisfy the set quality criteria, should be modified, dropped, or completely eliminated from the test.The Post-UTME test items in other subjects should be subjected to psychometric analysis to ascertain it's quality.

Test items in the Post-UTME should be made to pass through all process of standardisation and validation; test development and content experts should be involve in developing and validating the test items in order to obtain valid and reliable results which will lead to valid inferences. Finally, further study needs to be conducted with different testitems and should include differential item functioning analysis to ensure that valid and reliable measuring items are used in selecting prospective undergraduates into Nigerian universities.

ACKNOWLEDGEMENTS

The authors appreciatethe Tertiary Education Trust Fund (tetfund) Ministry of Education, Federal Republic of Nigeria for providing funding for the project through Institutional Based Research (IBR) scheme. Smilarly, authors aknowlegde the students who participated in the tests as well as all the cooperating schools.

REFERENCES

[1] S. P. Klein and L. S. Hamilton, "Large-scale testing: Current practices and new directions (IP-182)," Santa Monica, CA, RAND, 1999.
[2] L. F. Gurski, "Secondary Teachers' Assessment and Grading Practices in Inclusive Classrooms," A Thesis Submitted to the College of Graduate Studies and Research in Partial Fulfilment of the Requirements for the Degree of Master of Education, University of Saskatchewan, 2008.
[3] L. S. Hamilton, et al.,"Making Sense of Test-Based Accountability in Education. Santa Monica," CA, RAND, 2002.
[4] N. E. Gronlund and R. L. Linn, "Measurement and Evaluation in Teaching (6th Ed)," MacMillan, New York, 1990.
[5] I. K. Olayemi and O. S. Oyelekan, "Analysis of Matriculation and Post-Matriculation Examination Scores of Biological Science Students of Federal University of Technology, Minna Nigeria,"*Ilorin Journal of Education*, vol. 28, pp.11-18, 2009.
[6] J. Okoje, "State of Nigerian Universities," A Paper Presented at the University of Port Harcourt Alumni Association forum in Abuja, 2008.
[7] M. Amatareotubo, "Post-UME screening: Matters arising," 2006. Posted to the Web: 8/30/2006 5:49:17pm: amasmozimo@yahoo.co.uk. Retrieved January 10, 2016 from website: http://www.onlinenigeria.com/articles/
[8] U. N. Akanwa and P. C. Nkwocha, "Prediction of South Eastern Nigerian Students' under Graduate Scores with their UME and Post-UME Scores,"*IOSR Journal of Research & Method in Education*, vol/issue: 5(5), pp. 36-39, 2015.
[9] A. Ikoghode, "Post-UTME Screening In Nigerian Universities: How Relevant Today?"*International Journal of Education and Research*,vol/issue: 3(8), pp. 101-116, 2015.
[10] A. A. Bichi, *"Analysis of UTME and Post-UTME scores of education students at Northwest University Kano-Nigeria,"* Presented at the 1st International Conference on Education 2015, On 9th April, 2015 @ Novotel Beijing Xinqiao, Beijing China, 2015.
[11] S. O. Uhunmwuangho and O. Ogunbadeniyi, "The University Matriculation Examination as a Predictor of Performance in Post University Matriculation Examination: a Model for Educational development in the 21st Century," *Africa Research Review*, vol/issue: 8(1), 2014.
[12] K. Ebiri,"Post JAMB Basis for Admission says Obasanjo,"*Guardian*, pp. 49, 2006.
[13] R. P. Chalmers, "MIRT: A Multidimensional Item Response Theory Package for the R Environment,"*Journal of Statistical Software*,vol/issue: 48(6), pp. 1-29, 2012. http://www.jstatsoft.org/v48/i06. Assessed on 23 June, 2014.
[14] L. D. Trang, "Applying Item Response Theory Modeling in Educational Research," Graduate Theses and Dissertations, 2013.
[15] R. K. Hambleton and H. Swaminathan, "Item response theory: Principles and applications,"*Springer*, vol. 7, 1985.
[16] X. Fan, "Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item/Person Statistics,"*Educational and Psychological Measurement*,vol/issue: 58(3), pp. 357-381, 1998.
[17] G. R. Kiany and S. Jalali, "Theoretical and Practical Comparison of Classical Test Theory and Item-Response Theory,"*Iranian Journal of Applied Linguistics*, vol/issue: 1(12), 2009.
[18] R. K. Hambleton, et al.,"Fundamentals of Item Response Theory," Newbury Park, CA, Sage, 1991.
[19] T. G. Courville, "An Empirical Comparison of Item Response Theory and Classical Test Theory Item/Person Statistics, Ph.D Dissertation, Texas A & M University, 2004.
[20] R. E. Schumacker, "Item Response Theory," 2010. http://appliedmeasurementassociates.com/ama/assets/File/ITEM_RESPONSE_THEORY.pdf. Retrieved on 13 August, 2014.
[21] L. Crocker and J. Algina, "Introduction to Classical and Modern Test Theory," Holt, Rinehart and Winston, New York, USA, pp. 527, 1986.

[22] X. An and Y. F. Yung, *"Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It*," In Proceedings of the SAS Global Forum 2014 Conference, 2014. http://support.sas.Com/resources/papers/proceedings14/SAS364-2014.pdf.Assessed on June 9 2014.
[23] R. Guyer and N. A. Thompson, "User's Manual for Xcalibre item response theory calibration software, version 4.2," Woodbury MN, Assessment Systems Corporation, 2014.
[24] R. K. Hambleton and R. W. Jones, "Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development,"*Educational Measurement: Issues and Practice*, vol/issue: 12(3), pp. 38-47, 1993.
[25] J. C. Nunnally, "Psychometric Theory, 2nd Edn.," McGraw-Hill, New York, USA., pp. 701, 1978.
[26] D. M. Dimitrov, "IRT and True-Score Analysis of the NCA Tests for Teaching Skills," Technical Report, National Centre for Assessment in Higher Education, 2013.
[27] S. S. Pande, et al.,"Correlation between Difficulty & Discrimination Indices of MCQS in Formative Exam in Physiology," *South-East Asian Journal of Medical Education*,vol/issue: 7(1), 2013.
[28] Suruchi and S. R. Rana, "Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology,"*Paripex - Indian Journal of Research*,vol/issue: 3(6), pp. 56-58, 2014.
[29] A. A. Bichi, "Item analysis using a derived science achievement test data,"*International Journal of Science and Research (IJSR),* vol/issue: 4(5), pp. 1656-1662, 2015.
[30] M. R. Hingorjoand F. Jaleel, "Analysis of One-Best Multiple Choice Questions: The Difficulty Index, Discrimination Index and Distractor Efficiency,"*J. Pak. Med. Assoc*., vol. 62, pp. 142-147, 2012.

BIOGRAPHIES OF AUTHORS

Ado Abdu BICHI is a Lecturer in the Department of Arts and Social Sciences Education, Northwest University Kano, Nigeria. He obtained his M.Ed in Educational Assessment (Psychometric Methods) and Currently undergoing a graduate studies for the degree of Doctor of Philosophy (Ph.D) in Educational Assessment (*Psychometric Methods*) at the Universiti Sultan Zainal Abidin, Malaysia. His areas of teaching and research interest are Educational and Psychological Test, Measurement and Evaluation, Item Development and Analysis using Advanced Psychometric Models (*Classical Test Theory, Item Response Theory and Cognitive Diagnostic Models*), Research Methodologyand Educational Statistics. He is a member of Psychometric Society and American Statistical Association

Dr Hadiza HAFIZ is a lecturer in the Department of Arts and Social Sciences Education, Northwest University Kano.A respected academician with over 20 years teaching and research experience, obtained herPh.D in Curriculum and Instruction from the prestigious Bayero University, Kano-Nigeria in 2015. Her areas of teaching and research interest are Curriculum and Instruction, Social Studies Curriculum, Teaching Methodology, Entreprenuership Education and Research Methodology. Currently, she is theDeputy Dean Students' Affairs of the Northwest Universty, Kano. She is a professional teacher registered byTeachers Registration Council (TRC)of Nigeria

Samira Abdullahi Bello is a lecturer in the Department of Arts and Social Sciences Education, Northwest University Kano. A respected academician and researcher with over 20 years teaching and research experience in teacher education. Currently she is a Ph.D (Curriculum Studies) scholar atBayero University, Kano-Nigeria. Her areas of teaching and research interest are Curriculum and Instruction, Educational Technology, ICT in Education and Research Methodology. she is a professional teacher registered byTeachers Registration Council (TRC) of Nigeria