

# Parameters of Basic Science Test Item's of 2011 Basic Education Certificate Examination Using Item Response Theory (IRT) Approach in Delta State, Nigeria

*Oguguo, Basil C.E*

Department of Science Education,  
Faculty of Education, University of Nigeria, Nsukka

*Lotobi, Regina Awele*

Department of Educational Evaluation and Counseling Psychology  
University of Benin City, Nigeria

Doi: 10.19044/ejes.v6no1a2

URL:<http://dx.doi.org/10.19044/ejes.v6no1a2>

---

## Abstract

This study determined the psychometric properties of the examination items in 2011 Basic Education Certificate Examination for Basic Science. The design adopted was survey research design. The instrument for data collection was the 2011 Delta State Basic Education Certificate Examination (BECE) in Basic Science Multiple Choice Test Items. The IRT model for item selection was used to determine the estimates of the item parameters. The findings of the study revealed that 45, 45 and 40 items satisfied the IRT difficulty, discrimination and guessing parameter respectively. While 38 items satisfied the combined three IRT parameter estimates. This result revealed a significant difference in the number of the selected items that satisfied all three IRT parameter estimates in BECE Basic Science.

---

**Keywords:** Determination, test item's parameters, IRT, basic science, BECE.

## Introduction

The Nigerian government places great emphasis on the teaching and learning of science in our schools. Basic science is a revolutionary new introductory science curriculum developed for student's considering a career in science. Basic science formally known as integrated science is the first form of science a child comes across at the secondary school level. Hence, basic science prepares students at the junior secondary school level for core science subjects (physics, chemistry, biology, mathematics and geography) at the senior secondary school level. This implies that for students to be able to study

single science subjects at the senior secondary school level successfully such student had to be well-grounded in basic science at junior secondary school level. The subject has been made compulsory for all students because of its importance to everyday life. It plays key role towards the actualization of the scientific advancement in the society and enables students live efficiently within the global world. As a core subject, all students are expected to offer it, irrespective of their interest. Basic Science help learners develop skills needed for further learning for science in the senior classes and beyond. As important as this subject, there has been a downward trend in students' performance. The percentage of students who pass at credit level showed that from 2011 to 2014, the percentage of student with credit pass and above is less than 50% and occurring in a reducing trend (Ministry of Education Asaba, 2016). As a result of students' performances, researchers have done so many works on different teaching methodologies to determine if these methodologies could have positive influences on their performances but the issue of poor performances still remains the same.

Considering the fact that teaching methodologies could not be the sole factor responsible for poor performances of the students, there is need to determine item parameters of test items constructed by examination bodies in order to determine if the test items met the psychometric properties that is expected of the test item. Test as viewed by Omorogiuwa (2010) is an item or set of items presented to an individual to whom they are expected to respond under specific conditions, with the intent to determine the extent to which such trait is present or absent in the respondent(s). Test items are composed of series of questions designed to measure series of behaviors such as intelligence, aptitude, attitude, acquired skills and knowledge in all field of education. For test items to achieve its aim, the test items must meet up with the theoretical scale for item selection using the item parameters.

Item parameters are statistical indicators that define the quality of an item in the instrument employed (Orheruata, 2015). These statistical indicators are item difficulty, discrimination and guessing parameters. Item difficulty parameter (b) refers to the examinee's ability level at which approximately half of the examinees are likely to answer a particular item correctly. Item discrimination parameter (a) describes the strength of an item discriminating between examinee with trait level ( $\theta$ ) below and above the threshold (b) while guessing parameter (c) is the probability of getting the item correct by guessing alone (Embretson & Reise, 2010). A well-developed test needed to have its item parameters in conformity with the theoretical scale for item selection using test theories. McDonald, (1999) defined test theory as the fundamental collection of mathematical concepts that formalize and clarify certain questions about constructing and using tests, and then provide some methods

for answering them. The author, further explained test theories as tools for addressing inferential problem within the measurement framework.

There are two test theories used in determining test item parameters. They are the Classical Test Theory (CTT) and the Item Response Theory (IRT). The Classical Test Theory (CTT) and the Item Response Theory (IRT) are the two contrasting measurements frameworks which address measurement problem. Both CTT and IRT are scientific methods which have a pioneer role in educational measurement and psychometric process. It is a known fact that Basic Education Certificate Examination body has been determining their test items before now using CTT (Ministry of Education Basic and Secondary Examinations and Standards, Asaba).

Classical Test Theory (CTT) according to Allen and Yen (2002) is a body of related psychometric theory that predicts outcomes of psychological testing such as difficulty of items or the ability of candidates. The theory actually deals with the effect of both unsystematic and systematic influence on the observed test theories. Classical Test Theory is based on the decomposition of observed scores into true and error scores and also views that the observed score changes as the amount of random (unsystematic) error changes. The problem with Classical Test Theory estimators has to do with circular dependency. It is also on record that the Classical Test Theory estimators of item difficulty and item discrimination indexes are not generalizable across populations. Classical Test Theory is limited in the comparison of performance of different examinees. The examinees must be given either the same or parallel items (Omorogiuwa, 2009). Embretson and Reise (2010) reported that Classical Test Theory provides no bases for determining how an examinee might perform when confronted with a test item and that CTT assumes that the measurement error is the same for all examinees. These limitations have led to a new measurement theory which is Item Response Theory (IRT).

Item response theory is one of the statistical frameworks that generate a mathematical function to describe the relationship between student performance in a test and ability or trait level. Its procedure improves psychometric methodology and assessment instruments. It provides meaningful information about examinee when its methodology is focused on the relationship between each individual item and the underlying (latent) trait or ability assessed by the instrument. A group of items responded to by a group of examinees are used to estimate the item parameters in order to discover an item's measurement qualities. A test developed using IRT provide information about an item at its difficulty level, discrimination level, and guessing level for efficient procedure for estimating item parameters.

The name Item Response Theory is due to the focus of the theory on the item, as opposed to the test-level focus of Classical Test Theory, by modeling the response of an examinee of given ability to each item in the test.

The term item is used because many test questions are not actually questions; they might be multiple choice questions that have incorrect and correct responses. IRT is based on the idea that the probability of a correct/keyed response to an item is a mathematical function of person and item parameters. Item parameters include difficulty (location), discrimination (slope or correlation), and pseudo-guessing (lower asymptote). Considering the quality of IRT, this study was an attempt to determine item parameters in 2011 Basic Science test items in Delta State Basic Education Certificate Examination using IRT approach.

### **Theoretical Approach**

This study is hinged on item response theory (IRT) framework. The theory was adopted due to its applicability to sift through item level statistics. IRT generated new rules of measurement and presented as modern and superior alternative to CTT (De Boeck & Wilson, 2004; Embretson & Reise, 2010; Nering & Ostini, 2010, Zickar & Broadfoot, 2008). Three of the pioneers who pursued parallel research work independently were the Educational Testing Services psychometricians known as Frederic M. Lord, the Danish mathematician, George Rasch, and Austrian sociologist Paul Lazarsfeld (Millikarjuna, 2014). Lord (1952) brought the idea of latent trait or ability and at the same time differentiated this construct from observed score. Lazarsfeld (1950) only described the unobserved variable as accounting for the observed interrelationships among the item responses. While Rasch (1960) reported the need for creating statistical models that maintain the property of specific objectivity, the idea that people and item parameters be estimated separately but comparable on similar metric.

In using IRT, one can assess through the item characteristics within a multiple item test and estimate the examinee's ability given the item parameters and the response pattern to the test by that examinee. A more comprehensive approach to psychometrics that rectifies many of the perceived shortcomings associated with classical approaches is also provided by IRT. It implemented new concepts to describe tests (item characteristics curve, item/test information function and so on). It puts the focus on the estimation of item's operational characteristics like assessment of test dimensionality, estimating of the difficulty, discriminating, guessing parameters, item bias and differential item functioning. IRT methods differentiate error more finely, most especially with respect to characteristics of individual items that may affect their performance. A goal of IRT is to enable a researcher to establish certain characteristics of items that are independent of who is completing them. It examines the level of the attribute being measured that most strongly influences an item. The purpose of IRT is to estimate both the value of the latent trait for each respondent and the item parameters for each item. In IRT,

each item specifies three parameters that define an s-shape logistic curve, called item characteristic curve (ICC), linking probabilities of individuals into position of the individuals in the latent trait (or ability).

### **Statement of the Problem**

Over the years, the Basic Education Certificate Examination Chief examiner's report in Delta State has been showing a downward trend in the performance of students in Basic Science. In order to overcome this issue of poor performances, the government has been organizing workshop and seminars to improve on teachers' professional skills most especially in the area of teaching methodology but the issue still remains the same. The persistent issue of poor performances has shown that the teaching methodology could not be the sole factor responsible for students' poor performances. As a result of this, there is need to determine the psychometric properties of the examination items used to determine the performance of students. Psychometricians have relied so much on observed scores to represent the best estimate of a person's ability. Examinee ability is obtained from the score an individual obtains from a test and this really depends on the quality of the test taken. As a matter of fact, a student's score is due to some unspecified combination of his/her ability and the properties of the items, such as how difficult they are. When representative samples are carefully selected, reliable item and test statistics can be used to generate parallel forms measuring the same construct. Determination of the psychometric properties of examination items are important to ensure public examinations items are measuring good psychometric properties that are expected of the items used for examination purpose. Based on the above, the problem of this study is to determine item parameters in 2011 Basic Science test items in Delta State Basic Education Certificate Examination using Item Response Theory approach.

The main purpose of the study was to determine the item parameters of 2011 Delta State BECE Basic Science multiple choice test items using the Item Response Theory approach. Specifically, an attempt was made to:

- determine the number of items of the 2011 Delta State BECE Basic Science multiple choice items that satisfied the theoretical expectation of the IRT difficulty parameter estimate.
- determine the number of items of the 2011 Delta State BECE Basic Science multiple choice items that satisfied the theoretical expectation of the IRT discrimination parameter estimate.
- determine the number of items of the 2011 Delta State BECE Basic Science multiple choice items that satisfied the theoretical expectation of the IRT guessing parameter estimate.

- determine the number of items of the 2011 Delta State BECE Basic Science multiple choice items that satisfied the theoretical expectation of the combined IRT item parameter estimates.

### **Research Questions**

The following research questions were raised to guide the study:

1. How many of the 2011 Delta State BECE Basic Science multiple choice items satisfied the theoretical expectation of the IRT difficulty parameter estimate?
2. How many of the 2011 Delta State BECE Basic Science multiple choice items satisfied the theoretical expectation of the IRT discrimination parameter estimate?
3. How many of the 2011 Delta State BECE Basic Science multiple choice items satisfied the theoretical expectation of the IRT guessing parameter estimate?
4. How many of the 2011 Delta State BECE Basic Science multiple choice items satisfied the theoretical expectation of the IRT combined item parameter estimates?

### **Hypotheses**

The following hypotheses were formulated to guide the study:

H<sub>01</sub>: There is no significant difference in the number of items that satisfied the theoretical expectation of the IRT item difficulty parameter estimate and the total number of items in BECE Basic Science.

H<sub>02</sub>: There is no significant difference in the number of items that satisfied the theoretical expectation of the IRT item discrimination parameter estimate and the total number of items in BECE Basic Science.

H<sub>03</sub>: There is no significant difference in the number of items that satisfied the theoretical expectation of the IRT guessing parameter estimate and the total number of items in BECE Basic Science.

H<sub>04</sub>: There is no significant difference in the number of items that satisfied the theoretical expectation of all selected three IRT parameter estimates and the total number of items in BECE Basic Science.

### **Methods**

The survey research design was adopted in this study. This design is employed because it meets the expectation of this study for effective analysis. It is adopted to collecting data on and describing in a systematic manner, the characteristics, features and facts about the population of the study. The strength of this design is that it provides latitude for full description of relevant variables in relation to the given population. This survey research design is concerned with description of events as they are. It is purely for collecting and

interpreting information. It does not involve manipulation of information. Population of this study consist of Three Thousand, Nine Hundred and Thirty-one (3,931) students from Twenty-Three public schools preparing to write their 2014/2015 BECE in Oshimili North and Oshimili South Local Government Areas of Delta State. The number represents the entire population of JSS3 students in both Local Government Areas. The sample of students for this study comprise of One Thousand, Two Hundred and Ninety-Seven students (1,297). Proportionate stratified sampling technique was used to select 33% of 23 schools which is 8 schools. This includes 4 schools from Oshimili North Local Government Area and 4 schools from Oshimili South Local Government Area. Simple Random Sampling Technique was used by balloting to select the 1,297 students from the selected 8 schools in Oshimili North and South Local Government Areas.

The research instrument is the June 2011 Basic Science Multiple Choice Test Items of Basic Education Certificate Examination (BECE). It contains sixty (60) items with one correct answer and four distracters drawn across the Basic Science syllabus. The instrument was adopted whole without modifications. The validity of the instrument was done by the BECE Board being a standardized item by their originality. The Board has the credit and merit of adopting different approaches of determining validity. The reliability of BECE instrument used was also done by the Board and as such regarded as standardized items. The permission of the school principals and Basic Science teachers was sought and obtained before the administration of the instrument. The teachers informed the students of writing the examination in Basic Science in two weeks before the examination. The instrument was administered to the students with the help of 3 Basic Science teachers of the various schools selected serving as research assistants and the researchers personally supervised some schools that were used for the study. The instrument was administered as mock examination under similar conditions as given by examination body. Having retrieved all the data responses from one thousand two hundred and ninety-seven (1,297) students, the response sheets were sorted out. Correct responses were scored as "1" and incorrect as "0". The response sheets were numbered from 1 to 1297. Having generated the item parameter estimates, the selection of items using the IRT procedures were carried out. The selected items were used to answer the research questions 1 through 4. For research question 1 through 4, frequency count was adopted. The obtained number of satisfied and non-satisfied items of item difficulty parameter estimate, item discrimination parameter estimates, the guessing parameter estimate and all selected three parameter estimates were subjected to further analysis using the Statistical Package for Social Sciences (S.P.S.S). Hypotheses 1 through 4 were tested using chi-square test of goodness of fit at

an alpha level of 0.05. Selection of items was based on having satisfied all three IRT parameter estimate concurrently.

## Results

**Table 2; Number of items that satisfied the theoretical expectation of the IRT difficulty parameter estimate and those that did not**

Status of item	No of items	Items	%
Satisfied	45	1,3,5,6,10,12,14,15,17,19,20,21,22,23, 24,27,28,29,30,31,32,33,34,35,37,38,39, 40,41,42,43,44,46,47,48,49,50,51,53,54,55, 56,57,59&60	75.00
Not satisfied	15	2, 4,7,8,9,11,13, 16,18,25,26,36,45, 52 &58	25.00

Table 2 above shows that the number of items that satisfied the theoretical expectation of the IRT difficulty parameter estimate are 45 which is 75.00% of the items and 15 (25.00%) did not satisfy it.

**Table 3; Number of items that satisfied the theoretical expectation of the IRT discrimination parameter estimate and those that did not**

Status of items	No of items	Items	%
Satisfied	45	1,3,4,5,6,10,12,14,15,17,19,20,21,22,23,24, 27,28,29,30,31,32,33,34,35,37,38,39,40,41,42, 44,46,47,48,49,50,51,53,54,55, 56,57,59 &60	75.00
Not satisfied	15	2,7,8,9,11,13,16,18,25,26,36,43,45,52&58	25.00

Table 3 shows that the number of items that satisfied the theoretical expectation of the IRT discrimination parameter estimate are 45 which is 75.00% of the items and 15 (25.00%) did not satisfy it.

**Table 4; Number of items that satisfied the theoretical expectation of the IRT guessing parameter estimate**

Status of items	No of items	Items	%
Satisfied	40	1,2,3,4,5,6,10,12,14,17,20,21,22,23,24, 27,28,29,30,31,32,34,35,41,42,44,46,47, 48,49,50,51,53,54,55,56,57,58,59&60	66.67
Not satisfied	20	7,8,9,11,13,15,16,18,19,25,26,33,36,37, 38,39,40,43,45&52	33.33

Table 4 shows that the number of items that satisfied the theoretical expectation of the IRT guessing parameter estimate are 40 which is 66.67% of the items and 20 (33.33%) did not satisfy it.



**Table 5; Number of items that satisfied the theoretical expectation of the IRT combined item parameter estimates and those that did not**

Status of items	No of items	Items	%
Satisfied	38	1,3,4,5,6,10,12,14,17,20,21,22,23,24,27,28, 29,30,31,32,34,35,41,42,44,46,47,48,49,50, 51,53,54,55,56,57,59, &60	63.33
Not satisfied	22	2,7,8,9,11,13,15,16,18,19,25,26,33,36,37,38, 39,40,43,45,52&58	36.67

Table 5 shows that the number of items that satisfied the theoretical expectation of the combined IRT parameter estimates are 38 which is 63.33% of the items and 22 (36.67%) did not satisfy it.

**Table 6; Chi-square goodness of fit of IRT item difficulty parameter estimate**

Items	Observed	Expected	Df	Chi-square	Sig
Fit	45	30	1	15.0	0.295
Not Fit	15	30			

$\alpha = .05$

Table 6 shows a chi-square value of 15.0 and a p-value of .295. Testing at an alpha level of .05, the p-value is greater than the alpha level, so the null hypothesis which states that there is no significant difference in the number of items that satisfied IRT item difficulty parameter estimate and the total number of items in Basic Education Certificate Examination in Basic Science is retained.

**Table 7; Chi-square goodness of fit of IRT item discrimination parameter estimate**

Items	Observed	Expected	Df	Chi-square	Sig
Fit	45	30	1	15.0	0.295
Not Fit	15	30			

$\alpha = .05$

Table 7 shows a chi-square value of 15.0 and a p-value of .295. Testing at an alpha level of .05, the p-value is greater than the alpha level, so the null hypothesis which states that there is no significant difference in the number of items that satisfied IRT item discrimination parameter estimate and the total number of items in Basic Education Certificate Examination in Basic Science is retained. There is no significant difference in the number of items that satisfied IRT guessing parameter estimate and the total number of items in Basic Education Certificate Examination in Basic science.

**Table 8; Chi-square goodness of fit of IRT guessing parameter estimate**

Items	Observed	Expected	Df	Chi-square	Sig
Fit	40	30			
			1	3.574	0.059
Not Fit	20	30			

$\alpha = .05$

Table 8 shows a chi-square value of 3.574 and a p-value of .059. Testing at an alpha level of .05, the p-value is greater than the alpha level, so that null hypothesis which states that there is no significant difference in the number of items that satisfied IRT guessing parameter estimate and the total number of items in Basic Education Certificate Examination in Basic Science is retained.

**Table 9; Chi-square goodness of fit of all selected three IRT item parameter estimates**

Items	Observed	Expected	Df	Chi-square	Sig
Fit	38	30			
			1	4.939	0.026
Not Fit	22	30			

$\alpha = .05$

Table 9 shows a chi-square value of 4.939 and a p-value of .026. Testing at an alpha level of .05, the p-value is less than the alpha level. Therefore, the null hypothesis which states that there is no significant difference in the number of items that satisfied all selected three IRT parameter estimates and the total number of items in Basic Education Certificate Examination in Basic Science is rejected.

## Discussion

### Number of items that satisfied the theoretical expectation of the IRT difficulty parameter estimate

The findings revealed that there is no significant difference in the number of items that satisfied the IRT item difficulty parameter estimate and the total number of items in BECE Basic Science. What this means is that more than half of the number of items selected satisfied the item difficulty parameter estimate indicating that the items are good. Table 6 confirmed items that satisfied IRT item difficulty parameter estimate of BECE 2011 Basic Science multiple choice test items. The selected items are 45 (75%) of the 60 total Basic Science multiple choice test items. This implies that the achievement test met the required difficulty level because more than half of the items satisfied the item difficulty parameter estimate. The result could also mean that examinees really have the ability to answer the difficult items. It also showed that the number of items selected was good enough. Therefore, there is no significant difference in the number of items that satisfied IRT item difficulty parameter

estimate and the total number of items in Basic Education Certificate Examination in Basic science. This only means that generally, the items for the achievement test truly represent the learning ability of the test taker because most of the items can discriminate well between high and low performing groups. The findings are in agreement with the report of Bock, Muraki and Pfeiffenbergger (1998) who reported that only 63.22% of the physics achievement items satisfied item difficulty.

### **Number of items that satisfied IRT item discrimination parameter estimate**

The finding revealed that there is no significant difference in the number of items that satisfied IRT item discrimination parameter estimate and the total number of items in Basic Education Certificate Examination in Basic science. This shows that the test items satisfied the item discrimination parameter estimate. This also shows that the number of items selected were good enough. Table 7 confirmed items that satisfied the IRT item discrimination parameter estimate of BECE 2011 Basic Science Examination objective test items. The selected items 45 of the 60 total Basic Science multiple choice test items presented for BECE. From this result, one would rightly say that more than half of the items satisfied the item discrimination parameter estimate. This goes to show that the items discriminated well enough. It also showed in the results that the number of items selected were good. Therefore, there is no significant difference in the number of items that satisfied the IRT item discrimination parameter estimate and the total number of items in Basic Education Certificate Examination in Basic science. The findings disagree with the findings of Pedraza, Sachs, Ferman, Rush and Lucas (2011). Results showed that successive items do not necessarily reflect a monotonic increase in psychometric difficulty, some items are inadequate to distinguish individuals at various levels of naming ability, multiple items provide redundant psychometric information, and measurement precision is greatest for persons within a low-average range of ability.

### **Number of items that satisfied the IRT guessing parameter estimate**

The finding showed that there is no significant difference in the number of items that satisfied IRT guessing parameter estimate and the total number of items in Basic Education Certificate Examination in Basic science. This finding revealed that the test items also satisfied the guessing parameter estimate which also shows that the number of items selected was good enough. Table 10 confirmed items that satisfied the IRT item guessing parameter estimate of BECE 2011 Basic Science examination objective test items. The selected items are 40 of the 60 total Basic Science multiple choice test items presented for BECE. From this result, it showed that the number of items

selected is good. Therefore, there is no significant difference in the number of items that satisfied IRT guessing parameter estimate and the total number of items in Basic Education Certificate Examination in Basic Science. This is in line with Obinne (2012) who reported that the way an item is written can influence guessing on the item and advised that IRT method of item analysis should be employed to eliminate those items prone to guessing, so that when guessing occurs, it will not be blamed on the item.

### **Items that satisfied the IRT combined item parameter estimates**

Thirty-eight items satisfied the IRT combined item parameter estimates and twenty-two did not. This shows that the number of items selected was not good enough. Table 5 confirm items that satisfied the IRT combined item parameter estimates of BECE 2011 Basic Science Examination objective test items. The selected items are thirty-eight of the sixty total Basic Science multiple choice test items presented for BECE. The findings of this result is at variance with the study of Orheruata (2015) who reported that 35, 48 and 31 percent of WAEC items satisfied the combined parameter estimates for 2012, 2013 and 2014 examination years respectively. These findings disagreed with Omorogiwa (2009) who reported that only 46 percent of the 60 physics items satisfied the IRT combined parameter estimates. The result also revealed that there is a significant difference in the number of items that satisfied all selected three IRT parameter estimates and the total number of items in BECE Basic Science. This implies that when all selected three IRT parameter estimates were applied concurrently, the items selected did not satisfy these three conditions. This shows that the items failed the general test. Table 9 confirmed that the items selected did not satisfy these three conditions. Therefore, the result showed that the number of items selected were not good enough. This is because the p-value is less than 0.05. Therefore, there is a significant difference in the number of items that satisfied all selected three IRT item parameter estimates and the total number of items in Basic Education Certificate Examination in Basic Science. This therefore confirms that these items did not satisfy the selection procedures as required. Even though individually they satisfied some of these parameters when separated, but when the parameters were combined together, they did not satisfy these parameters for selection. This is because items can satisfy difficulty parameter and yet fail to satisfy discrimination parameter. Items can satisfy difficulty parameter and yet fail to satisfy the guessing parameter. Before any item can be selected, it must satisfy all conditions of all three IRT item parameter estimates, for the item to be desirable. The findings of this study confirms the views of McAlphine (2002) who reported that poor estimation of “C” parameter can degrade the estimates obtained for other parameters and unless limits are put on the item and ability parameter values, these procedures can fail, it is also possible that the test in

each item tends to confuse good students and that, can lead to a drop of performance which was not good enough. When the students have a 50% chance to answer the item correctly, the student has no advantage over the item. Examinees that did well on the total test tended to do well on the item as well. However, an interaction can exist between item discrimination and item difficulty. Item answered correctly (Or incorrectly) can have reduced power also to discriminate and thus, can have lower correlations.

Items with extreme difficulties can have lower discrimination values while Items with high discrimination power contributes more to the precision of measurement than items with lower discriminating power. An item provides less information when guessing is a factor in test performance. When guessing occurs, item provide their maximum information at a point slightly higher than the point corresponding to difficulty level. This was why Birnbaum (1968) questioned: Do the items in a test really differ from each other in discriminating power? Linacre, (2007), Lumsden (1978) argued that unequal item discrimination parameter indicates a violation of unidimensionality assumption governing the discrimination parameter and unidimensionality, need to be evaluated to determine the appropriateness of an item response model for a given data set.

## Conclusion

When the item parameters were tested differently, that is difficulty, discrimination and guessing parameters, the number of items selected was satisfactory but when the three IRT item parameters were used together, it was found out that it was not satisfactory. Therefore, there is a difference in the number of items selected in the total items when the three IRT item parameters were put together. Ministry of Education should ensure that even though the examinations are standardized by state and local government, they should also ensure that there is some central monitoring for further standardization and ensuring of standards and ensuring of accuracy, validity and reliability on the bases that they have been validated and worked upon by the state group. Item Response Theory (IRT) analysis should be integrated into the construction and analysis of public examinations by examination bodies in Nigeria.

## References:

1. Allen, M.J. & Yen, M.W. (2002). *Introduction to measurement theory*. Monterey CA: Brooke/Cole.
2. Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical*

- theories of mental test scores*. Reading, MA: Addison-Wesley. 397-472.
3. Bock, R.D., Muraki, E. & Pfiffenberger, W. (1998) Item pool maintenance in the presence of item parameter Drift. *Journal of Educational measurement* **25**(94): 275-278.
  4. De -Boeck, P. and Wilson, M. (2004). *Explanatory item response model: a generalized linear e non linear approach*. New York: Springer.
  5. Embretson, S. & Reise, S.P. (2000). *Item response theory for psychologists* Mahwah, NJ: Lawrence Erlbaum Associates.
  6. Embretson, S. E. & Reise, S. P. (2010). *Item response* (2<sup>nd</sup> ed.). New York: Routledge Academic.
  7. Lazarsfeld, P.F. (1950). The logical and mathematical foundation of late4nt structure analysis. In S. A. Stouffer, L. Guttman, Measurement and prediction. New York: Wiley. 362-412.
  8. Linacre, J.M. (2007). Estimation: Iteration and convergence. *Rasch Measurement Transactions*, **1**: 7-8.
  9. Lord, F.M. (1952). A theory of test scores. *Psychometric Monographs*, Whole N. 7.
  10. McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates. 288
  11. Mallikarjuna, G. (2014). Investigate into invariance properties of item response theory (IRT) by two and three parameter models in terms of item parameters /ability parameter of a group of online test takers and establish such invariance properties of IRT as opposed to variance of index of difficulty in classical test theory (CTT) of the same groups. *Thesis: Chitkara University Himachal Pradesh, India* pp.175.
  12. Neringe, M.L. and Ostini, R. (2010). *Handbook of polytomous item response theory models*. New York: Routledge.
  13. Obinne, A.D.E. (2012). Using IRT in determining test items prone to guessing. *Word Journal of Education*, 2 (1): 91-95.
  14. Omorogiwa, K.O. (2009). An empirical comparison of the classical test theory and item response theory in the selection of physics achievement test items. A Ph.D. Dissertation of the faculty of education. University of Benin, Benin City.
  15. Omorogiwa, K.O. (2010). *An introduction to educational measurement and evaluation*. Benin City: Perfect touch prints.
  16. Orheruata, M.U. (2015). Item Parameter Drift of 2012 to 2014 WAEC and NECO SSCE Agricultural Science multiple choice items using Item Response Theory. Unpublished Ph.D. Dissertation of faculty of education. University of Benin, Benin City.

17. Pedraza, O., Sachs, B.S., Ferman, T.J., Rush, B.K. & Lucas, J.A. (2011). Difficulty and discrimination parameters of Boston naming test items in a consecutive clinical series. *Archives of Clinical Neuropsychology*, 26: 434-444.
18. Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
19. Zickar, M.J. and Broadfoot, A.A. (2008). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C.E. Lance & R.J. Vandenberg (Eds.) *Statistical and methodological myths and urban legends*: New York: Routledge Academic