



RELEVANCE OF CLASSICAL TEST THEORY IN THE ASSESSMENT OF LEARNING IN TERTIARY INSTITUTIONS IN NIGERIA

Ebere Sampson Wagbara (Ph.D.)¹ and Nbame Letam Nna-Kue (Ph.D.)²

^{1&2}Faculty of Education, Department of Educational Foundations, Rivers State University, Nigeria.

Cite this article:

Ebere S. W., Nbame L. N. (2024). Relevance of Classical Test Theory in the Assessment of Learning in Tertiary Institutions in Nigeria. *British Journal of Education, Learning and Development Psychology* 7(2), 202-213. DOI: 10.52589/BJELDP-IQRLMLZZ

Manuscript History

Received: 13 Mar 2024

Accepted: 30 May 2024

Published: 01 Jul 2024

Copyright © 2024 The Author(s).

This is an Open Access article distributed under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), which permits anyone to share, use, reproduce and redistribute in any medium, provided the original author and source are credited.

ABSTRACT: *Over the years, measurement experts have been captivated by the description of students, which has resulted in the development of test theories such as Item Response Theory and Classical Test Theory. The traditional method of item analysis, known as "Classical Test Theory," asserts that an individual's observed score on an exam is equal to their true score and an error score, with all items in the test contributing equally to student performance. Assessment, in this context, refers to any method used to gauge a learner's current knowledge. The significance of Classical Test Theory in teaching, learning, and evaluating learning outcomes has spurred academic inquiry. This paper explored the application of Classical Test Theory in tertiary institution assessment, emphasizing its relevance in evaluating learning outcomes. Some notable points included the simplicity of mathematical procedures in classical test analysis and the straightforwardness of model parameter estimation. Additionally, this paper advocated for the utilization of statistical sophistication inherent in Classical Test Theory to interpret undergraduates' performance effectively. Lecturers were encouraged to familiarize themselves with its application to provide meaningful insights into students' performance.*

KEYWORDS: Assessment, Classical Test Theory, Education, Item Response Theory, Tertiary Institution.



INTRODUCTION

Education is the means through which an individual realizes their potential and utilizes them for their wellbeing and also contributes their quota to nation building. According to Ocho (2005), acquiring knowledge and skills, as well as developing one's character, is the main goal of education. However, this cannot be comprehended by simply weighing it on a regular scale. Rather, it must be understood using the mathematical tools of measurement and assessment; which are essential to the accreditation of all educational programmes. As opined by Asuru (2015), measurement is the systematic process of determining the characteristics or behaviour of an individual and reporting the same quantitatively. It is the process of using numbers or quantities (category system) to describe a person's educational or psychological behaviour or characteristics. However, the category systems that lend themselves to quantification may be used. One possible approach is to employ a classification system with terms like A, B, C, D, or excellent, good, poor, and so forth. Stated differently, measurement is the methodical process of employing numbers or numerals (or category system) to ascertain the degree to which an attribute is present in an individual to describe the attribute. Going further, Asuru (2015), stated that measurement is the act of determining the degree to which an individual possesses a certain attribute. Being systematic implies that both the actual measurement and its description must conform to certain acceptable standards. For example, a lecturer may want to find out how much ordinary level mathematics a student has learned, how much skill he has acquired in playing the keyboard, how much interest a prospective student has in schooling, his level of motivation, attitude towards teaching as a profession, attention span, and the likes. There is no precise instrument the teacher will place on the student's head or heart to determine the percentage of ordinary level mathematics he knows or his attitude, interest or attention span. The lecturer may use tests to obtain numbers such as 20%, 60% or category systems like A, B, or excellent, poor, and so on to indicate the extent of the behaviour. Generally, measurement is quantitative in nature and deals with status determination; hence, it answers questions such as: How much? How many? Estimation possesses a focal spot in the school program as it decides the degree to which instructive targets are accomplished. One more idea firmly connected with estimation is evaluation.

The term "assessment" refers to the wide range of guidelines, procedures, or tools that educators employ to gauge, assess and document students' learning progress, experience growth, mastery of material, and information needs (Kuldos, 2013). Assessment is an undertaking to know an individual; it is seen as something that occurs whenever an individual is related with one kind of coordinated effort clearly or indirectly with another person or individuals. Assessment as a methodology used to all the more promptly understand the continuous data that an understudy has (Ogidi and Wagbara, 2017). It might be basically all around as direct as a teacher's personal judgment considering the view of the understudy's presentation. As noted in Wagbara (2020), cognizant approach describes expected understudy achievement using pre-chosen inside standards and external benchmarks, all centred around unrivalled learning.

Evaluation is without uncertainty a critical part of schooling. It is utilized to give significant criticism on an individual's or gatherings' comprehension of examples, acknowledgment of showing strategies embraced by the instructor and others (Kocdar, Karadag & Sahin, 2016). One of the techniques for acquiring such input is by means of testing. In educating and learning, evaluation is unmistakable on the grounds that it gives relevant data concerning the



distinguishing proof of students regardless of exceptional advancing necessities, reason for reviewing, observing educational viability and arrangement of students' accomplishment. As affirmed by Obilor (2019), evaluation is an indispensable piece of homeroom guidance that improves, engages, and loves understudies learning. Utilizing an assortment of evaluation methods, instructors can accumulate data about what understudies know and can do, and give positive, steady criticism to understudies. The data is utilized to analyse individual necessities of understudies and work on their educational program, which thus assists understudies with learning all the more, particularly in tertiary organizations. He further expressed that evaluation is a connected part of a mental test.

A mental exam consists of an objective or a normalised percentage of a behaviour sample. According to Asuru (2015), psychometrics is the technical term for the science underlying mental testing. A test is a systematic approach to observing a person's behaviour and representing it using a class framework or a mathematical scale. As opined by Obilor (2019), testing could likewise include a more thorough evaluation of a person. It is a course of controlling the instruments to students or members; such instruments estimate the degree of greatness, capacity, expertise, dominance, information, interest or disposition, among others, that gives significant information to evaluation.

It is worth noting that testing, when applied, can reveal certain hidden mental abilities or skills about individuals that would not have been widely known. Since the test assumes a fundamental part in the instructive biological system, it is, in this manner, basic that test advancement particularly in the tertiary establishment of learning ought to be of elevated expectation that should be satisfactory, compelling and proficient in finding out a student's presentation.

As stated by Anene and Ndubuisi (2003), schools, particularly tertiary establishments, give significant level HR to driving the economy and guaranteeing fast change. It is a middle where educators are prepped and ready to confront the universe of work, particularly in the instructing venture. All in all, tertiary organization is a panacea to the ills of society and how much a country advances is a component of the nature of the instructive fulfilment of its residents. It is in this basic way that test advancement and development that conform to global best norms are followed to create graduates that can work ideally in the realm of work. Besides, investigating test papers from tertiary establishments in Nigeria, it will be seen that some test items developed by teachers particularly those without foundation in schooling come up short on psychometric properties. This highlights the need to look at the pertinence of Classical test theory (CTT) in the evaluation of learning in tertiary organizations in Nigeria.

Test engineers are fundamentally worried about the nature of test things and how examinees answer it while building tests. Psychometric hypothesis offers two methodologies in breaking down test information: Classical test theory (CTT) and Item Response Theory (IRT). The two hypotheses empower one to anticipate results of mental tests by distinguishing boundaries of thing trouble and thing segregation.

According to Asuru (2015), item difficulty as the percentage of persons who answer an item correctly. Item difficulty is simply referred to as the p value, the proportion of examinees who answer an item correctly. Item p-values ranges from 0.0 to 1.00. The easier an item is for an examine group, the higher its p-value. The harder an item is, the lower is its p-value. In test construction, it is a practice of arranging items in order of their difficulty. This procedure gives



confidence to individuals approaching the test and reduces time wasted on items that are beyond an individual's ability. This method gives certainty to people moving toward the test and lessens time squandered on things that are past a singular's capacity.

For most extreme data on a test, it is great to choose things with a moderate spread of trouble, however, whose typical trouble is 0.50. Assuming that a thing is very hard or very simple for a gathering of examinees, there will be little variety in the scores for that thing. The nearer the trouble of a thing approaches 1.00 or 0, the less differential data about the examinees. The nearer the trouble level is to 0.50, the more separation the thing makes. Thing trouble may likewise be communicated in things of Z-scores. Maier (2001) expressed that Z-scores are relegated to thing trouble by looking into the Z-scores in a standard typical Z table related with the thing's p-esteem. For instance, assuming a thing's p-esteem is 0.46, the related z-score is the point that separates the scale under the ordinary bend so that 54% of the region under the bend tumbles to its left and 46% tumbles to the right. Things with p-esteems under 50 have positive z-scores and things with p-upside of more than 50 have negative z-scores. Since thing trouble communicated as far as expected bend units include negative qualities and decimals, they are typically changed over completely to a more reasonable scale, as utilized by the Educational Testing Service (ETS) in its test improvement; this change is called Delta Score. Thing trouble and thing segregation are the two factual techniques utilized in thing examination.

As noted in Hambelton (2000), Item discrimination is the ability of an item to differentiate among students on the basis of how well they know the subject matter they are being tested on. It is the degree to which item scores are indicative of an internal criterion (i.e. total score) or an external criterion (i.e. examinee age). The index of discrimination is computed by identifying the number of persons passing each item in U (upper) and L (Lower) orientation groups. Typically, the upper 27% and the lower 27% are chosen for maximum efficiency. The values of P_u and P_L are biserial correlation. The index of discrimination is expressed as $D = P_u - P_L$. He further states that if the D-values are greater than or equal to 40, the items are quite effective; D-values of 30 to 39 are satisfactory; D-values between 20 and 29 indicate that revision may be warranted and D-values of less than 19 have cause for revision or elimination.

In addition to the aforementioned methods for calculating the Index of Discrimination, item discrimination can also be expressed as the Pearson's Product Moment correlation, which calculates the degree of relationship between item score and total score. Psychometricians, however, favour biserial correlation over point biserial correlation due to its lack of correlation with item difficulty and its ability to accommodate examinees' often varying degrees of incomplete knowledge that was used to derive the point biserial (Crocker & Aigina 2008). On the other hand, the IRT is model-based and describes item performance at every student ability level.



Application of Classical Test Theory in Assessment in Tertiary Institution

Classical test theory (CTT) has been the establishment for estimation hypothesis for more than 80 years. The reasonable establishment, presumption and expansion of the essential premises of CTT have considered the improvement of a few fantastic psychometric social scales.

Awopeju and Afalabi (2016) made sense of that Old Style Test Hypothesis expects that an examinee test score is the composite of a genuine score and irregular blunder, that all things in an evaluation instrument make an equivalent commitment to the presentation of understudies. They further expressed that CTT just thinks about an assortment of examinees and decides their prosperity rate on a thing. CTT measurements include thing trouble and the typical p-esteems and conjures no complex hypothetical model to relate an examinee capacity to outcome in a specific thing. All things considered, it thinks about a pool of examinees and exactly inspects their prosperity rate on a thing.

It is essential to improve standard test items for the overhaul of academic data and capacities of understudies in the tertiary groundwork of learning and, as a result, making quality graduated class that can compete in the overall labour force. As seen in Prieler (2007), old style test speculation essentially revolved around test level information. CTT takes into understanding the response of each and every examinee of a given limit as it relates to the test. Utilization of CTT is stressed over the saw grade (x), real score (T) and inconsistent bungle (E). The saw score is the suggested genuine score in a test, however, the real score is the certified limit. It is the score an examinee would have achieved if there were certainly no botches in the assessment. On the other hand, the mix-up is the irregularity among saw and veritable scores—how much botch is found in a test or measure that ought to be a mistake in the test (Public Board of Trustees on Assessment in Preparing, 2017). This subsequently infers that the error distorts the night out of the certified score and saw score. Right when the bumble is killed, an individual's veritable score and saw score will be identical when assessed on and on. Conventional grade is about trustworthiness; it implies how unsurprising a test or measure is.

Unwavering quality is hypothetically characterized as the proportion of the fluctuation of the genuine score to the proportion of the noticed score. Mathematically, it is expressed by Booker (2018) thus:

$$P2 \text{ } xx = \frac{02 \text{ } t}{\sigma x2}$$

This implies that reliability tells the extent to which the observed score variance is close to true score variance. A perfect reliable test is one with zero error score, in which the observed score and true score are equal. The reliability of the test is +1. As the error increases, the reliability reduces. The equation for this process is as follows:

$$\text{VAR} (x) = \text{VAR} (T) + \text{VAR} (E).$$

Given this, it can be shown that the variance of the observed scores $\text{VAR} (x)$, that is due to true score variance $\text{VAR} (t)$, provides the reliability index of the test.

$$\text{VAR} (t) / \text{VAR} (X) = R$$

At the point when the change of genuine scores is high comparative with the difference of the noticed scores, the unwavering quality (R) of the action will be high (e.g., $60/70 = 0.85$), while



assuming the fluctuation of genuine scores is low comparative with the fluctuation of the noticed scores, the dependability (R) of the action will be low (e.g., $30/70 = 0.42$). But the idea of unwavering quality is subject to the gathering used to foster the test. In the event that the gathering has an extensive variety of expertise or capacities, the unwavering quality will be higher than if the gathering has a limited scope of expertise or capacities. Hence, unwavering quality is not invariant with regard to the example of the test-takers, and is consequently not a quality of the actual test. Unwavering quality qualities range from 0.00 to 1.00. Revamping the terms from the above conditions, it tends to be shown that:

$$R = 1 - [\text{VAR} (E)/\text{VAR} (X)]$$

That is, unwavering quality is equivalent to 1 - the proportion of arbitrary mistake fluctuation to add up to score difference. Further, there are investigations that consider an assessment of R (dependability), and, obviously, working out the noticed difference of a bunch of scores is a straightforward process. Since R and VAR (X) can be determined, VAR (T) can be displayed for with the accompanying condition:

$$\text{VAR} (T) = \text{VAR} (X)$$

It merits emphasizing here that CTTs are generally keen on displaying the irregular blunder part of a crude score. Some mistakes are not arbitrary; they are precise. Much time and exertion has been spent to distinguish and manage deliberate blunder with regards to test legitimacy.

In Osun State, which has a population of 13, 262, Awopeju and Afolabi (2016), provided an example of 6,000 understudies who took the senior optional endorsement assessment. Using the ex-post facto research plan and a maths accomplishment test with 60 questions and four possible answers, the results showed that the IRT and CTT were nearly equal in evaluating the thing qualities of measurable and psychometric tests, and that they should therefore be used to enhance public assessments.

Cheon, Law breakers and Tune (2012) completed an observational review named "old style test hypothesis examination of the elements idea stock." Various relapse investigations were led for information gathered with two instruments utilizing old style test hypotheses. These examinations gave understanding into Dynamic Idea Inventories (DCI's) reasonable setting, estimation properties, and relative legitimacy given its planned use. By and large, proof showed that the instrument was appropriate for low stakes developmental evaluation use, but may have restrictions for high stakes involved in its ongoing structure.

Adedoyin and Mokihi (2013) led a concentrate on two methodologies in psychometric cycle CCT and IRT. In the improvement of estimation hypothesis, there are two principal measurable ways to deal with depict qualities of an individual and to break down capacities and idle credits of a subject, which are CTT and IRT. This study gave data about the fundamental properties of the two speculations, decided a psychometric course of estimation with subtleties, looked at models of hypotheses and clearly communicated the benefits and hindrances of the two speculations.

Haiyang (2010) surveyed Traditional Test Hypothesis: A Prologue to direct displaying ways to deal with test and thing investigation. The primary motivation behind the review was to give an exhaustive outline of the CTT and its methodology as applied to test thing improvement and investigation. It was a logical structure which had a trailblazer job in instructive estimation and



psychometric cycle. CTT has served the estimation local area for quite a long time other than portraying the straightforwardness of the CTT model according to numerous perspectives; different impediments are point by point in thing and an individual capacity level. In spite of the deficiencies credited to CTT, it was suggested that the CTT approach of thing examination ought to be kept up with in test advancement and assessment, in light of its benefit and straightforwardness in examination of unwavering quality and in limiting estimation mistakes. CTT enjoys an upper hand over Item Response Theory (IRT) as a result of its hypothetical suppositions, making it simple to apply many testing circumstances (Hogg & Vaughan, 2005).

Item Response Theory (IRT) is an illustration of an estimation hypothesis that has the qualities of information procurement at the proportion scale level, example free characteristics and students' capacity that can be accounted for on both thing and all out instrument levels. According to Ayala (2009), the limitations of the CTT contributed to the development of IRT. Asuru (2015) acknowledged the aforementioned and maintained that IRT is an improvement over the CTT, which tackles a fundamental change in the plan, information analysis, and test instrument scoring. Additionally, Joshua and Ikiroma (2012) referred to IRT as a "solid genuine score." Mental Test Hypothesis and Dormant Quality Model: Dormant quality model attempts to sort out whether or not there is a fundamental characteristic that records for an examinee test execution.

Item Response Theory (IRT) is about a guinea pig execution on a test and how it connects with the singular things on a test; essentially, it zeroed in on the lab rat capacities and demeanor utilizing numerical models. The utilization of test speculations, particularly the Item Response Theory (IRT), is unquestionably crucial for enhancing standard test items aimed at enhancing the academic knowledge and skills of tertiary-level students, thereby producing high-quality graduates capable of competing in the global workforce. The Item Response Theory, also known as the Item Attribute Hypothesis, is a paradigm focused on the design, analysis, and scoring of test items and similar instruments used to measure achievement, abilities, attitudes, and other essential factors. This theory is rooted in mathematical models for testing knowledge (Hogg & Vaughan, 2005). Unlike the CCT, which focuses on the level of the test, the Item Response Theory concentrates on the test items, implying that the theory considers the response of each examinee to a given item in relation to the test.

The Item Response Theory (IRT), also known as the Dormant Quality Hypothesis, the Solid Genuine Score Hypothesis, or the Current Mental Test Hypothesis, is a theory used in psychometrics that guides the design, analysis, and scoring of assessments, questionnaires, and other similar instruments that estimate abilities, viewpoints, or other variables. Information testing relies on the application of relevant numerical models (Kuldas, 2013). It is generally thought to be superior to conventional test hypotheses and works well for high-stakes exams, like the Joint Admission and Matriculation Board (JAMB).

Use of Classical Test Theory in Assessment in Tertiary Institutions

Instructors should be familiar with the strategy of development to empower them to build legitimate, solid and usable tests. This is on the grounds that an inadequately built test could be negative not exclusively to the schooling system, but also to the general public overall. In the Methodology, Classical Test Theory (CTT) was engaged, as talked about by Asuru (2015).



Test planning, item creation, trial testing, and item analysis constitute assessment practices. Assessment encompasses a wide array of methodologies or techniques utilized to evaluate, gauge, and report students' academic readiness, learning progress, skill acquisition, or educational needs (Ogidi & Wagbara, 2017). They further assert that assessment serves as a means to better understand the ongoing knowledge that a student possesses; it is a pathway to acquiring valuable knowledge, skills, and values that ensure harmonious and productive living in society. The outcome of any tertiary educational institution should reflect in its members. Based on the aforementioned, the education of students, with a focus on enhancing cognitive development, fostering deeper intellectual abilities, acquiring technical skills, and fostering ethical character development encompassing values, respect, appreciation, and a sense of responsibility for the advancement of a sustainable nation, lie at the heart of teaching and learning in tertiary institutions.

As opined by Haiyang (2010), the objectives of tertiary institutions as follows: (i) Instilling and nurturing appropriate values and a positive outlook for individual survival in society (ii) Cultivating deep intellectual capacities in individuals to comprehend the dynamics of the environment and consequently appreciate the intrinsic value of the environment (iii) Acquiring both intellectual and physical skills necessary for functional living and evolving into responsible members of society (iv) Developing an objective perception of the local and global environment.

As seen in Maio and Haddock (2010), teachers in tertiary establishments are key impetus in accomplishing the required outcome for public financial development of a country. Tertiary organizations in Nigeria incorporate monotechnics, polytechnics, schools of training, schools of innovation, schools of nursing and colleges, among others. The three cardinal focuses through local area administration examination and showing should be fulfilled in compatibility with higher establishment of figuring out how to discover quality schooling for students. As noted in Iweka and Abbott (2017), the commitment of tertiary establishments to local improvement is a subject which has drawn in expanding consideration as of late. Showing in tertiary establishments includes diverse cycles that require the participators (speakers) to be proficient in the space of estimation and evaluation expertise that are important in arranging and development of tests, reviewing of grades, understanding of experimental outcomes utilization of appraisal results to illuminate educating and learning, translation of state administered tests and conveying results to pertinent partners. In any case, Dedicate (2002) featured that for one to see himself or herself as a powerful educator, information on instructive appraisal techniques is principal.

Relevance of Classical Test Theory in Learning

A few advantages are realistic through the utilization of objective informative goal and thing composing utilizing old style test examination. The most well-known approaches utilized by psychometrics are Classical Test Theory (CTT) and Item Response Theory (IRT), and both can be utilized to fabricate a decent estimating instrument. Traditional test hypothesis is frequently utilized in the homeroom because of its relative effortlessness and the lower level of ability expected to embrace examination. Above all, CCT must be utilized to investigate the exhibition of one gathering of understudies on one appraisal instrument. It concentrates on just the capacity of interest, any remaining wellsprings of the variety, for example, outside variables of the encompassing or physical and psychological circumstances of the examinee are consistent



all through rehashed normalization method. CTT is broadly utilized as it addresses a more conservative and functional method for producing insights, particularly when the evaluation included does not convey as much weight. The utilization of Traditional Test Hypothesis assists with deciding the nature of test things in school to guarantee understudies' exhibition are appropriately estimated. It gives recurrence to every one of the reactions to examine distractors for their usefulness and the connection among higher and lower accomplishing gatherings of understudies. This would assist speakers with enhancing their instruction and getting legitimate criticism on regions they need to adapt. A decent test will actually want to give quality criticism on the expected development, and to decide if the things used to construct are excellent; they should be examined as far as their trouble and how well they can recognize or separate between the understudies (Marcus & Joseph, 2014)

Joshua and Kiroma (2012) outlined the significance of Classical Test Theory in learning as follows:

1. Classical test analysis utilizes relatively simple mathematical procedures, and model parameter estimation is conceptually straightforward.
2. Classical test analysis is often referred to as a weak model because the assumptions are easily met by traditional testing procedures, making it easy to apply in many testing situations.
3. Compared to item response theory models, analysis can be conducted with smaller representative samples of examinees, which is important when field testing a measuring instrument.
4. It does not require a complex theoretical model to relate an examinee's ability to success or a particular item.
5. It collectively considers a pool of examinees and empirically examines their success rate in an item for dichotomous score items.
6. An examinee's raw score is the sum of the scores received in the test.
7. Classical Test Theory can be used for both high-stakes and school-based examinations.
8. It explains and addresses errors to enhance the reliability of the test.
9. It is based on true scores and thus depends on the examinee's aggregate score in the test.

Moreover, educating, learning, and evaluation are key ideas in training and their relationship should be visible as members of a three-legged race. In such a manner, homeroom evaluation practices, for example, educator-made tests are significant when they support understudies' learning. After regulating a test, an educator/teacher should have the option to decide test things' quality, and whether the things mirrored the exhibition of the understudies concerning those specific learning goals shown over a period (Odili, 2010).

According to Haladyna (2004), 50 percent or a greater amount of test things educators and test engineers write in tertiary foundations neglect to proceed as expected. So thing examination should be acted on to get helpful information for question improvement, and can be incorporated into the test creation and audit process. The utilization of traditional test



hypothesis assists with inspecting the degree of troublesomeness, segregation files, and distracter working of arithmetic numerous decision tests planned by teachers. The trouble and the separation lists of a thing change with the example of the guinea pig that passes through the examination; likewise the capacity of the lab rat fluctuates with the example of the test. Accordingly, the lists will rely upon the examples of the guinea pigs and test things utilized in getting information.

CONCLUSION

In summary, this paper emphasizes the relevance of Classical Test Theory (CTT) in assessing learning in Nigerian tertiary institutions. Despite the emergence of alternative methodologies, CTT remains valuable for constructing tests. The study aims to deepen understanding of CTT and advocates for its continued use in evaluating students' performance. CTT primarily focuses on test-level information, which is crucial for developing standardized test items that enhance intellectual knowledge and skills among learners in tertiary education, thereby producing competitive graduates for the global workforce.

Despite its limitations, CTT's approach to item analysis should be upheld in test development and evaluation due to its advantages and simplicity in assessing reliability and minimizing measurement error. CTT plays a vital role in interpreting students' performance, underscoring the importance for lecturers to be familiar with its application to provide accurate interpretations of student performance.

SUGGESTIONS

1. Classical test theory should be adopted in test development, validation and standardization in Nigerian tertiary institutions.
2. Classical test theory needs to be developed through higher level educational studies and practice in Nigerian tertiary institutions.
3. Classical test theory should be used by the lecturers in the interpreting of performance of undergraduate students in Nigerian tertiary institutions.
4. Knowledge of the statistical sophistication inherent in CTT should be increased in Nigerian tertiary institutions.



REFERENCES

- Adedoyin, O. O. & Mokihi, 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.
- Anene, G.U & Ndubuisi, O., G. (2003). Test development process. In B.G Nworgu (Ed). *Educational measurement and evaluation, theory and practice*, Sukka; University Trust Publishers.
- Asuru, V. A. (2015). *Measurement and evaluation in education and psychology*: Minsim Press.
- Awopeju, O. A. & Afolabi, E. R. (2016). Comparative analysis of CTT and IRT based item comparative estimate of senior secondary school certificate mathematics examination. *European Scientist Journal*, 12 (28), 1857-1881.
- Ayala, R. J. (2009). *Theory and practice of item response theory*. Guide ford Press.
- Booker, J. (2018). Statistical analysis of reliability data. *Journal of quality technology* 25 (3), 227-228.
- Cheon, J. IEE, S., Crooks, S. M., & Song, J. (2012). An investigation of mobile learning readiness in higher education based on the theory of planned behaviour. *Computer & Education*, 59(3), 1054-i064.
- Christensen, R. J. (2002.) Effect of technology integration on the attitude of teachers and students. *Journal of research on Technology in Education*. 3(4), 411-433.
- Croker, I. & Algina, J. (2008). Introduction to classical and modern test theory. Fort Worth Harcourt: Brace Jovanovich Press.
- Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. Routledge. <https://doi.org/10.4324/9780203850381>
- Haiyang, S. (2010). Application of classical test theory and many facet Rach measurement in analyzing the reliability of an English test for Non-English major graduates. *Chinese Journal Applied Linguistics* 33 (2), 12-24
- Hambleton. R. K. (2000). Emergence of item response modelling in instrument development and data analysis medical care. Retrieved from www.jstor.org
- Hogg, M. & Vaughan, G. (2005). *Social psychology*. Prentice-Hall.
- Iweka, F. O. E. & Abbott, T. N. (2017). Attitudes of teachers towards application of item response theory in tedimeal changes in Rivers State. *British Journal of Education*, 5(6), 39-56.
- Joshua, M. T. & Ikiroma, B. (2012). Differences between classical test theory and item response theory via derived test data. *Journal of Educational Assessment in African*, 7 (4), 210-217
- Kocdar, S., Karadog, N, & Sahin, H. T. (2016). Analysis of difficulty and discrimination indices of multiple-choice questions according to cognitive level in an open and distance learning context. *The Turkish Online Journal of Educational Technology*, 1 (5), 16-24
- Kuldass, S. (2013), Unconscious learning process; Mental integration of verbal and pictorial instructional materials. Retrieved on 29th September, 2018 from www.nebi.nih.gov/pmc/article.
- Maier, K. S. (2001). A rash hierarchy measurement model. *Journal of Educational and Behavioural Statistics*, 2 (6), 307-331.
- Maio, G. & Haddock, U. (2010). *The psycholK. Sogy attitude and attitude chary*. Sagi Publication.



-
- Marcus, A. C. & Joseph, E. A. (2014). Science teachers and continuous assessment implementation in secondary school: Competence and effects. *Journal of Research and Method in Education*, 4 (4), 36-41.
- National Council and Measurement Education (2017). *Classical of important assessment and measurement tonus*.
- Odili, J. N. (2010). Effect of language manipulation on differential item functioning of test items in Biology in a multicultural setting. *Journal of Educational assessment in Africa*. 4-268.
- Obilor, I. E. (2019). *Essentials of measurement and evaluation*. Sabars Printers and Publishers.
- Ocho, I. O. (2005). *Issues and concerns in education and life institute for development studies university of Nigeria Enugu campus*.
- Ogidi, R. C. & Wagbara, S. E. (2017). Teachers' use of objective based assessment practices in secondary schools in rivers west senatorial district, Rivers State. *Nigeria Journal of Education In Developing Areas*, 25 (3), 381-390.
- Prieler, J. A. (2007). So wrong for so long: changing our approach to change. *the psychologist*, 20 (12), 730-732.
- Wagbara. S. E. (2020). Assessing teachers' competencies in the use of continuous assessment in rebuilding trust in students' academic achievement. *World Journal of Interactive Research* 3 (1), 115-125.