

Relevance Of Item Response Theory In The Assessment Of Learning In Tertiary Institutions In Rivers State.

Dr Ebere Sampson Wagbara

Faculty of Education, Department of Educational Foundations, Rivers State University

Abstract

Appropriate description of the learner has agitated the mind of measurement experts over the years. This has resulted in the Test Theories namely: Classical Test Theory and Item Response Theory. The Item Response Theory is an improvement of the classical test theory. The importance of Item Response Theory in teaching and learning as well as evaluation of learning outcomes gave impetus to this study. This paper examined the following with regard to the Item Response Theory: Assessment in Tertiary Institutions, Application of Item Response Theory in Assessment in Tertiary Institution, Relevance of Item Response Theory in assessment of learning outcomes which includes: estimating examinee ability and how the contributions of error might be minimized; reporting true scores or ability scores and associated confidence bands; correlations between variables and flexibility afforded by the model, the study recommended among others: there should be increased awareness of Item Response Theory in assessment in tertiary institutions in Rivers State, Item Response Theory should be adopted in test development, validation and standardization and Item Response Theory should be utilized in the interpreting the performance of undergraduates.

Keywords: Evaluation, Assessment, Item Respond Theory, Classical Test Theory and Tertiary Institution.

Introduction

Education is an instrument par excellence, Ocho (2005) maintained that the central aim of education is character training, knowledge and skill acquisition, however this cannot be understood by mere putting it on an ordinary weighing scale but through the instrumentality of mathematical application referred as measurement and assessment which is fundamental certification of any educational programme. Therefore, measurement and assessment play important purpose in evaluation of educational process that is primarily teaching and learning. In the teaching and learning process, assessment is unique because it provides relevant information concerning the identification of learners with or without special learning needs, purpose of grading, monitoring instructional effectiveness and classification of learners' achievements. Malcom (2003) asserted that assessment is an integral part of the teaching enterprise as it is essential for the generation of information that is required for decision making process in the educational terrain especially at tertiary institutions of learning.

A psychological test is an objective and a standardized measure of a sample of behaviour. Anene and Ndubtiisi (2003) denotes that a well-constructed test is believed to reflect a psychological construct such as the cognitive ability, aptitude, emotional functioning and personality among: others. The technical term for the science behind psychological testing is psychometrics. A test is an instrument designed to measure unobservable construct also known as latent' variables .Hambleton (1989) assert that testing could also involve a more comprehensive assessment of an individual. It is a process that involves checking the integration of information from multiple sources such as test of normal and abnormal personality, test of ability or intelligent, test of interest or attitude among others, that provides important data for assessment.

Importantly of note is that testing or test when employed is to unravel some hidden psychological abilities or aptitudes about individuals that would have not been known ordinarily.

Since test plays an important role in the educational ecosystem, it is therefore imperative that test development especially in the tertiary institution of learning should be of high standard that must be adequately, effectively and efficiently in ascertaining learners' affective, cognitive and behavioural skills. The tertiary institutions are the centre where "would be" or pre-service teachers are groomed and prepared to face the world of work especially in the teaching enterprise. Tertiary Institution is the most sensitive as it is the level of education that train the trainers and the nation manpower. It is therefore holds that procedures of test development and construction that comply with international best standards are followed in order for best upgrade in the production of quality graduates in Nigeria tertiary institutions. Therefore, looking at test papers from tertiary institutions in Rivers State, one notices that some test items constructed by lecturers' especially those that did not study education are lacking in some psychometric properties. Here lies the credence to examine the relevance of Item Responses Theory (IRT) in assessment of learning processes in tertiary institutions in Rivers State.

Assessment in tertiary Institutions

The tertiary institutions are synonymous to an environment where the building and enhancement of human resources is achieved through a conscientious formal setting and training. It is a pathway towards the acquisition of desirable knowledge, skills and values that ensure cordial and productive living in the society. The product of any tertiary institution of learning should be the deposition in its members' the capabilities for self-reliance expected for personal survival in the immediate environment and contributing to the continuous survival of others in the larger society. Fundamentally, the tertiary institutions are saddled with the responsibility of producing the required technical and professional manpower required to control and transform all areas of national development. From the foregoing, the education of the learner in terms of enhanced cognitive development that is linked with deeper intellectual skills, possession of technical skills and positive character building associated with values, respect, appreciation and feelings for the development of a sustainable nation is the core of teaching and learning processes in the tertiary institutions. Haiyang (2010) highlighted the aims of tertiary institutions as the;

1. Acquisition, inculcation and development of right values and positive orientation for the
 2. Survival of the individual in the society.
- In-depth development of intellectual capacities of individuals to understand the dynamics in the environment and consequently appreciate the inherent value of the environment.
1. Acquisition of both intellectual and physical skill necessary for functional living
 2. Developing into responsible members of the society.
 3. Acquiring objective perception of in view of the local and external environment.

Maio and Haddock (2010) stated that for the higher institution of learning to ascertain quality of education for her learners, the three cardinal points which are community service, research and . Teaching must be satisfied. There is no gainsaying that the tertiary institution teachers are key catalyst in achieving the required result for national economic growth of a nation. The university academic staff is of great significance in the educational process not only in achieving the desired aims and objectives of education but in promoting quality education. It is important to state that the university lecturers' attitudinal sense is a significant factor in enhancing students' intellectual capacity. Therefore, lecturer as professionals are experts in the art of communication, knowledge dissemination and importantly in the development of test items which is imperative for assessment and a yardstick for the production of students that can actively

compete in the global settings.

Teaching in the tertiary institution comes along with its multifaceted processes that require the participators (lecturers) to be competent in the area of measurement and assessment skills that are relevant in planning and construction of test, grading of test scores, interpretation of test results, use of assessment results to inform teaching and learning; interpretation of standardized tests; and communicating results to relevant stakeholders. Nenty, Adedoyin, Odili, and Major (2007) posited that the instructional approach and most fundamentally the assessment procedures of academic stall are the pathway through which the educational system can be enhanced and defined in terms of quality and sustainability.

In the tertiary institutions, academic members of staff adopt variety of assessment practices to evaluate the learner learning outcome. However, Christensen (2002), enumerated that in other for one to consider himself or herself an effective teacher, knowledge of educational assessment procedures is fundamental. As such, two important assessment information must be an integral part of the teaching and learning process which are;

1. The content and character instructional assessment must be improved
2. Collecting, collating and application of appropriate item analysis.

Application of item Response Theory in Assessment in Tertiary Institution

Item Response Theory (IRT) is an example of a measurement theory that have the characteristics of data acquisition at the ratio scale level, sample independent attributes and learners' ability that can be reported on both item and total instrument levels. Troy-Gerard (2004) asserted that the development of IRT was due to the limitations of the Classical Test Theory. In recognition of the above, Asuru (2015) supported that IRT is an improvement on the CTT which represents a fundamental shift for the purpose of design, analysis of data and scoring of test instrument. Several authors (Joshua & Ikiroma, 2012) have also referred to IRT as Mental Test Theory, Latent Trait Theory and Strong True Score.

Item Response Theory (IRT) is basically focused on the testee's abilities and attitude using mathematical models. There is no doubt that application of test theories especially the Item Response Theory (IRT) is essential towards the development of standard test items required for the enhancement of intellectual knowledge and skills of learners in the tertiary institution of. learning and subsequently producing quality graduate that can compete in the global work force.

The item response theory otherwise known as the latent trait theory is the paradigm that is focused on the design, analysis and scoring of test questionnaires another similar instruments which are used in measuring achievement, abilities, attitude and other fundament variables. This theory is hinged on mathematical models for testing of data (Hogg, & Vaughan, (2005). As opposed to the classical theory which is focused on the level of the test, item theory on the test items which implies that the theory takes into cognizance the response of each examinee of a given ability as it relates in the test.

Asuru.(201 5) noted that the Item Response Theory is premised on the assumption that the probability for an individual to obtain the correct response or key response is a function of the individual including the item parameters. The item parameters are attitude, strength or intelligence, while the parameters upon which the items are construed are item difficulties, item discrimination, validity and reliability. Ayala (2009) commented that the aim of the application of Item Response Theory is to provide a platform essential for evaluating the outcome of assessment. He further mentioned that IRT gives the probability that an individual who has a given ability level will produce a correct answer to an item, while the one with less ability has less chance to produce the correct answer. It is no gainsaying that emphasis by the psychometric and measurement community from classical to item response models is as a consequence of the benefits

obtained through the application of item response models to measurement problems.

The fundamental differences between IRT and CTT as identified in Asuru (2015 : 117) are:

S/N	Item Response Theory (IRT)	Classical Test Theory (CTT)
1	Models the relationship between the ability of a test taker in terms of his response to a test and the probability of correctly responding to a test item. Judges examinee's responses at the item level. Focuses on individual item (item parameter)	Concerned with the relationship of X, T and E in a population. It has judge's testees responses at the test level – focuses on the test i.e all the items taken together.
2	Does not depend on the true score model, but also indicates or exposes the unseen psychometric properties of the test and the examinee	Is based on true score model and thus depends on examinee's aggregate score in a test.
3	Provides robust information on examinee's responses to each item that constitutes the test	Does not provide adequate information on examinee's responses to a particular test item
4	Has complex theoretical models hence difficult or complex to apply. Is more theory grounded	Has simple theory assumptions, easy to apply in many testing situations
5	Strictly computer based	Could be applied manually and also electronically
6	Is robust in providing objectivity in trait measurement	Has an inherent inability to sustain specific objectivity in trait measurement
7	This is taken care of by sample invariance properly, as (items can be calibrated by a process of test of fit of the model.	Item analysis is sample dependent.
8	Has inherent quality to detect item bias or differential item functioning	Does not perform this function.

9	Tailored testing is achievable and operational as items Administered are the exact item that suit the examinees ability.	Tailored testing is not achievable as the same sets of text items are given to all examinees.
10	Item banking is easier and items are easily tested for goodness of fit	Item banking is difficult and achievable mainly manually.
11	Features on item-level information	Focuses on test level information
12	the primary interest is not on the raw score but on whether an examinee got each individual item correct or not	An examinee's raw score is the sum of the scores received on the items in the test.
13	Conceptualizes the probability of an individual responding to any particular item. It is therefore useful for describing test items in an item bank	Item statistics are dependent on the particular group used in the item calibration process.
14	Ideal for high stake tests, tests for relatively large population such as those conducted by public examining bodies like WAEC, JAMB, NECO etc.	Could be used for both high stake and school-based examinations.

These benefits include:

1. Item statistics that are independent of the groups from which they were estimated.
2. Scores describing examinee proficiency that are not dependent on test difficulty.
3. Test models that provide a basis for matching test items to ability levels.
4. Test models that do not require strict parallel tests for assessing reliability.

There are three basic assumptions or principles underlying the item response theory. These assumptions include:

1. Dimensionality of the latent space
2. Local independence
3. Item characteristics curve

The dimensionality of space enumerates that the dimension and direction of latent ability is imperative and it accounts for the examinees test performance. This simply implies that a single latent ability is adequate in explaining the performance of a testee. In a situation where the above statement occurs,, it is known as unidimensionality of the latent space while those that assumes more than one traits are called multidimensionality.

The Local independence assumes that the probability of a testee to provide the right answer for a particular test item is not affected by his/her performance in other test item on the same test. The item characteristics curve shows the relationship between the true score and the ability score. It proposes that the probability of success on an item is related to the ability measured by the set of the item.

Cheon, Lee, Crooks and Song (2012) has explained that the item response theory that can be applied in measuring unobservable traits like interest, also measuring growth or developmental level score, in detection of bias, tailoring testing, estimating power scores, equating test scores and for item banking for test development Item Response Theory has many models which include;

1 Dichotomous models that comprises of

- one parameter logistic model (1 -PL.)
- two parameters logistic model (2-PL.)
- three parameters logistic model (3-PL)

2 Polytomous models which comprises of

1. The Partial Credit Model (PCM)
2. Generalized Partial Credit Model (GPCM)
3. the Rating Scale Model (RSM)
4. the Nominal Response Model (NRM)
5. Graded Response Model (GRM)

One Parameter Logistic Model (1-PL)

This model of the item response theory employs the **b** parameter assumes that test item discriminates equally without guessing, although varies with item difficulties. Mathematically, the one parameter logistic model stated that the probability of an individual having ability level denoted with O respond correctly to an item i is given as;

$$P_1(O) = \frac{1}{1 + e^{-ai(O-b_i)}} \quad \text{Equ. 1}$$

Where $P_1(O)$ is the probability of the testee

O = ability level

i = item b = difficult parameter

Two Parameter Logistic Model (2-PL)

The two parameter logistic model uses 2 parameters which are a and b . This implies that the item varies in difficulties and discrimination without guessing. The mathematical equation is given as;

$$P_1(O) = \frac{1}{1 + e^{-ai(O-b_i)}} \quad \text{Equ. 2}$$

Three Parameter Logistic Model (3-PL)

The three parameter logistics maintains that the low ability testee will provide an answer to a difficult item by guessing. The mathematical model is given as;

$$P_1(O) = c_i / (1 + e^{-a_i(O-b_i)}) \quad \text{Equ. 3}$$

Progar and. Socan (2008) in a study conducted in six European countries (Hungary, Latvia, Netherlands, Norway, Scotland, and Slovenia) titled “An empirical comparison of Item Response Theory and Classical Test Theory”. Based on the outcome of the study, it was revealed that IRT person parameters are more invariant across different item sets, and that the CTT item parameters are at least as much invariant in different item sets as the IRT item parameters. The results furthermore demonstrate that, with regards to the invariance property, IRT item/person parameters are in general empirically superior to CTT parameters, but only if the appropriate “1W!” model is used for modeling the data.

Crocker and Algina (2008) investigated an empirical study titled “Introduction of Item Response Theory (IRT) Models in the Development and Validation of College Mathematics in Attaining Quality Education for National Values” using a sample size of 1000 students through multistage random sampling technique. The data were analyzed using Factor Analysis using the principle component analysis (PCA) and rotated component matrix (RCM) and Chi-square goodness of fit of the MULTI LOG. The findings of the study revealed that IRT models should be adopted in test construction and validation over CTT, since CTT has varieties of shortcomings.

Iweka and Abbott (2017) conducted a study on attitudes of teachers towards application of item response theory in technical colleges in Rivers State. Using a stratified random sampling technique to draw the sample of size of 212 technical college instructors were selected for the study. The instrument for data collection was titled “Survey of Teacher’s Attitudes Towards Application of Item Response Theory” (STATA-IRT). The findings of the study showed that strong linear relationship exists between attitudes of teachers and the application of Item Response Theory in Technical Colleges Rivers State. It was also found from the multiple regression results that the attitudes of teachers are predictors of effective application of item response theory in (the technical colleges in Rivers State).

Awopeju and Afolabi (2016) sampled 6000 students who sat for senior secondary certificate examination in Osun State out of a population of 35,262. Using the ex-post-facto research design and a mathematics achievement test comprising of 60 questions with 4 options each, the result revealed that both CTT and IRT were comparable in estimating item characteristics of statistical and psychometric test and therefore should be used for the development of national examinations.

Relevance of Item Response Theory on Learning

Item Response Theory and related models are important to the practice of educational and psychological measurement because they provide a framework for considering issues and addressing technical problems. One of the most important issues is the handling of measurement errors. It can help in understanding the role that measurement errors play in:

1. Estimating examinee ability and how the contributions of error might be minimized (e.g., lengthening a test).
2. Correlations between variables
 - Reporting true scores or ability scores and associated confidence bands.
1. Flexibility afforded by the model. For example, different sets of items could be administered to individual test-takers and yet comparable estimates can be estimated from these different tests.
2. Representing the ability of the test-takers and the difficulty of the items as independent parameters.

Allen (2007) posited that a good Item Response Theory can also provide a frame of reference' for doing test design work or solving other practical problems. He further explained that if specific the precise relationships among test items and ability scores so that careful test design work can be done to produce desired test score distributions and errors of the size that can be tolerated.

Recommendations

- There should be increased awareness of Item Response Theory in assessment in tertiary institutions in Rivers State.
- Item Response Theory should be adopted in test development, validation and standardization.
- Item Response Theory should be utilized in the interpreting the performance of undergraduates.
- Lecturers should ensure that test or examinations should be appropriate to the level of the undergraduates in tertiary institutions in the state.
- The performance of undergraduates should be transformed to standard scores and then interpreted.
- There should be increased computer literacy in order to make IRT effective in the universities.
- There should be increased knowledge of the statistical sophistication inherent in IRT.

Conclusion

Item Response Theory is very important in interpreting the performance of learners. It is necessary 'that lecturers should be aware of its relevance in order to give appropriate interpretation of learners' performance.

References

1. Adedoyin, O.O. & Mokibi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test item. *International Journal of Asian Social Science*, 3 (4), 992-1011.
2. Allen, D.D. (2007). Validity and reliability of the movement ability measure: A self- report instrument proposed for assessing movement across diagnoses and ability levels. *Journal of Pin's Ther*, 8(7), 899-916.
3. Anene, G.U & Ndubuisi, O.G. (2003). Test development process. In B.G Nworgu (Ed). Educational measurement and evaluation. Theory and practice. Nsukka: University Trust Publishers.
4. Asuru, V. A. (2015). *Measurement and evaluation in education and psychology*. Pearl Publishers.
5. Awopeju, O.A. & Afolabi, E.R. (2016). Comparative analysis of CTT and IRT based item parameter estimate of senior school certificate mathematics examination. *European Scientific Journal*, 12 (28), 1857-1 881.
6. Ayala, R.J. (2009). *The theory and practice of item response theory*. Guide ford Press.
7. Cheon, J, Lee, S., Crooks, S. M., & Song, J. (2012). An investigation of mobile learning readiness in higher education based on the theory of planned behavior. *Journal of Computers & Education*, 59(3), 1054 -1064.
8. Christensen, R. (2002). Effects of technology integration on the attitude of teachers and students.
9. *Journal of Research on Technology in Education*, 3(4), 411 -433.
10. Crocker, I: & Algina. J. (2008). Introduction to classical and modern lest theory.. Brace Jovanovich press.
11. Haiyang, S, (2010). Application of classical test theory and many facet Rasch measurement in analyzing; the reliability of an English test for Non-English major graduate. *Chinese in Journals Applied Linguistics*, 33(2), 12-24.
12. Hambleton, R. K. (1989). *Principles and selected applications of item response theory*. In R. L. Linn (Ed.), Educational Measurement: Macmillan.

13. Hogg, M., & Vaughan, G. (2005). *Social psychology*. Prentice-Hall.
14. Iweka, F.O.E & Abbott, T.W. (2017). Attitudes of teachers towards application of item response • theory in technical colleges in Rivers State. *British Journal of Education*, 5(6) 39-56.
15. Joshua, M.T & Ikiroma, B. (2012). Differences between classical test theory and then, response theory via derived test data. *Journal of Educational Assessment in Africa*, 7(4), 210-217.
16. Maio, G. & Haddock G. (2010). *The psychology attitude and attitude change*. SAG 1' Publications .
17. Ocho, L. O. (2005). *Issues and concerns in education and life institute for development studies*. University of Nigeria Enugu campus Press.