



Psychometric Properties of a Test: An Overview

Solomon Chukwu Ohiri¹, Dr. Ihebom, Chidinma Nnennaya²

¹Directorate of Academic Planning, Alvan Ikoku University of Education, Owerri, Nigeria E-mail: ohiri.chukwu2018@gmail.com

²Examinations & Records Unit, Alvan Ikoku University of Education, Owerri, Nigeria E-mail: ceerimaconcept@gmail.com

DOI: <https://doi.org/10.55248/gengpi.5.0224.0539>

ABSTRACT

Tests are useful tools in the measurement of educational or psychological attributes. They are indispensable tools in the collection of important data in which educational decisions are based. In order to obtain the vital data required, tests must possess acceptable psychometric characteristics or properties. Psychometric properties are most important and fundamental features in the evaluation of any measuring instrument or tool. Validity and reliability are considered the main psychometric properties of such instruments. Validity refers to the degree of accuracy of an instrument (test) to measure exactly what it is intended to measure. Reliability on the other hand, is the ability of an instrument (test) to reproduce a result consistently in time and space. The purpose of this paper was to review the validity and reliability as psychometric properties of tests in some details. The determination of tests' (instruments') psychometric properties is imperative in order to ensure the quality of the instruments used.

Key words: Psychometric Properties, Test, Measurement, Validity and Reliability

Introduction

Psychometrics is a field that deals with the theory and techniques of psychological measurement. It is the science of psychological assessment, and is a foundation of assessment and measurement (Rust, 2007). According to Borsboom and Molenaar (2015), psychometrics is a scientific discipline concerned with the question of how psychological constructs (e.g., intelligence, neuroticism, or depression) can be optimally related to observables (e.g., outcomes of psychological tests, genetic profiles, neuroscientific information). Generally, it refers to a field of study within psychology and education devoted to testing, measurement, assessment, and related activities. The field is concerned with the objective measurement of skills and knowledge, abilities, attitudes, personality traits, as well as educational achievement. Psychometrics is a highly interdisciplinary field, with connections to statistics, data theory, econometrics, biometrics, measurement theory, and mathematical psychology.

The intrinsic components of a test are precisely called its psychometric properties. These properties are typical characteristics of test that identify and define critical aspects of an instrument, such as its reliability for use in a specific circumstance. In a nut shell, psychometric properties do reveal information about a test's adequacy, relevance and usefulness.

The psychometric properties of a test are associated with the data generated from the assessment to determine how well it evaluates the interest construct. The development of a valid test is conditional on the fact that it has been subjected to statistical analysis, which ascertains that it has adequate psychometric properties.

Tests are very vital tools in educational measurement and evaluation. They are used in gathering valuable data in which educational decisions are based. Such decisions should be made carefully based on authentic and accurate data. Certainly, erroneous data will lead to wrong decisions. To obtain authentic and accurate data, the test, employed for that purpose should possess some essential characteristics (psychometric properties) (Ezeh, 2003). Two of the most widely discussed psychometric properties of tests are validity and reliability, which are broad categories of related concepts.

Test Validity

Validity is arguably the most important quality of a test because it has to do with the fundamental measurement issue of what our measuring instruments are really measuring (Bandalos, 2018). The validity of a test is the degree of accuracy with which the test measures what it is intended to measure (Iwuji, 1997). Messick (1995) described validity as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment". Allen and Yen (1979) referred to validity as the appropriateness, meaningfulness and application of test scores as well as the usefulness of judgments made from test scores. Validity gives meaning to the test scores and gives authority to the link between how an individual performs in the test and the test's stated measurement criteria. It tells

us about the degree to which it is possible to draw specific conclusions or predictions, based on an individual's test score; in other words, it provides information on the efficacy of the test.

Test has been seen as a measuring instrument designed for the measurement of educational or psychological attributes. Most of these attributes are constructs whose existence we cannot see, but rather infer from manifested behaviours associated with individual constructs. The validity of a test, therefore is the degree of accuracy with which a test measures the specific attribute it is constructed to measure.

Types of Validity

There are three major types of validity – content, criterion-related and construct validity. However, there is another type of validity, usually referred to as "face validity", which in a strict or technical sense is not considered as a type of validity (Ezeh, 2003).

Content Validity: Content validity is most often evaluated during the development of an instrument rather than after the instrument has been created. The content validity of any given test refers to the extent to which the test measures both the subject matter content and the instructional objectives designed for a given course. It is the extent to which the questions on the instrument and the scores from these questions represent all possible questions that could be asked about the content or skill (Creswell, 2005). According to DeVellis (2003), content validity refers to the degree to which the instrument comprehensively assesses the underlying construct of interest. It is the most appropriate form of validity for achievement test. According to Nwana (1982), to ensure a sound construction of a valid achievement test, the following principles have been recommended:

1. Questions should be set from all parts of the syllabus.
2. The number of questions set in each section of the syllabus must reflect the relative importance of each section.

Content validity usually depends on the judgment of experts in the field. The unclear and obscure questions can be amended, and the ineffective and non-functioning questions can be discarded by the advice of the reviewers (Mohajan, 2017). Expert judges may be asked to rate the 'fit' of each item on the measuring instrument to its respective domain using a Likert-type rating scale (Briere, 2011). Examining the variance in judges' fit ratings would then allow the instrument developer to assess the degree of agreement for item fits across all domains and judges. The higher the degree of agreement, the higher the content validity.

In order to ensure a high content validity, a "table" is made to map out the content or learning materials which the students have been taught as well as the levels of behaviour objectives at which they have learnt them. This provides a basis for the construction of test items. This table is known as table of specification.

Criterion-Related Validity: Another name for criterion-related validity is empirical validity. The criterion-related validity of a test can be defined as the degree of correlation that exists between the scores obtained by a group of testees in a test and their scores in another test known to be a standard measure of the characteristics the test under study claims to measure (Nkwocha, 2019). The correlation coefficient estimated is called the validity coefficient. The closer the validity coefficient is to one (1), the higher the level of validity.

Gronlund in Nkwocha (2019) states that "criterion-related validity is demonstrated validity by comparing the test scores with one or more external variables considered to provide a direct measure of the characteristics or behaviour in question." There are two types of criterion-related validity: concurrent-related validity and predictive-related validity.

a) Concurrent-criterion-related validity: A test is said to have concurrent-criterion-related validity when there is correlation between the scores obtained by a group of testees in the test and scores they obtained in a criterion test which they took within the same period (Nkwocha, 2019). It is the degree to which the scores on a test are related to the scores on another, already established as valid, designed to measure the same construct, administered at the same time or to some other valid criterion available at the same time (Mohajan, 2017). It is established by correlating one question with another that has previously been validated with standard setting (Okoro, 2002). The two tests to be correlated are taken concurrently. They can be described as equivalent tests.

The following steps should be adopted as procedures for estimation of concurrent-criterion-related validity of a test.

- i. Identify a criterion test, that is test which has been recognized as a standard measure of the characteristics the new test is designed to measure.
- ii. Give the test under study and the criterion test to the same testees simultaneously within a short interval.
- iii. Compute the correlation coefficient of the scores obtained in the two tests using appropriate correlation procedure. If the calculated validity coefficient (correlation coefficient) is high and positive, the test is said to have high concurrent criterion validity.

b). Predictive criterion-related validity: It is often used in programme evaluation studies, and is very suitable for applied research (Mohajan, 2017). It is a test constructed and developed for the purpose of predicting some form of behaviour (Allen & Yen in Mohajan, 2017). It indicates the ability of the measuring instrument to differentiate among individuals with reference to a future criterion. This implies that it indicates the degree of correspondence between scores on the test in question and future outcomes that are expected to be related to the characteristics measured by the test (Tuckman in Okoye, 2015). A test is said to have predictive criterion-related validity if individuals' present scores in a test can be used to predict their future performance in another test which measures a similar criterion. Examples of tests which are expected to give predictive validity are entrance examination results such as

Joint Admission and Matriculation Board (JAMB) examination results. They are expected to predict the candidates' grades in the subjects at graduation. If the entrance scores in a subject correlate with the graduation scores in that subject, the test for entrance exam is said to have predictive validity.

Predictive validity is most applicable to aptitude or intelligence tests whose scores are intended to predict what the taker will be able to do in future. It should be noted that there is a big interval between the time of obtaining scores from the test (i.e. predictor) and the time of obtaining scores from the criterion. The higher the correlation between the criterion and the predictor indicates the greater the predictive validity. If the correlation is perfect, that is 1, the prediction is also perfect. Most of the correlations are only modest, somewhere between 0.3 and 0.6 (Mohajan, 2017).

Construct Validity (Logical Validity): This refers to the extent to which the test measures a psychological construct or trait, which it is supposed to measure. Constructs are terms, which have no direct representations in the empirical world but which are used to describe and explain certain aspects of our behavior. Examples of psychological constructs or traits include intelligence, speed of reading, honesty, anxiety, stability, sociability, verbal fluency, etc. For each of these constructs, a theory may be available aimed at explaining what it is and how people who possess it are likely to behave. If there is a test that is designed to measure any of such construct, we expect scores obtained by people in it to agree with the theoretical prediction.

Construct validity therefore deals with the degree to which scores obtained from a test agree with the theory underlying the construct. It involves testing a scale in terms of theoretically derived hypotheses concerning the nature of underlying variables or constructs (Pallant, 2011). The evidence of construct validity can be ascertained by obtaining convergent validity and discriminant validity. An instrument is said to have a convergent validity if scores obtained from it are highly and positively related to scores from an instrument that measures a similar construct. On the other hand, an instrument has discriminant validity, if scores obtained from it have little or no relationship with scores from an instrument that measures a construct that is dissimilar to it. Evidence of convergent validity and discriminant validity can be obtained using factor analysis.

Face Validity: According to Okoye (2015), face validity refers to whether the test looks valid on its surface. It can be described as the extent to which the test looks like an instrument on the variable the test is designed to examine (Nkwocha, 2019). It is the simplest and least precise method of determining validity which relies entirely on the expertise and familiarity of the assessor concerning the subject matter (Nwana, 2007). Face validity is not computed. It is usually checked through surface observation of items. It is usually used to describe the appearance of validity in the absence of empirical testing (Cook & Beckman, 2006). This type of validity is needed for acceptance of the test by evaluators and for sustaining the motivation of the test users. A mathematics test should contain mathematical signs and symbols that would make it look like mathematics test. If it is composed of a composition passage like an English comprehension test, testees who prepared for a mathematics test may go away with the notion that it is a test for candidates who want to be examined in English comprehension. Before a test is considered for other types of validity, it must first have face validity.

Importance of Test Validity

Test validity is relevant to education in the following ways:

1. Test validity helps in the achievement of test purposes.
2. Test validity guarantees collection of authentic data. It ensures that tests designed to measure specific attributes, measure them.
3. Validity of a test enhances accurate evaluation of capabilities and potentialities of individuals based on test results. Placement, promotion, and certification of individuals are adequately done when tests measure exactly what they are designed to measure.

Factors Affecting the Validity of Tests

1. Poor sentences structuring in tests and the use of inappropriate vocabularies affect the testee's comprehension of the task required in the tests.
2. Poor constructions of items.
3. Ambiguous and misleading statements.
4. Improper arrangement of the test items
5. Unclear instructions.
6. Most tests are specific in terms of the behaviour being measured. Inclusion of items from different behaviours other than that intended for such tests, reduces its validity.

Reliability

A reliable test is one we can trust or we can use to measure a person's performance approximately the same way each time. Reliability is used to evaluate the stability of tests administered at different times to the same individuals and the equivalence of sets of items from the same test (Kimberlin & Winterstein, 2008). According to Iwuji (1997), the reliability of any measuring instrument is the degree of consistency of the results of its repeated measures of an attribute it is measuring. Nkwocha (2019) saw reliability as the degree to which test items consistently measure any phenomenon they measure. Anastasi in Nkwocha (2019) opined that reliability refers to the consistency of scores obtained by the same person when reexamined with the same test on different occasions. It measures consistency, precision, repeatability, and trustworthiness of a test (Chakrabarty, 2013). It is the degree to which an assessment tool produces stable (free from errors) and consistent results (Mohajan, 2017). It indicates that the observed score of a measure reflects the true score of that measure. The term reliability in the context of measurement theory refers to the dependability or consistency of measurements across conditions (Bandalos, 2018).

Reliability of a test relates to the degree of consistency or stability, which the test exhibits. It is that attribute of a test which is responsible for making its results equal (Kulwinder, 2012). In other words, it is concerned with how reproducible such test results are when the measurement is repeated on different occasions. Assuming that Adaeze and Chizzy obtained scores of 60% respectively in a given test. Three days later, the same test was re-administered to the same class, and their scores were 75% and 30% respectively. As the scores of these two testees on the same test are inconsistent over time, such test is said to be unreliable. Reliability therefore connotes the accuracy, trustworthiness or dependability of a test (Ezeh, 2003).

Coefficient of Reliability

The degree of consistency of a test is expressed as a coefficient called the coefficient of reliability. The coefficient of reliability falls between 0 and 1, with perfect reliability equaling 1, and no reliability equaling 0. Thus, in most cases it is determined by correlating two sets of scores independently obtained by the test.

Reliability coefficient takes value between 0 and 1. The following guidelines can be used for the interpretation of the values of reliability coefficients according to Yolonda in Ohiri and Okoye (2023):

- 0.9 and greater = excellent reliability
- 0.8 - 0.9 = good reliability
- 0.7 – 0.8 = acceptable reliability
- 0.6 – 0.7 = questionable reliability
- 0.5 – 0.6 = poor reliability
- 0.5 and less = unacceptable reliability

Ways of Establishing Test Reliability

The reliability of a test is usually determined by comparing the reliability coefficient. This coefficient can be obtained using various methods, thereby giving rise to different types of reliability. The measures commonly obtained in the bid to ascertain reliability are:

- i. measures of stability often referred to as the test-retest method,
- ii. measures of equivalence usually referred to as equivalent form or parallel form method, and
- iii. measures of internal consistency which include split-half, Kuder-Richardson and Cronbach's coefficient alpha methods.

1. Test-Retest method: This involves estimating the correlation coefficient of two sets of scores, obtained from the same group of people, on the same test which was given to them at two different occasions, under the same conditions, after a short time interval. The scores the testees obtained at the two different occasions are correlated using an appropriate type of correlation method. The closer the correlation coefficient index is to one (1), the more reliable the test. If the reliability coefficient is high, for example, $r = 0.98$, we can suggest that the instrument is relatively free of measurement errors. If the coefficients yield above 0.7, are considered acceptable, and coefficients yield above 0.8, are considered very good.

Limitations of Test-Retest Method

The conditions under which the test is taken at the two occasions may differ, particularly if invigilated by different people.

- a) The testee's state such as emotional state may differ on the two occasions.
- b) Some testees may not be available during the second test.
- c) More learning may take place before the next testing.
- d) Forgetting may also influence the response made during the second testing.
- e) It is expensive in terms of time and economy.

2. Equivalent form method: It is a measure of reliability obtained by administering two different versions of an assessment tool to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions. If they are highly correlated, then they are known as equivalent-form reliability.

Limitations of Equivalent form method

- a) It would be fatiguing and time consuming for testees who are required to write two tests almost at the same time.
- b) It is costly to produce two different tests at the same period.
- c) An equivalent test may not be easily available.

3. Split-half method: In this case, the test is administered to the same group of testees once. The common procedure is to divide the test into two groups, with odd-numbered items usually placed in one group and even-numbered in the other. Each testee gets two scores, one from each half of the test. The

scores on one half of the test are correlated with the scores on the second half. The computation is done with Pearson's product moment correlation method. Since the score of each testee has been divided into two, the correlation index estimate is the split-half reliability coefficient. To calculate the reliability index for the full test, Spearman-Brown computation formula is used based on the split-half. Spearman-Brown formula is given as:

$$r_f = \frac{n r_s}{1+(n-1) r_s}$$

Where r_f = Reliability of the full test
 n = The number of times the test was shortened or elongated
 r_s = The reliability of the shortened or elongated test

Split-half method is used when the items of the test are homogenous. That is, when the items measure one construct.

Limitations of Split-half form method

- The two halves may actually not be properly split. Hence, the correlation coefficient it produces may be unreliable and misleading.
- It would not be appropriate for tests which measure different constructs.
- Chance error can arise in the test, which increases the amount of r in the two parts

4. Kuder-Richardson (K-R) method: Kuder-Richardson's reliability is a method that makes use of the full test. It is used for frequency scores. Kuder-Richardson's approach avoids the problem of how to split the items, and it has two procedures - KR-20 and KR-21. The KR-20 is best used for a test that does not have many items because the formula requires computation of the proportion of those who passed each item and the proportion of those failed each item. KR-21 does not require such rigor and hence is used when the test items are many.

In the computation of K-R reliability coefficient, a single test is administered to a group of testees. It estimates the consistency of responses to all the items in a test.

$$\text{The formula for KR-20} = r = \left\{ \frac{K}{K-1} \right\} \left\{ 1 - \frac{\sum pq}{SD_t^2} \right\}$$

Where K = number of items the whole test is composed of

P = proportion of those who passed each item

q = proportion of those who failed each item

SD_t = square of standard deviation of testees scores of the whole test

$$\text{The formula for KR-21} = \left\{ \frac{K}{K-1} \right\} \left\{ 1 - \frac{\bar{X}(K-\bar{X})}{K(SD_t)^2} \right\}$$

Where K , SD_t are defined as in KR-20

\bar{X} = mean of the summated scores

Kuder-Richardson method provides an estimate of the average reliability found by taking all possible splits without actually having to do so.

Limitations of Kuder-Richardson (K-R) method

- Its application is not recommended for scales that offer multiple choice formats.
- In this method, the reliability coefficient obtained is somewhat less than that obtained by other methods.
- The Kuder-Richardson method is anchored on the basic assumption that the difficulty level of all items will be the same, which is not possible in practice.
- K-R method cannot be applicable in speed tests, because the parts of the test are not independent.

5. Cronbach's coefficient alpha: This is a lower-bound estimate of reliability under the assumption that items are with uncorrelated errors. It can be used for any mixture of binary (true/false) and partial credit items (true/sometimes/false). This is computed by correlating the score for each scale item with the total score for each observation (usually – individual test takers), and then comparing that to the variance for all individual item scores.

The formula for Cronbach's coefficient alpha according to Nkwocha (2019) is

$$r = \left\{ \frac{K}{K-1} \right\} \left\{ \frac{V_t - \sum V_i}{V_t} \right\}$$

Where r = coefficient alpha (the reliability index)

K = number of items that compose the test

V_t = variance of total scores of each respondent on the test

V_i = variance of scores obtained by all respondents on each item.

$\sum V_i$ = Sum of total variance of scores for all items.

Limitations of Cronbach's coefficient alpha method

- It does not account for measurement error, which can lead to an overestimation of reliability
- It assumes that all items in a test are measuring the same attribute, which may not always be true.
- Cronbach's alpha coefficient are highly influenced by the number of items of the measurement instrument.

Importance of Reliability

Some values of test reliability are:

- Test reliability makes a test dependable for future planning.
- Reliable tests can be used to establish criterion-related validity and equivalent form reliability for new tests that measure the same traits.
- Promotion, placement, certification, all these are based on test reliability.

Factors that Affect Test Reliability

- Use of different administration procedures: The results obtained when the test is conducted in a well arranged, spacious and conducive environment would differ from a result obtained in a crowded examination hall that gave room to examination malpractice.
- Momentary fluctuations such as level of anxiety, motivation, fatigue and lack of concentration on the part of both the testee and the scorer can reduce the degree of the test reliability.
- Ambiguity: An ambiguous question can be interpreted by testees in different ways at different occasions. The responses they will give at different occasions may therefore differ.
- Learning and forgetting: when test-retest method is involved, test scores may be affected by the fact that the testee may acquire new knowledge or forget some of the things he earlier knew, after the initial testing.

Differences Between Validity and Reliability of a Test

S/N	Validity	Reliability
1.	Test validity always includes test reliability	Reliability does not always imply validity
2.	It deals with relationship between what is measured and test objectives	It does not deal with relationship between what is tested and the test objectives
3.	The concern is not on consistency in the responses given to the test items	Reliability deals with consistency in giving the same results.
4.	Test validity is concerned on how truthfully a test measures what it is supposed to measure.	Reliability is concerned with stability in measuring whatever it measures.

5.	Low test validity reveals the presence of items unrelated to test objectives and measurement of constructs unrelated to the test objectives	Fluctuation in an individuals' score indicates low test reliability which reveals errors in the instrument.
6.	Validity of a test is estimated mainly through a statistical comparison of scores on the test with values on some outside variables	No outside variable is involved in estimation of reliability. It is only concerned with what a test can measure.

Similarities Between Test Validity and Reliability

- i. Validity and reliability are indispensable qualities a good test must possess
- ii. Both are concerned with measuring the attribute of whatever or whoever they are measuring, not the person or object itself.
- iii. Ambiguity of test items reduces both test validity and reliability.
- iv. Chance factors such as involvement in examination malpractice can affect both validity and reliability
- v. Valid test tends also to be reliable.

Conclusion

Psychometrically sound tests are indisputable tools in the measurement of educational or psychological attributes. As a result, ascertaining whether or not a test produces valid and reliable scores should be of great concern for test developers and users. Evidence of validity and reliability are not only crucial when developing new testing instruments but are also imperative when choosing instruments to be used in educational and psychological settings. Irrespective of the circumstances, maximum accuracy in measurement is always the goal. Thus, choosing the most valid and reliable measure of the construct of interest helps to ensure maximum accuracy.

References

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. New York: The Guilford Press.
- Borsboom, D. & Molenaar, D. (2015). Psychometrics. *International Encyclopedia of the Social and Behavioral Sciences*, 2(19), 418 – 422.
- Briere, J. I. (2011) *Determining the Psychometric Properties of the Retrieval-Induced Forgetting Procedure*. Unpublished Masters' Thesis, Department of Educational Psychology, College of Graduate Studies and Research, University of Saskatchewan, Saskatoon.
- Chakrabarty, S. N. (2013). Best Split-Half and Maximum Reliability. *IOSR Journal of Research & Method in Education*, 3(1), 1- 8.
- Cook, D. A. & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine*, 119 (2), 7 – 16.
- Creswell, R. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. California: Sage Publications, Inc.
- DeVellis, R. F. (2003). *Scale development: Theory and applications (2nd ed.)*. California: Sage Publications, Inc.
- Egereonu, A. C. (2009). *Measurement and evaluation in psychology*. Owerri: Career Publishers.
- Ezeh, D. N. (2003). Validity and reliability of tests. B. G. Nworgu (ed). *Educational measurement and evaluation: Theory and practice*. Nsukka: University Trust Publications.
- Iwuji, V. B. C. (1997). *Measurement and evaluation for effective teaching and learning*. Owerri: Joe Mankpa's Print.
- Kimberlin, C. L. & Winterstein, A. G. (2008). Validity and Reliability of Measurement Instruments Used in Research. *American Journal of Health-System Pharmacists*, 65(1), 2276 - 2284.
- Kulwinder, P. (2012). Reliability – Types, methods and usability. P. Kulwinder (ed). *Educational measurement and evaluation*. New Delhi: Usi Publications.
- Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational measurement (3rd ed.)*. New York, NY: Macmillan.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, Winter, 5 - 8.
- Mohajan, H. (2017). Two Criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University*, 17(3), 58-82
- Nwana, O. C. (1982). *Educational measurement for teachers*. Lagos: Thomas Nelson.

Nwana, O. C. (2007). *Textbook on educational measurement and evaluation*. Owerri: Bomaway Publishers.

Nkwocha, P. C. (2019). *Basic educational measurement and evaluation*. Owerri: Mercy Divine Publishers.

Ohiri, S. C. & Okoye, R. O. (2023). Application of classical test theory as linear modeling to test item development and analysis. *International Research Journal of Modernization in Engineering Technology and Science*, 5(10), 2152 – 2159.

Okoye, R.O. (2015). *Educational and psychological measurement and evaluation*, 2nd ed. Awka: Erudition Publishers.

Pallant, J. (2011). *A step by step guide to data analysis using the SPSS program: Survival manual*, (4th ed.). Berkshire: McGraw-Hill.

Rust, J. (2007). Discussion piece: The psychometric principles of assessment. *Research Matters*, 3, 25 – 27.