

THEME: THE THREE MOST IMPORTANT CONSIDERATIONS IN TESTING: VALIDITY, VALIDITY AND VALIDITY

**TOPIC: SURVEY OF TEACHERS' ATTITUDE TO THE VALIDITY OF
INSTRUMENTS USED FOR CONTINUOUS ASSESSMENT OF BASIC EDUCATION
IN NORTH CENTRAL GEO-POLITICAL ZONE OF NIGERIA**

**Paper presented at the 41st International Association for Educational Assessment (IAEA)
Conference hosted by the City of Lawrence, Kansas, and the University of Kansas' Center
for Educational Testing & Evaluation (CETE), 11 – 15 October 2015**

BY

**Prof. Charles M. Anikweze
Dept. of Educational Foundations, Faculty of Education
Nasarawa State University, Keffi, Nigeria
0803 666 4269; anikweze@yahoo.com**

Abstract

Continuous assessment (CA) is applied at all levels of education in the Nigerian school system but its implementation particularly at basic education levels is still fraught with hiccups. At federal level, attempts are being made to harmonize the implementation strategies for CA and particularly test development procedures to improve the validity of the instruments used by teachers for assessing learning, but the actualization remains an issue for perturbation. This survey sought to explore the attitude of basic education teachers towards establishing the validity of the instruments applied for continuous assessment of learners. Three research questions and two hypotheses guided the study. The multistage stratified random sampling procedure was adopted to select a sample of 1,500 teachers from three States of the North Central geo-political zone plus the Federal Capital Territory. A researcher-developed 25-item structured Likert scale was validated through expert consensus-based appraisal and used for data collection. Data were analysed using descriptive statistics for answering research questions and ANOVA and Phi correlation for testing the hypothesis at the 0.05 alpha. The results show that most of the professionally qualified teachers are sensitive to the relevance of validity for any test that will produce useful results. Nonetheless, only a negligible proportion of teachers irrespective of sex and level of educational qualification actually bothered to establish the validity of the instruments they use for obtaining CA scores. Findings from the study reflect the inevitable gap between policy prescription and its implementation. The dynamics of politics coupled with the diversity of ethnic nationalities and the associated multi-lingual confederacy seem to have acted as stumbling blocks to policy implementation. It is recommended that the Federal Ministry of Education should intensify efforts to harmonize the implementation strategies for CA particularly in ensuring the validity of the instruments used by teachers for assessing learning.

Key words: continuous assessment, rational validity, validation measures

Introductory Background

In the Nigerian school system, continuous assessment (CA) is applied, in part, for the determination of the overall achievement of learners at all levels of education, but particularly at

the basic education levels. By 1999, Nigeria introduced universal basic education which consists of 9 years of formal education comprising 3-year lower basic of junior primary schooling, 3-year middle basic of senior primary schooling and 3-year upper basic schooling which obtains in junior secondary schools.

Universal education at the basic levels is expected to be 'free and compulsory'. Universalization of education recognizes the influence of environmental variables on learning and the responsibility of the teacher to help as many learners as possible to gain from instruction. The assumption is that all individuals can attain mastery of a learning task provided they are given opportunities and time. Thus, with properly developed instructional sequence, every child could reach 100% mastery of any objective. Based on this assumption, the revised National Policy on Education (FRN, 2004) reviewed the weight of CA for basic education from 30% to 60% while the end of term examination accounts for 40% of the cumulated scores that determine the learners' ultimate annual achievement. Arguably, when CA is logically implemented, it is expected to encompass the developments of learners in the cognitive, psychomotor and affective domains following planned instructions. CA thus becomes an inevitable precursor of educational evaluation by which defensible judgment is passed on the worth of the learning that has occurred during the process of schooling. Indeed, with the heavy weight of 60% CA as policy directive, its implementation is rationally treated as fundamental to the demands of accountability from teachers. Therefore, the validity of the instruments used in conducting continuous assessment of learners is an issue of serious concern to responsible stakeholders in Nigerian education.

However, education in Nigeria is in the concurrent list; therefore the Federal Ministry of Education at best prescribes criteria for maintaining national standards while the State Ministries of Education implement the policy relying on their disaggregated variable manpower and technical expertise. Attempts have been made to harmonize the implementation strategies for CA but ensuring the validity of the instruments used by teachers for assessing learning as it obtains, remains an issue for perturbation.

Test validity connotes the appropriateness and comprehensiveness of the testing instrument in accurately measuring what it is supposed to measure. It establishes the meaning ascribed to the test score or assessment result as determined by evidence from the content, response processes, internal structure, and relationships to other variables and test consequences (Braden, 2009). The internal structure and response processes refer to reliability while the test consequences and relationship with other variables refer to usability. Thus, a good testing instrument should possess the qualities of validity, reliability and usability, in that order of critical importance (Anikweze, 2013). Most teachers in the Nigerian school system are involved in implementing CA; hence, as a matter of routine responsibility, they inevitably construct, administer, and score teacher-made tests. Whether consideration is given to validity issues is a matter of concern. What does validity mean to basic education teachers? Do teachers generally bother to establish the validity of the instruments they use in producing CA scores for their learners? What proportion of basic education teachers are knowledgeable about instrument validation? These and similar questions prompted conducting a survey to determine basic education teachers' attitude to validation of tests used for CA exercises.

Current Scenario

An annual review meeting of Registrars of States and FCT examination bodies took place on 21st April 2015. The purpose was to share experiences through the State reports on implementation of continuous assessment leading to the basic education certificate examination (BECE) at the end of the 9-year basic education. The thrust was towards harmonization of operations and synergizing for improvement and national standards. Already, most of the states in the country adopt similar strategies for the implementation of CA and BECE. Such States have computerised their operations and engage in standardization of scores and the good practice of giving regular orientation to teachers on the conduct of comprehensive CA, record keeping and standardization of scores to make them comparable across subjects. There is also the involvement of subject teachers in the generation of questions for the items bank at State level. Nevertheless, there are a number of questions whose answers might facilitate rational harmonization procedures. For instance:

- a. How are valid CA scores obtained on affective behaviours and character development since the National Policy on Education stresses that CA should encompass the cognitive, affective and psychomotor domains?
- b. What should be the best way of scheduling tests in various subjects to really make CA an aspect of formative evaluation FOR learning?
- c. How logical is the argument in favour of States implementing variable weights of CA and end-of term examination? Some States adopt 30% CA to 70% end-of term examination whereas the revised policy prescribed 60% CA to 40% end-of term examination.
- d. What is the justification for 35% - 40% raw score as pass mark practiced in some States? Can a student that scores over 40% from examination but having no CA score be promoted?
- e. How fair is it to deny re-sit candidates of graded performance in the basic education certification for failure to pass English Language and/or Mathematics at first instance? How can the potentials manifested in subjects other than English Language and Mathematics by such candidates be identified by institutions for further educational venture?
- f. Although learners at basic education are exposed to as many as 19 subjects, yet since only six credit passes including English language and Mathematics are required for basic education certification. What then is the justification for insisting that candidates take a minimum of 12 subjects and a maximum of 16 when nine would have sufficed?

Statement of the Problem

One of the greatest problems of education in Nigeria is the gap between policy formulation and the implementation of policy prescriptions. Of course, policy actors generally translate, adjust and work differently to implement policy prescriptions based on their convictions and preferences (Braun, Maguire and Ball, 2010). In the case of Nigeria, the dynamics of politics and the problems of diverse ethnic nationalities and the multi-lingual confederacy often act as stumbling blocks to policy implementation. The Universal Basic Education (UBE) programme as conceptualized by the democratic administration of Nigeria in 1999 is supposed to be compulsory from primary to the end of junior secondary school level. However, the implementation is still fraught with

thwarting problems and justified skepticism over the plausibility of its full implementation. Faithful implementation of the free, compulsory Universal Basic Education law was intended to, among others: (a) Improve education infrastructure, (b) Expand institutional capacity to produce quality manpower, and (c) Expand total school enrolment to increase the literacy level. More importantly, policy provides that assessment of learning achievement should be based on continuous assessment with a weight of 60% while the end-of-term examination accounts for 40% of the cumulated scores that determine the learners' ultimate annual achievement. When logically implemented, CA is expected to encompass the learners' developments in the cognitive, psychomotor and affective domains following planned instructions. But most teachers at the basic education level, as a matter of routine, develop and administer tests to learners on regular bases but scarcely do many of them give thought to issues of test validity. Could it be a matter of ignorance or one of attitudinal indifference? As contemporary educators express concern for validity of testing instruments, it becomes necessary to investigate the attitude of basic education teachers in Nigeria towards the validity of the tests they design for continuous assessment of their learners.

Research Questions

1. What proportion of the basic education teachers, segregated according to educational qualifications, in North Central geo-political zone of Nigeria have the correct conception of test validity?
2. How often do the teachers, moderated by sex, bother about establishing the validity of tests they use for assessing learners' achievement in public and private schools?
3. What is the frequency of tests given by basic education teachers in North Central geo-political zone of Nigeria?

Statement of the Hypotheses

H₀1: The level of educational qualification has no significant influence on teachers' conception of test validity as measured by their frequency of establishing the validity of tests they use for assessing learners' achievement.

H₀2: There is no significant relationship in the distribution of basic education teachers regarding the relevance of establishing test validity between public and private schools segregated according to sex.

Literature Review

Over the years, the term validity has attracted the attention of scholars whose conceptualizations differed as their disciplines and preferred emphases equally differed. In consequence, taxonomy of validity into various types has arisen depending on the purposes to be served and confirming the position that validity is always specific to a particular curriculum or a particular job (Thorndike and Hagen (1995). A distinction has also been made between traditional and modern concepts of validity (Messick, 1996a; Amy, 1999). The traditional means of accumulating validity evidence

have been categorized into three groups, namely: content-related, criterion-related, and construct-related evidence of validity (Amy, 1999).

Perhaps the commonest type of validity in the traditional parlance is *content validity* which pertains to the adequacy of the sample as representative of a domain. It is the extent to which an instrument measures what it is supposed to measure and performs as it is designed to perform. It is a matter of degree since it is practically impossible to attain 100% validity of an instrument (Biddix, 2009). Mehrens (1997) described the evidence regarding sample adequacy as *content validity evidence*. Construct validity emerges when one considers the extent to which two instruments that measure conceptually related properties agree. Construct validity is used to indicate the appropriateness of an inference using the adequacy of the behaviour as an indication of some hypothetical construct. For instance, a test has construct validity if it accurately measures a theoretical, non-observable construct or trait. Put as a question, construct validity asks whether scores or data obtained with the instrument relate to other qualities with which they are expected to be related. Attempts to provide answers to this question have given rise to convergent validity, either using the correlational approach or the contrasted groups approach, and divergent validity when the test scores are unrelated to behaviours and tests that they should be unrelated to. Convergent validity is the extent to which multiple attempts to measure the same concept are in agreement (Bagozzi, 1993). Thus, if two or more measures of the same concept highly correlated, it implies they are valid measures. Discriminant validity is the degree to which measures of different concepts are distinctively dissimilar.

Often the types of validity refer to terms that differentiated sample from sign inferences and predictability from scientific explanation inferences. Thus, we have Criterion validity referring to a standard by which a test or test performance may be judged. It is a set of scores, ratings, et cetera that a test is designed to measure, predict or correlate with. There are two types of criterion-related validity generally identified based on time reference. It becomes predictive validity when the score from a test, for whatever reason, predicted success in a future-oriented criterion. Shuttleworth (2009) posits that predictive validity is an important subtype of criterion and as a stalwart of science applied in educational and psychological employment tests to predict future performance. It becomes concurrent validity when it shows the extent to which different tests of the same property are in agreement.

Some individuals are promoting even more “progress” by suggesting that the validity of an assessment should be evaluated based on the consequences; hence the emergence of Consequential

validity. Shepard (1997) reasoned that based on the centrality of test use and consequences for validity, the debate should not be whether consideration of consequences is worthwhile but whether it should be an integral part of validity theory and practice. She contended that it is essential to evaluate test validity through the examination of the effects following from the use of test. However, the elevation of consequential validity has been described as “right concern of a wrong concept (Popham, 1997).

Messick (1996b) distinguished six aspects of validity that serve as the reference points for the modern conceptions of validity. These are:

1. The content aspect as determined by the knowledge, skills, and other attributes to be revealed by the assessment tasks.
2. The substantive aspect whose emphasis is on the verification of the domain processes to be revealed in assessment tasks.
3. The structural aspect of construct validity that insists on scoring models being consistent with the structural relations inherent in behavioral manifestations of the construct in question (Loevinger, 1957).
4. Generalizability aspect that emphasizes representativeness of content coverage and processes of the construct domain.
5. The external aspects of validity that stresses conformance between assessment scores and interactive relations implicit in the specified construct.
6. Consequential aspects of validity which suggests that the validity of an assessment should be evaluated based on the consequences.

However, Mehrens (1997) suggested that the psychometric community should narrow the use of the term validity rather than expand it. To him, the term should be reserved for determining the accuracy of inferences about (and understanding of) the characteristic being assessed, not the efficacy of actions following assessment. In any case, Wolming and Wikström (2010) have identified a gap between validity theory based on modern perspectives and practice. They concluded that validity theory is difficult to put into practice; therefore they suggested a need for guidance on how to prioritize validity questions and the interpretation of validity evidence.

With regard to continuous assessment carried out by basic education teachers in Nigeria, a few reviewed literatures suggest that teachers are generally competent in the development of tools for assessing the cognitive domain but deficient in psychometric properties and the construction

of non-cognitive measures (Ugodulunwa and Musa, 1996; Anakudo, 2001). Bulus (2007) investigated the perceived competence of secondary school teachers in Jos area of Plateau State in Nigeria and found that teachers were more competent in constructing cognitive tests than in developing non-test instruments and only few teachers could analyse and interpret scores using higher order statistical tools such as standard deviation, percentile rank and z-scores.

Recently, in a cross-sectional survey of a sample of 300, Jugbo (2015) examined the assessment competence of primary school teachers in Nasarawa North senatorial district of Nasarawa State. Using questionnaire and rating scale validated through consensus-based appraisal, she obtained information on the teachers' test construction capabilities for the continuous assessment of their learners. Her findings indicated that teachers were generally competent in the development of cognitive and psychomotor tools; yet a significant difference existed among the teachers segregated by academic and professional qualifications. Expectedly, better qualified teachers with professional qualification in teaching were more competent in test development than graduates without teaching qualification. The hypothesis that assessment competence of teachers will not differ significantly due to their educational qualification was rejected ($p > 0.05$; 299df; $F_{cal.} = 4.44 > F_{cri.} = 2.10$). This confirms an earlier finding by Anakudo (2001) that the assessment competence of teachers is determined by their educational qualification. Jugbo (2015) also found out that primary school teachers were deficient in development and use of affective tools.

Methodology

The study was a cross-sectional survey of basic education teachers in North Central geo-political zone of Nigeria. The study area covers six States and the Federal Capital Territory (FCT) situated in the middle belt region of the country. The population of the study consisted of 169,437 registered basic education teachers in six states of the North Central geo-political zone of Nigeria plus the FCT which constitute the Middle Belt region (TRCN, 2014).

Multistage stratified random sampling strategy was applied to select a sample of 1,500 teachers from 30% of the Local Government Areas in the three randomly selected States of Kaduna, Plateau, and Nasarawa. Stratification enabled selection of samples from both public and private schools. The sample size was based on a little less than 9% sample size precision of population estimates on the 0.05 alpha (Cozby, 2001). All selected teachers for the study teach at basic education level. Randomization enabled teachers of both sexes to be selected as well as research participants with different educational qualifications.

The instrument for data collection consisted of the researcher-developed 20-item structured Likert scale. Content validation of the instrument was through the logical and quantified appraisal of experts whose consensus gave a mean rational validity index of 0.89. Further validation was obtained through trialing using a pilot sample of 30 teachers randomly drawn from Keffi North educational area which was part of the population but outside the study sample. The analysis of the obtained data gave internal consistency index of 0.87 using Cronbach Alpha method.

The administration of the instrument was facilitated by research assistants consisting of part-time postgraduate students of the Faculty of Education, Nasarawa State University, Keffi. By adopting the wait-and-take method, 100% return ratio was assured. Data were analysed using descriptive statistics for answering research questions and the ANOVA for testing the first null hypothesis and Phi correlation for testing the second null hypothesis, both at the 0.05 alpha.

Results

On the first research question: *What proportion of the basic education teachers, segregated according to educational qualifications, in North Central geo-political zone of Nigeria have the correct conception of test validity?* Data in Table 2 in the Appendix suggest that a good proportion of the teachers (over 60%) have a positive concept of what test validity implies. Indeed, 70% of the respondents agreed that every test used for assessing learners ought to be valid. Similarly, 66% agreed that a valid test should not contain any errors or mistakes, even as 68% expressed the conviction that a test should be long enough to be valid and short enough to be reliable. It is also noted that 76% rejected the proposition that validity is necessary for standardized test not for teacher-made test. It is equally noteworthy that only 25% imagined that reliability of a test is more important than its validity.

Hypothesis 1: *The level of educational qualification has no significant influence on teachers' conception of test validity as measured by their frequency of establishing the validity of the tests they use for assessing learners' achievement.* This hypothesis was tested using the ANOVA statistic. The evidence from Table 3 in the Appendix shows that the calculated F ratio of 0.712 is less than 2.60 (the F critical) at the 0.05 level of significance. The hypothesis was therefore accepted since the teachers generally manifested positive conceptualization of validity irrespective of differentials in their educational qualifications.

Research Question 2: *How often do the teachers, moderated by sex, bother about establishing the validity of tests they use for assessing learners' achievement in public and private schools?*

This research question was answered by considering the extent of relevance the teachers attached to issues of test validity in their assessment practices. Table 4 in the Appendix portrays the results. Evidence from the study shows that the male teachers give greater consideration to the validity of tests they use for CA than their female counterparts. Findings indicate that 81% of male teachers as against 39% of females make use of table of specification to obtain content validity. Similarly, 75% of male teachers as against 58% of females claimed to obtain face validity of their tests and still obtained their reliability through trialing. To corroborate this claim is the finding that 87% of male teachers as against 37% of their female colleagues show their tests to their heads of department to determine the appropriateness or otherwise. However, both male (72%) and female (74%) teachers seem to lay equal claim to injecting validity into their tests by sticking to what they have taught the students. But whereas 92% of male teachers bothered about the validity of their tests as much as they considered their reliability, only 59% of the female teachers felt bothered.

With regard to public/private dichotomy, the study suggests no clear distinction in both perspectives and practices. For instance, for establishing content validity with the use of table of specification or blueprint, it is 62% public and 50% private; for botheration about the validity of tests as much as consideration for their reliability, it is 80% public and 84% private; for the demands of continuous assessment as an impediment against establishing the validity of tests, it is 38% public and 37% private; and for expressed ignorance about test validation, it is 39% public and 37% private.

Test of Hypothesis 2: *There is no significant relationship in the distribution of basic education teachers regarding the relevance of establishing test validity between public and private schools segregated according to sex.* Phi correlation (ϕ) was employed to test the association between sex and school type as influencing variables regarding the relevance that teachers accord to test validity. The choice of ϕ statistic was because both variables are genuinely dichotomized. Based on the details shown in Table 5 in the Appendix, the obtained ϕ index of 0.094 represents very negligible association. Therefore, school type could not be relied upon to explain the extent of relevance that teachers give to test validity when segregated by sex.

Research Question 3: *What is the frequency of tests given by basic education teachers in North Central geo-political zone of Nigeria?* Based on the data portrayed in Table 6 in the Appendix, 32% of the teachers claimed that they administer one test per week. Another 34% of the respondents conduct tests fortnightly. It is doubtful if these two groups would give sufficient

validity considerations to incessant tests. Another 23% of teachers administer tests once per month. This group of teachers has sufficient time to construct and validate their tests before administration. Only 11% acknowledged that they construct tests used for continuous assessment once per term. This is grossly insufficient for formative evaluation of learning.

Discussion of the Findings

Findings from the study indicate that there is no significant difference among the teachers concerning the concept of validity based on educational qualifications. This is contrary to the findings of Anakudo (2001). However, it is one thing to have awareness of a concept and an entirely different thing to apply the concept. There is evidence that professionally qualified teachers showed greater concern for test validity than graduates without professional qualification in education. This agrees with the findings of Jugbo (2015).

Part of the findings indicate that generally few teachers give thought to instrument validation partly out of ignorance of validation measures and partly due to poor skill in test construction meant for ‘assessment FOR learning’. It is equally probable that there was wrong conceptualization of CA implementation methods as identified by Ugodulunwa and Musa, (1996). About 80% of male teachers claimed that they make use of table of specifications for constructing tests, yet as much as 73% confessed ignorance of how to validate testing instruments. What then is the use of test blue print if not for ensuring content validity? It is probable that many teachers lack faith in their peers’ ability to provide dependable consensus-based rational validity. Evidence from the study suggests that male teachers bother about the validity of tests used for continuous assessment of learners more than their female counterparts.

Findings from this study suggest compulsive adherence to certain frequency of tests often imposed by the officials of State Ministries of Education. This is compounded by the operational difficulty associated with assessing affective and psychomotor behaviours.

Conclusion

Many teachers of basic education in North Central geo-political zone of Nigeria are yet to be abreast of validity issues in test construction. Findings from the study reflect the inevitable gap between policy prescription and its implementation. The dynamics of politics coupled with the diversity of ethnic nationalities and the associated multi-lingual confederacy seem to have acted as stumbling blocks to policy implementation.

Recommendations

Based on the findings from this study, it is recommended that the Federal Ministry of Education should intensify efforts to harmonize the implementation strategies for CA particularly in ensuring the validity of the instruments used by teachers for assessing learning. It is also recommended that greater emphasis be given to the production of professional assessors and evaluators to take charge of measurement and evaluation courses in teacher education institutions so that fresh graduates of education will be masters of valid test construction. For the existing teachers in the field, there should be regular workshops to update them with strategies for effective CA implementation strategies and item writing.

References

- Amy, B. (1999). *Traditional and Modern Concepts of Validity*. ERIC Clearinghouse on Assessment and Evaluation Washington DC.. Retrieved on 16th July 2015 from <http://www.cehd.umn.edu/EdPsych/rips/Documents/Measurement/TraditionalandModernConceptsofValidity>
- Anakudo, E. N. (2001). Perceived Competency of Secondary School Teachers in the Operation of Continuous Assessment in Jos North, Plateau State. Unpublished M. Ed. Thesis, University of Jos, Nigeria.
- Anikweze, C. M. (2013). *Measurement and Evaluation for Teacher Education, 2nd Edition*. Ibadan: Constellation (Nigerian) Publishers.
- Bagozzi, R. P. (1993). Assessing Construct validity in Personal Research: Applications to Measures of Self-esteem. *Journal of Research in Personality*, 27, 49-87.
- Biddix, J. P. (2009). Instrument, Validity, Reliability: Uncomplicated Reviews of Educational Research Methods. Retrieved on 16th April 2015 from <https://researchrundowns.wordpress.com/quantitative-methods/instrument-validity-reliability/>
- Braden, Jeffery (2009). Historical and contemporary definitions of validity: critical issues in assessment validity. Retrieved on 14th March 2015 from <http://www.education.com/reference/article/validity/>
- Cozby, P.C. (2001). *Methods in Behavioural Research, 7th Ed*. Boston: McGraw-Hill
- Federal Republic of Nigeria (2004). *Compulsory, Free Universal Basic Education Act, 2004; Official gazette of the Federal Republic of Nigeria*, No 66, Vol. 91, A113 – A124. Lagos: Federal Government Press.
- Federal Republic of Nigeria (2014). *National Policy on Education*. Abuja: Nigerian Educational, Research and Development Commission
- Jugbo, Joyce (2015). Assessment of Primary School Teachers' Competence in test Construction for Continuous Assessment in Nasarawa North Senatorial Zone. Unpublished M. Ed. Dissertation submitted to the Faculty of Education, Nasarawa State University, Keffi, Nigeria.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. In B. Amy: *Traditional and Modern Concepts of Validity*. ERIC/AE Digest

- Braun, A., Maguire, M., & Ball, S. J. (2010). Policy enactments in the UK secondary school: examining policy, practice and school positioning. *JOURNAL OF EDUCATION POLICY*, 25(4), 547 - 560. [10.1080/02680931003698544](https://doi.org/10.1080/02680931003698544)
- Mehrens, W. A. (1997). The Consequences of Consequential Validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8, 13, 24.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement, 3rd Ed.*, (pp. 13-103). New York: Macmillan.
- Messick, S. (1996a). Standards-based score interpretation: Establishing valid grounds for valid inferences. Proceedings of the joint conference on standard setting for large scale assessments, Sponsored by National Assessment Governing Board and The National Center for Education Statistics. Washington, DC: Government Printing Office.
- Messick, S. (1996b). Validity of Performance Assessment. In Philips, G. (1996). *Technical Issues in Large-Scale Performance Assessment*. Washington, DC: National Center for Educational Statistics.
- Popham, W. 3. (1997). Consequential validity: Right concern-wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8, 13, 24.
- Shuttleworth, M. (2009). Predictive validity and University Selection. Retrieved Nov.18, 2014 from <https://explorable.com/predictive-validity>
- Teachers Registration Council of Nigeria (2014). Statistical Digest of Teachers in Nigeria. Abuja: Author
- Thorndike, R.L. & Hagen, Elizabeth (1995) *Measurement and Evaluation in Psychology and Education*, (5th Ed.). New York: John Willey & Sons
- Wolming, S. & Wikström, C. (2010). The concept of validity in theory and practice. In *Assessment in Education: Principles, Policy & Practice*, 17(2). 117-132

Appendix A

Table 1: Distribution of the target population by State and Sex

S/N	States	No of Basic Education Teachers		
		M	F	MF
1.	Benue	21,359	12,559	33,918
2.	Kaduna	16,648	14,538	25,186
3.	Kogi	18,294	10,941	29,235
4.	Nasarawa	9,105	4,376	13,481
5.	Niger	10,990	8,206	19,196
6.	Plateau	10,636	9081	19,717
7.	FCT	7,612	13,455	21,067
Total		88,305	81,132	169,437

Source: Statistical Digest of Teachers in Nigeria, TRCN (2014)

Table 2: Distribution of Respondents by Highest Qualification and Concept of Validity

The Items	Distr. by Highest Qualification and Proportion with concept of Validity				
	NCE	Degree/HND + Educ.	Degree/HND Only	PG degree in Educ.	Total
N	375	555	400	170	1,500
Every test used for assessing learners ought to be valid	210 (56%)	525 (95%)	210 (52%)	100 (59%)	1,045 (70%)
Valid tests should have comprehensive items based on topics taught to learners	280	355	170	70	875 (58%)
Valid tests need consider the topics in the syllabus not the level of the students	130	395	125	145	795 (53%)
The language of valid tests should be appropriate to the level of the students	260	380	185	127	952 (64%)
A valid test should not contain any errors or mistakes	220	475	145	150	990 (66%)
Errors in a test affect its reliability not the validity	100	155	140	50	445 (30%)
A test should be long enough to be valid and short enough to be reliable	325	370	185	140	1020 (68%)
Validity is necessary for standardized test not for teacher-made test	95	140	90	70	395 (26%)
Reliability of a test is more important than its validity	130	110	105	30	275 (25%)
Talking about test validity should be for only mathematicians	175	220	115	55	565 (38%)

Table 3: ANOVA for the influence of educational qualifications on teachers' concept of validity

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	46.776	3	15.592	.712	.545
Within Groups	32768.741	1496	21.904		
Total	32815.517	1499			

Table 4: Relevance of test validity to the teachers by Gender and School Type

The Items	Gender			School Type		Total
	Male	Female	Total	Public	Private	
N	625	875	1,500	715	785	1,500
I always base my tests on a table of specification or blueprint	505 (81%)	340 (39%)	845 (56%)	380 (53%)	485 (62%)	865 (50%)
I take care in constructing tests but I don't know how to validate the tests that I set	430 (69%)	190 (22%)	620 (73%)	225 (31%)	310 (39%)	535 (37%)
I know how to obtain validity of my tests through consensus-based appraisal by experts but there is no time for that	270 (43%)	295 (34%)	565 (38%)	310 (43%)	415 (53%)	725 (48%)
I obtain face validity of my tests and still obtain their reliability through trialing	470 (75%)	505 (58%)	975 (65%)	330 (46%)	485 (62%)	815 (54%)
I inject validity into my tests by sticking to what I have taught the students	450 (72%)	645 (74%)	1095 (73%)	440 (62%)	635 (81%)	1075 (72%)
The demands of continuous assessment does not allow me to insist on establishing the validity of my tests	345 (55%)	205 (23%)	550 (37%)	260 (36%)	300 (38%)	560 (37%)
I show my tests to my HOD who determines whether they are appropriate or not	545 (87%)	325 (37%)	870 (58%)	395 (55%)	480 (61%)	875 (58%)
I bother about the reliability of my tests as much as I consider their reliability	575 (92%)	520 (59%)	1095 (74%)	625 (87%)	630 (80%)	1255 (84%)
I copy questions from textbooks, so validity should be the responsibility of the authors	365 (58%)	160 (18%)	525 (35%)	215 (30%)	330 (42%)	545 (36%)

Table 5: Phi Correlation test of significant relationship between male and female teachers segregated by school location regarding the relevance of establishing test validity

		SCHOOL TYPE		
		Private	Public	Total
SEX	Male	423 (a)	452 (b)	875
	Female	362 ©	263 (d)	625
	Total	785	715	1,500

The formula for computing the Phi correlation is
$$\phi = \frac{bc - ad}{\sqrt{(a+b)(c+d)(b+d)(a+c)}}$$

$$\begin{aligned} \text{Substituting } \phi &= \frac{452 \times 362 - 423 \times 263}{\sqrt{(423 + 452)(362 + 263)(452 + 263)(423 + 362)}} \\ &= \frac{163624 - 111249}{\sqrt{(875)(625)(715)(785)}} = \frac{52375}{\sqrt{546875 \times 561275}} = \frac{52375}{749.18 \times 739.51} = \frac{52375}{554026.08} = 0.094 \end{aligned}$$

Table 6: Frequency distribution of the number of tests used by basic education teachers for CA

S/N	Description of items	Frequency	Percent	Cumulative %
1.	One test per week	485	32.3	32.3
2.	One test fortnightly	505	33.7	66.0
3.	One test per month	350	23.3	89.3
4.	One test per term	160	10.7	100.0
5.	Total	1500	100.0	