USE OF MULTIPLE MEASURES THROUGH ASSESSMENT AS A PANACEA FOR QUALITY SHAPING OF NIGERIAN LEARNERS

BY
Sofeme R. Jebson, PhD
Federal University of Kashere,
Gombe State, Nigeria
email: sofemejebson@yahoo.com

Abstract
Individual's and institution's activities are mostly purpose-driven that require decision-making. Quality decisions are dependent on assessment measures. To rely on a single measure is inviting trouble both in terms of how it will distort the identity of the learner because of its limited validity. Assessment plays a central role in Nigeria's educational practices especially in generating, analyzing and supplying information for decision-making. This paper focuses on the need to use multiple measures through quality assessments in decision-making for selection and placement of learners in Nigerian institutions. Qualities of measuring instrument and the theories employed in assessment are also addressed. Among others the paper recommends the use of multiple measures rather than single measure in selection or placement assessment. To this end valid assessment data based on modern test theories should be used by decision-makers for predictability or accountability purposes.

Key words: Multiple Measures, Assessment and Learners

Introduction
Education is an important instrument for development of any individual or nation in the world. Assessment is vital to any educational enterprise. Assessment is a term which includes all the processes and products which describe the nature and extent of student's learning. According to Federal Government of Nigeria (FGN, 2014) educational assessment and evaluation shall be liberalized by their being based in whole or in part on continuous assessment of the progress of the individual. This implies that in Nigeria being it for selection or placement purposes no single measure of assessment should be used in decision making process. Assessment plays a central role in Nigeria's educational practices especially in generating, analyzing and supplying information for decision making. Most summative assessment outcomes are competitive, or at least comparative between students, and it supplies a sort of seal of approval or disapproval on the student's efforts. Such an assessment exerts a potentially powerful effect on the student's self –esteem and growing sense of identity as a learner in a particular subject area (Satterly, 1981). To rely on a single measure is inviting trouble both in terms of how it will distort the identity of the learner because of its limited validity. The quality of any decision is determined by the validity of the information available to the decision maker. Quality educational decisions are test dependent on assessment measures. Faulty formulation and implementation of policy by parents, schools, societies and governments may result from invalid assessment data from the measuring instruments and also test theories employed in such assessment. According to Nenty (2001) the persisting problems in African education and the failure of policies and reforms to meet their desired aims and objectives tend to tell us that there are things we are not doing right and cast huge doubt on the validity of the information supplied by assessment to decision makers from classroom level to the level of national governments.

The paper, therefore, focuses on the measurement theories and of qualities of assessment instruments that underlies the generation of information for decision making in selecting or placing learners for their educational pursuits. Such quality assessment would shape and define the learners positively in terms of their self-esteem.

Measurement Theory Use for Assessment Measures in Decision Making

Fundamental to the problem of invalid scores from assessment exercise, are the nature of the theory that underlies our test construction, analysis and the estimation of the examinee and item parameters. The two main sources of information as input use for decisions involve in formulation and implementation of educational policies results from research and the one from assessment.

These two sources share something in common - that is the measurement of human latent characteristics. In the case of measurement in the physical world there are well-established measuring instruments and scales that provide us with useful information. Unlike measurement of psycho-social attributes is difficult which are often concepts or notions which lack clear definitions. Psycho-social measurement involves constructing measures (or scores and scales) from set observations (indicators). There is a need for a theory to guide us to get what we want from what we have. For Hambleton and Jones (1993), a test theory provides a general framework linking observable variables, such as test and item score to unobservable variables, such as true score and item score to unobservable variables, such as true score and ability scores. In Classical Test Theory (CTT) also called true score theory, the key methodology is about how to summarize (or aggregate) a set of data into a score to represent the measure on the latent trait. Based on CTT inferences on student ability, measures are made using test scores. In Item Response Theory (IRT), the methodology could involve a weighted sum score where different items have different weights when item scores are summed up to form the total test score. The weights may depend on the importance of the items. Alternatively, the item scores can be transformed scores using a mathematical function before they are added up. The transformed item scores may have better measurement properties than the raw scores. IRT has provision for summarizing a set of observed ordinal scored into a measure that has interval properties. Any test developed based on CTT takes on its own subjective but unique scale, on which performance on another test of same ability cannot be objectively mapped. While the IRT provides the basis for direct operationalization of our naïve understanding of what measurement is meant to be. Objective measurement requires a calibrated scale or measurement line with an equally spaced interval and units representing what is to be measured and on which both the item and person parameters could be objectively mapped (Jebson, 2004). The theory that underlies a test or an instrument development says a lot about its psychometric characteristics in terms of information generation for decision making.

CTT-based Reliability and Validity

Any test that gives consistent information reduces the level of uncertainty to decision makers. In statistics, information means reduction in uncertainty, and the analogous measure in CTT is reliability (Brannick,nd). Within the context of CTT, reliability is theoretically defined as the ratio of true score variance to observed score variance, that is, a measurement is reliable if it reflects mostly true score, relative to error. The component of observed score that underlies consistency across repeated measures with the same or a parallel test is the score, so operationally, reliability refers to consistency or repeatability in measurement across:

i. Time;
ii. Parallel/ alternate forms;
iii. Parts or portions of same test; and across
iv. The items of the test.

In other words, it is an indication of how well items of the test hung together across time, across forms, or across parts/portions of the same or parallel test as measure of whatever it- is measuring. Depending on which of these repeatability one is concerned involved with, reliability may be estimated through either a single or multiple administration of test. Each of these methods ends up with different estimate of reliability of the same test because each is sensitive to different sources of error. This index too is sample dependent as derived because it is not a bona fide property of the test itself but the scores. It is a product of the interaction of test and examinees' characteristics.

The ability of the testing process to give us valid or true information reduces the level of our error in educational decision-making. CTT-based testing has been reported to be very weak at providing valid information on the ability underlying learners' performance to the different arena of decision-making. Nenty (1996) asserted that the level of confidence with which scores from test items could be used to infer the ability under measurement depends on both (i) how well the test items represent the totality of the ability and only the ability under measurement, and (ii) the extent to which it is only this ability that sustains responses to these items. CTT has no provisions to ensure and check any of these two conditions, hence, securing valid information from CTT-based test is always a herculean task.

IRT- based Reliability and Validity

IRT replaces the concept of reliability as a single index for whole test with a function called the information function. Earlier on we saw that a test's ability to give us consistent information reduces the level of our uncertainty in decision-making. In statistics, information means reduction in uncertainty and the analogous measure in CTT is reliability. With IRT during the estimation of examinees standing on the trait, for example the ability Θ, each item in a test contributes some amount of information that helps to reduce the amount of uncertainty about the resulting estimate IRT determines item and test information in place of CTT's reliability. Information is reciprocal of the standard error of measurement at a given trait level. The more information provided by the items in a test a particular ability level, the smaller the error of measurement associated with the estimation of that ability level. Therefore, each Θ can be used to illustrate how much information an item contributes at each point of the theta scale. With a large bank of pre calibrated items for measuring a given trait, items could be selected to give a test information function shaped to control measurement error very accurately. IRT was developed basically to improve the validity with which latent traits are defined and measured. With IRT-based scores, one can with certainty answer the question 'what exactly does a student mean? While CTT-based performance score means significantly more or less than the ability it was intended to represent, IRT-based logit means exactly the trait or ability level under measurement possessed by the person.

Three models of IRT are applicable to dichotomously scored cognitive test and with two assumptions that underlie the IRT's strength as a desirable measurement theory. The assumptions are that of unidimensionality and that of local independence. In a non-technical language, these assumptions have it that for the responses to a test to be amenable to IRT analysis, all its items must be developed to measure one, and only one, ability; and the examinee's responses to one item is independent to their responses to another item in the test. In order words, if the ability under measurement is held constant across the items, responses to the items are uncorrelated. The person score or parameter that results from IRT analysis is invariant across tests of the same ability wherever it is given. So it gives a reliable and valid

data based on which decisions could be reliably made. IRT has a number of advantages over CTT in assessing learning, in developing better measures and in assessing changes over time.

Assessment in Nigerian School System

In most cases, the journey of a Nigerian child starts with nursery school. At this level, the child is confronted with his/her first form of formal assessment which could be oral or written. At the primary school, in compliance with the Nigerian National Policy on Education, students are introduced to Continuous Assessment (CA) and culminate at adulthood in a Ph.D Programme. Continuous assessment is officially defined as a mechanism whereby the final grading of students in the cognitive, affective and psychomotor domains of behaviour takes account in a systematic way of all their performances during a given period of schooling. Such an assessment involves the use of a variety of evaluation techniques for the purpose of guiding and improving the learning and performance of the students (FME, 1980). Awotunde and Ugodulunwa (2001) stated that the advocacy of CA in Nigeria arose from the belief that it will:

1) Give teachers greater involvement in the total assessment of learners
2) Provide a more valid and reliable assessment of the learners' overall ability and performance
3) Enable teachers to be more flexible and innovative in their teaching
4) Provide a basis for more effective guidance of learners
5) Reduce examination malpractice and
6) Provide a basis for teachers to improve their instructional methods

It is an issue of concern to the stakeholders the quality of graduates from Nigerian educational institutions considering their knowledge and skills when compared with their counterparts from other parts of the world. Among other reasons on this disturbing development is teachers' lack competence in the operation of CA (Atsumbe, 1996). The assessment practice in Nigeria emphasizes the acquisition of certificates and is in sensitive to the improvement of teaching, learning and competencies. This type of assessment does not utilize the strengths of authentic Curriculum – based and competence – based assessment (Carew & Hamman-Tukur, 2001). It appears at all levels of Nigerian educational assessment as there is too much desire for obtaining good grades than acquisition of deeper knowledge and skills.

It should be noted that presenting a wrong picture of a child has damaging effect not only of his personality but on his career aspirations. Teachers and other persons involved in operation of continuous assessment system are not very honest in their assessment and cannot be exonerated for distorted identity of Nigerian learners as represented by the assessment measures. Assessment in Nigeria and many developing countries are highly predictable from year to year. This type of assessment encourages teaching practice that focuses on past papers and practising answers to previous papers. This results to surface learning by students that hardly last beyond the assessment; it is a case of teaching and learning for assessment.

Effectiveness of Assessment Measures in Decision Making

Educational institutions serve primarily as selection and certification agencies, whose job is to measure and label people and as such should be kept out of politics. Assessment with a prime purpose of obtaining information becomes an indispensable tool in educational decisions.

Assessment measures could be obtained from tests or non-tests integrations of such measures as they give reliable and valid information for selection replacement of students.

The most important psychometric property that relates with effectiveness of assessment measure is the criterion in related validity which according to Denga (1987) aims at prediction and generalization from one score to another.

Goge and Berline (1991) defined prediction, as an attempt to forecast an outcome on the basis of data or information considered relevant to the observed event. In education, predictions are involved in making decisions about grouping of student employment for jobs seekers admission of students into colleges, formulation of vocational and educational choices in retention or promotion of students.

Ugodulunwa and Ugwuanyi (1999) described prediction as the primary purpose of all scientific inquiries. They stated that prediction in education is concerned with the learning capacity, the potential growth, success and adjustment of the organism. They identified problem areas of prediction as that of an accurate classification, selection and guidance.

Classification is assignment of the individual into category which he/she best belongs and where he/she has the highest potential for self-expression. Selection, on the other hand, involves categorization of individuals in a way that would indicate the probability of success on a desired task. Guidance – oriented examination provides information regarding abilities, interest and the chances of success in reaching at various goals. Kerlinger (1973) asserted that prediction has a very close association with the future. Considerable amount of investigations had been done on the validity of examinations in Nigeria and other countries. Some of these studies relate intelligence/aptitude tests with achievement tests while others relate achievement tests with other achievement tests. Few of such researches are considered in this paper.

Houston (2000) used stepwise multiple regression analysis, examined the relationship between college grade point average (GPA) as the dependent variables. The verbal Aptitude Test (SAT-V and SAT-M) were evaluated on a sample of black female undergraduates who initially had not been accepted but later were admitted into a predominantly while Liberal Art College. The results showed that after semesters, 19 per cent of the students withdrew or were transferred to another institution, 52% were graduated and 27% were dismissed for academic reasons.

Halpins and Schaer (2011) studied the relative effectiveness of the California Achievement Test (CAT), America College Testing School Grade Point Average (GPA) in prediction college freshman GPA. The incremental and differential effectiveness of CAT, ACT and SAT in addition to high school GPA was also studied. Although a high school GPA was the best single predictor, the CAT was an effective predictor as was the ACT or the SAT. The use of the ACT, SAT or SAT resulted in an 18.45% increase in predictive efficiency over that obtained using high school GPA alone. The increase in predictive efficiency was very nearly the same within rounding error for the three tests (CAT, SAT and ACT), they failed to demonstrate differential increment validity.

Hamman-Tukur and Amin (1996) studied performance in 'O' level examinations and achievement in Doctor of Veterinary Medicine (DVM) programme and found that 'O' level subjects were significantly related to and can predict success in three levels of the DVM degree programme namely pre-clinical, clinical and overall DVM. Their results showed that performance in Biology, Chemistry, and composite 'O' level was significantly related to achievement pre-clinical, clinical, and overall DVM. However it was only performance in Biology and English that could significantly predict success in the DVM (R= 0.530, 0.512, and 0.521) for overall, pre-clinical and clinical DVM respectively. Performance in 'O' level Mathematics and Physics was significantly related to achievement in any of the DVM levels.

In 1976, the Federal Government of Nigeria established Joint Admissions and Matriculation Board (JAMB) as a central examination body saddled with the responsibility of conducting

selection and placement examination into Nigerian higher institutions of learning. The first examination of this body was conducted in 1978 and since then entrance examination into Nigerian Universities had continued to be handled by JAMB examination which is a combination of students' test of aptitude and achievement in the different subject areas.

The validity of UTME scores, as demonstrated in the correlation coefficient between of UTME scores and universities, CGPA has been a concern for users of UTME scores and critics. One the criticisms of the scores is that they do not adequately predict individual student university grades. Obioma and Salau (2007) observed that UME scores were the least predictors of both the first CGPA and the final year CGPA and accounted for 1-5% of the variance in first year CGPA and 0-6% of final year CGPA. In a similar study, Ojerinde and Kolo (2009) observed that UME scores accounted for a very low amount of variance (2-8%) in the final CGPA.

In a related study, Jebson (2017) examined the relationship between UTME scores and CGPA of 200 and 300 level undergraduates admitted in 2011/2012 and 2010/2011 academic sessions respectively in Modibbo Adama University of Technology, Nigeria. Using ex-post facto design and Pearson Product Moment Correlation statistic, the results of the analyses showed that there was no significant relationship between UTME scores for undergraduate students and their CGPA.

Research studies revealed that students JAMB scores as a single measure for selection and placement into Nigerian tertiary institution of learning has a low predictive validity. However, concerns about accurate placement have recently led some institutions to introduce other measures such as post UME screening to validate placement decisions. Belfied and Crosta (2012) asserted that there is a growing recognition that assessing students using a single standardized measure may contribute to misplacement and that more appropriate decisions can be achieved by using more than one measure of student readiness.

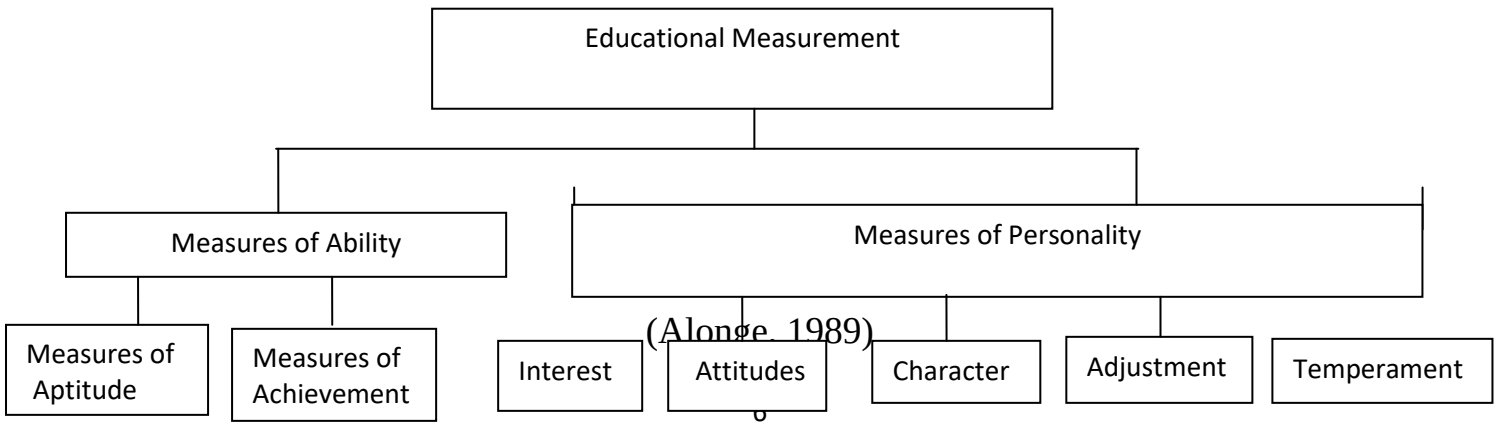
The use of assessment techniques in Nigerian educational institutions has been misleading in the sense that students and parents have the impression that all matters in the institutions are to obtain a certificate at the end of a course. This has led so many students to be involved in examination malpractice. Assessment techniques are more concerned with assessing the job or educational course. It is expected that the assessment should be valid, it must measure what is the set objectives of the course and it must measure it accurately. Thus, what to measure include:

Measure of ability (cognitive domain) and

Measures of personality (affective and psychomotor domains)

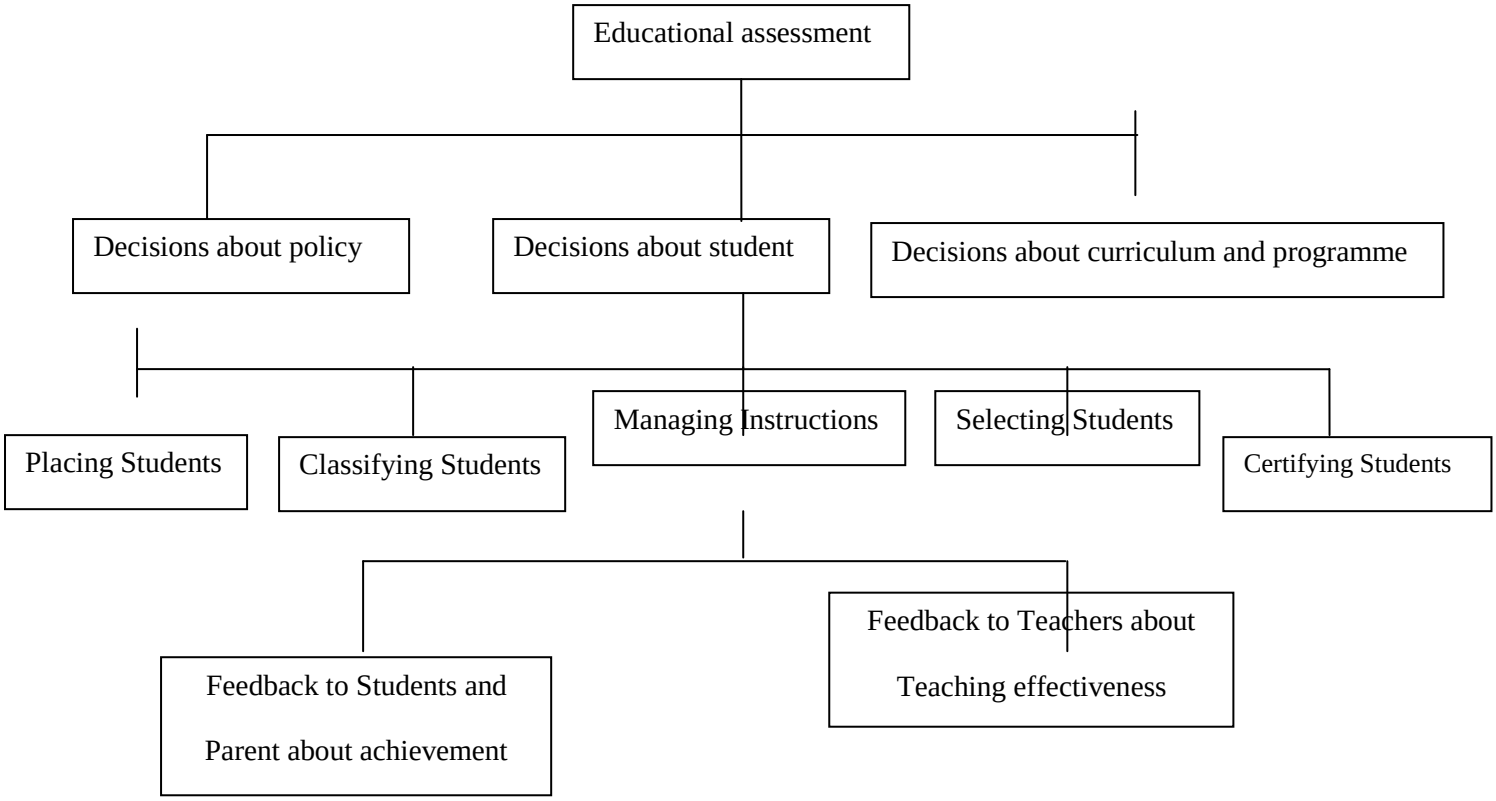
Measures of ability have to do with what a person or group of people are capable of doing in terms of their intellectual capabilities.

Measures of personality deal with such variables like attitudes, interest, character, temperament and adjustment. The overall domain can be illustrated as follows:



(Alonge, 1989)

Research in educational psychology further suggests that an array of factors beyond cognitive intelligence and skills are predictive of college success and future outcomes. Sedlacek (2004), for example, argues that non-cognitive measures of adjustment, motivation, and perception are strong predictors of success, particularly for under-represented minority students. Principally assessment is mostly concerned with decision making. The decision can be about policy, students, curriculum and programme. Similarly, the overall domain of assessment can be illustrated as follows:



(Capper, 1996)

Modern trend in assessment in most developed nations has moved from a singular high stake examination to a system of multiple assessments. In Africa and Nigeria in particular, what can be closely related to this system is CA. However, multiple assessments appear to be more engaging as rigorous modes of assessment other than the paper and pencil C.A. as used. Multiple assessment systems are designed to include formative and summative components, along with technological innovations to assess more complex application of learning. While such technologies may not be found in must educational institutions in Africa, multiple assessments have begun to gain popularity in classroom because they provide teachers with resources for more formative assessment (Hidden Curriculum, 2014). Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning in order to improve student's achievement of intended instructional outcomes (UNESCO, 2016).

Conclusion

Information as input into decision- making is useful to the extent that it is valid. Theories on which the generation of such information is based must itself be valid to be able to sustain valid results, but CTT, as a theory-based on which information on testing is generated, has been shown to be fundamentally flawed. IRT is a modern test theory with advantages that overcomes the weaknesses of CTT. It improves the amount of information the assessment measures and conveys about the estimation of learner's ability. Assessment measures shape and define learner. Using the right theory in constructing an assessment technique will result in obtaining correct estimation of learner's ability.

It is observed that IRT has the ability to accept good items and at the same time reject or redeem bad ones. At all levels of education in Nigeria, the policy on continuous assessment provides basis for multiple assessment. It is expected that the multiple measures obtained through assessment should comprise cognitive and non cognitive measures. Using data obtained from multiple measures rather than single measure for selection or placement would improve decisions about learners. When such decisions on students are done appropriately, the chances of their success would be high and hence reduces phobia for education.

Recommendations

Based on the issues raised in this paper, the following recommendations were made:

1. Examining bodies in Nigeria and other African countries should use IRT in constructing and administering the high stake examinations. This will improve the information function of such examinations.

2. Continuous assessment system should be strengthening to ensure the inclusion of non cognitive measures.

3. If a single measure is to be used for decision making, then it should be repetitive so that an average measure is utilized to make it more reliable.

4. Selection and placement decisions should be based on data from multiple measures.

5. Using multiple sources of information about students for their placement into programmes at the university would reduce their desire for examination malpractice and they are likely to be successful.

6. Countries without multiple measures policy on placement decisions should formulate such policy so as to improve placement policy.

References

Alonge, M. F. (1989). *Measurement and Evaluation in Education and Psychology.* Ado – Ekiti. Adebayo Publishing Company (Nig.) Ltd

Atsumbe, B. N. (1996). *Issues in the implication of Continuous Assessment in Secondary School introductory technology Programe.* In G. A. Badmus and P. I. Odor (Eds.) Challenges of Managing Educational Assessment in Nigeria (pp. 235 – 244). Kaduna Atman Limited

Awotunde, P. O. & Ugodulunwa C.A. (2001). *An Evaluation of the administration of continuous Assessment in Nigeria Secondary School* The Nigerian Academy of Education Proceeding of the 16th Annual Conference held at University of Jos, Jos 12-16 November, 2001

Belfied, C. & Crosta, P (2012). *Predicting success in college: The Important of Placement t. Tests and High School Transcripts. (CCRC working Paper No. 42)* New York, NY Columbia University, Teacher College, Community College Research Centre

Brannick, M. T. (nd). *Items response Theory*. Retrieved on August 8, 2004 at http:/una.cas.ufs.edu (2) umbrannic/files.pnet/irt.htm

Capper, J (1996). *Testing to Learn Learning t Test*: Washington DC Academy for Educational Development

Carew, P.F.C & Hamman-Tukur, A. (2001). Recent Trends in the Assessment of Achievement: *Implication for the Nigerian School System.* The Nigeria Academy of Educational Proceedings of the 16th Annual Congress Held in the University of Jos, Jos 12 – 16 November, 2001

Federal Republic on Nigeria (2014). *National Policy on Education* Abuja: NERDC Press

Federal Ministry of Education (1980). *A Handbook on Continuous Assessment.* Lagos: Government Printers

Goge, N. L. & Berline D. C. (1991). *Educational Psychology*. Boston Houghton: Mifline Company

Halpins, G. & Schaer, B. B. (2011). *Relative effectiveness of California achievement tests in Comparison with ACT Assessment average.* Journal of Educational Psychology measurement 39 (4) 821-825

Hambleton, R. K. & Jones, R.W. (1993). *Comparison of Classical test and item response theory and their application to test development.* An NCME Instructional Module 16 Fall 1993 Retrieved on 06/7/04 from http://www.acme.org/pubs/items/24 PDF

Hamman-Tukur, A. & Amin, J.D. (1996). Performance in 'O' level examinations and achievement in doctor of veterinary medicine programme: Implications for admission criteria. Educational Forum 2 (2), 92-99

Hidden Curriculum (2014, August 26). Assessment in S. Abbott (Ed.) *the glossary of Education reform*. Retrieved from http://edglossary.org/hidden-curriculum

Houston, L. N. (2000). *Predicting academic achievement among specially admitted black female college students*. Journal of Educational Psychology measurement 39 (4) 1189-1195

Jebson, S. R. (2010). Review of classical Test Theory (CTT) and Item Response Theory (IRT) and the Applicability of IRT in Reducing the politics of Assessment *Journal of Education and Applied Psychology 3 (1) 133-142*

Jebson, S. R. (2017). Validity of Unified Tertiary matriculation Examination Scores*: implication of Test Theories application to Stakeholders' Educational Assessment Literacy.* African Journal of Theory and Practice of Educational Assessment (AJJPEA) Vol. 5, November, 2017

Kerlinger, F. N. (1973). *Foundation of behavioral research,* New York: Harper & Raw Publishers

Nenty, H. J. (2001). Research in Educational assessment in Africa: *Challenges for the 21st Century*. In J.P. Lefka and Matsoso (Eds) Educational research for development: *Challenges for the 21st Century pp 35-59,* Roma Lesotho: Institute of Education National University of Lesotho

Nenty, H. J. (1996). Advance in test validity. In G.A. Badmus and Odor, P. I. (Eds) *Challenge of Educational Assessment.* Kaduna, Nigeria: JAMB, NABTEB & NBEM

Oboima, G & Salau M. (2007). *The predictive validity of public Examination: A case study of Nigeria,* A Paper Presented at the 33rd Annual Conference of International Association of Educational Assessment (IAEA) held in BAKU Azerbaijan, 16-21 September, 2007

Ojerinde, D. & Kolo, T. N. (2009). *Prediction of first year University education performance from entry academic performances. Being a Paper presented at the 27th annual conference of AEAA* Yaounde Cameroun 24-28 August, 2009

Sedlacek, W.E. (200s4). Beyond the Big Test: *Non cognitive Assessment in Higher Education,* San Francisco, CA: Jossey – Bass.

Satterly, D. (1981). Assessment in Schools (*Theory and Practice in Education (*No.1 London: Basic Blackwell Publishers Ltd

Ugodulunwa, C.A. & Ugwuanyi, C. L. (1999). *Understanding Educational evaluation* (2nd ed) Jos: Fab Educational Books

UNESCO (2016). Out in the Open: *Educational sectors Responses to violence based on sexual orientation and gender identity/expression (PDF).* Paris, UNESCO p. 54