

What Teachers Know About Validity of Classroom Tests: Evidence from a University in Nigeria

Christiana Amaechi Ugodulunwa¹, Sayita G. Wakjissa²

^{1,2}*Department of Educational Foundations, Faculty of Education, University of Jos, Nigeria*

Abstract: *This study investigated university teachers' knowledge of content and predictive validity of classroom tests. A sample of 89 teachers was randomly selected from five departments in the Faculty of Education of a university in Nigeria for the study. A 41-item Teachers' Validity Knowledge Questionnaire (TVK-Q) was developed and validated. The data collected were analysed using descriptive statistics, t-test and ANOVA techniques. The results of the analysis show that the teachers have some knowledge of content-related evidence, procedure for ensuring coverage and adequate sampling of content and objectives, as well as correlating students' scores in two measures for predictive validation. The study also reveals that the teachers lack adequate knowledge of criterion-related evidence of validity, concept of face validity and sources of invalidity of test scores irrespective of their gender, academic disciplines, years of teaching experience and ranks. The implications of the results were discussed and recommendations made for capacity building.*

Keywords: *validity, teachers' knowledge, classroom tests, evidence, university, Nigeria*

I. Introduction

There is increased desire for school effectiveness and improvement at all levels of education in Nigeria. The central role of assessment in the school system for improving teaching and learning therefore demands that classroom tests should be valid and reliable measure of students' real knowledge and skills and not testwiseness or test taking abilities. Classroom tests are made and used by teachers for formative and summative purposes. They serve formative purpose when the results are used to monitor teaching and learning progress and summative purpose when the results are used for grading, promotion and certification. In Nigeria, teachers are required to adopt continuous assessment mode of evaluation and the most commonly used techniques are written tests, performance tests and projects at the university level. These assessment techniques are expected to be valid and reliable measures of abilities. The extent to which the university teachers are able to develop and use valid assessment instruments depends on what they know about validity to help ensure quality in classroom assessment.

At the university level, different techniques are adopted in operation of continuous assessment mode of evaluation by the teacher. These include assignments, projects, term paper, quizzes, tests, practical examinations, oral and end of semester examinations. The end of semester examinations are internally moderated, while the final year examinations are both internally and externally moderated for quality assurance. The internal mechanisms for assuring quality consist of those of academic departments, faculties, senate and external examiner system (National Universities Commission, 2002). The external examiner provides additional assurance that quality of academic programmes of a university is acceptable to academic peers across the university system in the country; while the university's examination regulations guide the conduct and processing of examination materials and results (Author, 2005). In the entire process of quality assurance in assessment, the issue of validity is paramount and critical.

Validity is a concept that requires to be fully understood by teachers in any effort to improve quality of classroom tests. It is the most important technical quality of an assessment instrument. Validity is a general term that refers to the extent to which an instrument measures what it claims to measure. This definition implies that for an instrument to be valid, it must measure what it purports to measure and nothing else. Validity also refers to the appropriateness of inferences drawn from test scores or other assessment instruments. According to Rudner and Schafer (2002), test validity refers to the degree with which the inferences based on test scores are meaningful, useful and appropriate. This definition implies that validity is expressed in degree from low, moderate to high. It is not measured but inferred from available evidence and depends on many types of evidence. Validity requires that an instrument is reliable but an instrument can be reliable without being valid. Most literature (Author, 2008; Professional Testing Inc, 2006; Rudner & Schafer, 2003) identifies different types of validity based on scope, relevance, predictive quality and association. These types of validity are content validity, construct validity, criterion-related validity and face validity. However, content validity is considered to be of most importance for classroom teachers. Also important is criterion-related validity since teachers not only conduct assessment for learning and of learning, as well as prepare students for labour market.

Content validity requires judgment of experts to determine if the test is representative of knowledge and skills that are supposed to be measured. This involves consistency between curriculum content, test objectives and content of the test. The degree of content validity depends on the test's coverage of necessary objectives and content as well as adequate sampling of important curriculum content. This is what Popham (1990) refers to as item relevance and content coverage. Item relevance and content coverage help in providing evidence from which valid inferences can be drawn. Constructing table of specifications is one of the practical ways of achieving content validity of a test. On the other hand, predictive validity is the extent to which a student's current performance on a test estimates the student's later performance on a criterion measure. Although face validity is not a type of validity in a technical sense, it is the degree to which an instrument appears to measure what it measures. It is usually confused with content validity by teachers. Important evidence of validity to be sought by teachers is content-related evidence and criterion-related evidence.

It has been argued that the traditional conception of validity is fragmented and incomplete because it fails to take into account evidence of meaning of scores as a basis of action and the social consequences of the use of score (Messick, 1996). Messick's modern concept of validity views it as a unified concept, which lays more emphasis on the use of a test. He identified six aspects of validity that are implicit in the notion of validity as a unified concept. The six aspects are content, substantive, structure, generalizability, external factors and consequential. The six aspects are viewed as interdependent and complementary forms of validity evidence and not separate entities. These imply that evidence for assessing validity should include evidence of content relevance and representativeness, extent to which scores are consistent with theoretical predictions, evidence on extent to which scores and their interpretations generalize to and across groups, settings and tasks. Other evidence are the fidelity of scoring structure to the structure of the construct being assessed; evidence from criterion-related studies, and consequential aspect of test use and score interpretation especially the issues relating to bias and fairness.

It is clear from the views of Messick that one cannot validate a test but can only validate the inferences that are drawn from students' scores in the test as observed by Killen (2003). This point is usually overlooked by teachers who think about validity of their questions and not inferences drawn from them. Teachers concern should therefore not be to develop valid tests but to develop tests that will provide evidence from which valid inferences can be drawn about students' learning. This is an important challenge for teachers in order not to ignore the most characteristics of assessment, which is appropriate the use of test results in making instructional decisions as observed by Killen (2003).

Considering the importance of validity in classroom testing, teachers need to possess necessary knowledge of validity and how to gather validity-related evidence for their tests and other forms of assessment to enable them draw valid inferences and take relevant decisions based on students' assessment scores. Unfortunately, it has been observed (Dosumu, 2002; Adeola & Fajonyimi, 1999; Imo, 2012; Agu, Onyekuba & Anyichie, 2013) that teachers lack test construction skills. Most teachers receive little or no training or support after certification. Although, teachers are not expected to be experts in educational measurement and evaluation to construct valid and reliable tests, they need some basic knowledge on how to develop and validate their classroom tests to enable them use results of their classroom assessment for taking relevant decisions about the students. The situation may be worse at the university level where most lecturers did not have any formal training in educational assessment, except those in the faculty of education. Most of the efforts are usually directed to assessment of teachers' competence in test construction at the primary and secondary levels. There is therefore need to find out what university teachers know about validity of classroom tests in order to provide baseline data for capacity building in test development and validation for the quality assurance in assessment of learning outcomes. This study was designed to find out aspects of validity teachers are knowledgeable in and the aspects they lack knowledge, find out if their responses differ due to gender and academic discipline and the areas they need capacity building.

II. Method

The study was a survey research that was geared towards identifying what university teachers know about validity of classroom tests. The sample of study consisted of 89 university teachers (46 male; 43 female) drawn using random sampling technique from a population of 165 teachers in the Faculty of Education of a university in Nigeria. The 89 teachers belong to five academic disciplines (Arts Education=13; Educational Foundations=31; Science and Technology Education=23; Social Science Education= 6; Special Education and Rehabilitation Sciences=16). The teachers' ranks vary from Assistant Lecturer cadre to Professorial cadre (Assistant Lecturer=7; Lecturer II= 13; Lecturer I =35; Senior Lecturer= 16; Reader/Professor= 18). Their years of teaching experience at the university level also vary (1-5years=20; 6-10years=25; 11-15years=17; 16-20years=8; 24years and above= 19).

The instrument for data collection was the Teachers' Validity knowledge Questionnaire (TVK-Q) developed and validated by the researchers. The TVK-Q consists of 41 items that are organized in three

sections, A, B and C. Section A consists of four items that sought information on teachers' gender, academic discipline, years of teaching experience and rank. Section B consists of 30 items, which sought information on the teachers' knowledge of different aspects of validity. Section C consists of seven items, which required teachers responses on possible aspects of validity they need capacity building. The teachers were required to make their responses in Section B using a 4-point modified Likert Scale: Strongly Disagreed (SD)=1, Disagree (D) =2, Agree (A) =3, and Strongly Agree (SA)=4. Eighteen out of the 30 items are positively structured while 12 were negative statements about validity of classroom tests. High mean score for a negative statement was regarded as lack of knowledge of validity, while high mean score on positive statements was considered as aspects of validity that teachers know. The mean responses of teachers were categorized as follow: 1.00-1.49= SD; 1.50-2.49= D; 2.50-3.49= A; and 3.50- 4.00= SA. The instrument was subjected to the scrutiny of two experts in educational measurement and evaluation to judge the appropriateness and clarity of the items. The items were considered adequate for assessment of knowledge of validity of classroom tests. The TVK-Q was administered to the teachers within two weeks.

The coefficient of reliability of the instrument was 0.76; obtained using split-half method and correction for length using Spearman Brown formula. The Cronbach coefficient alpha obtained was 0.79. The instrument was therefore considered to be a reliable measure of teachers' knowledge of validity of classroom tests. The data collected for profiling teachers' knowledge of issues relating to validity were analysed using mean and standard deviation. Those for testing hypothesis of no gender difference in the teachers' responses were analysed using t-test for independent samples; while significant differences due to academic disciplines, ranks and years of teaching experience were tested using one way analysis of variance (ANOVA) statistical technique. The results of the analyses are presented in Tables 1 to 7.

III. Results

Table 1 Mean and Standard Deviation of Profile of Teachers' Knowledge of Validity of Classroom Tests

S/N What teachers know about Validity of test	Mean	SD
1. Validity indicates how well a test measures content and objectives it is supposed to measure	4.55	0.74
2. Tasks assessed for validity does not include extraneous factors	2.80*	1.20
3. Validity is not about a test itself but about meaning of test scores	3.40*	1.30
4. Even if a test is reliable, it may not provide valid results	3.66	1.57
5. Test validity is about appropriateness of inferences made about test scores	3.66	1.22
6. Content validity is usually established by experts	4.43	0.72
7. Content validity deals with adequate sampling of course content and objectives of assessment	4.25	0.98
8. A table of specifications relates content to objectives according to their relative importance	2.76*	0.93
9. Using Bloom's taxonomy in developing tests helps assure content validity	4.18	1.04
10. Criterion-related evidence is the extent to which scores on two tests are in agreement	3.43*	1.39
11. Predictive validity of continuous assessment should be determined by correlating it with examination scores	3.61	1.09
12. It is necessary for predictive validity of University Post JAMB examination to be established	3.88	1.02
13. Criterion-related validity involves comparing scores on a test with performance in another current or later measure of performance	3.42*	1.35

Table 1 shows the 13 aspects of validity the teachers are knowledgeable in. The mean of their responses ranged from 2.76 to 4.55 which were above the midpoint of the scale (2.50) between agreeing and disagreeing with different aspects of validity. The standard deviation ranged from 0.72 to 1.57, which shows that their responses were not too variable. The aspects of validity the teachers agreed that they were knowledgeable in are items 2, 3, 8, 10 and 13; while the other items were the ones they strongly agreed that they know. Generally, the aspects of validity the teachers know were those that deal with content-related evidence, procedure for ensuring coverage and adequate sampling of content and objectives, as well as correlating students' scores in two measures for predictive validation.

Table 2 Mean and Standard Deviation of Profile of Teachers' Lack of Knowledge of Validity

S/N What Teachers do not know about validity of tests	Mean	SD
1. Negative consequences of assessment should not be minimized in		

evaluating validity of test	3.92	1.31	
2. A valid test is should actually reflect test taking skills		2.58*	0.94
3. Validity is not a matter of degree		2.84*	1.32
4. Validity is not the most important quality of a test		3.89	1.10
5. A valid test should not consider background of testees		4.06	1.22
6.Face validity is the same as content validity		4.13	1.17
7. Assuring face validity is not necessary in classroom assessment		3.58	1.41
8. Content validity is expressed as coefficient		3.00	1.57
9.A table of specification is not necessary for achieving representativeness of a test	3.90	1.40	
10. Content validity is not primarily judgmental		3.52	1.19
11. Moderating examination questions at unit or departmental level is not a way of evaluating coverage of content and objectives taught		4.38	1.03
12. Predictive validity is not concerned with correlation of classroom tests with criterion measure of performance at a later time		3.30	1.66
13. It is not necessary for continuous assessment scores of students to be in agreement with their examination scores		3.96	1.10
14. Predictive validity is not a necessary quality desired for a classroom assessment		3.67	1.21
15. Predictive validity is judgmental		2.99*	1.11
16. Predictive validity does not allow a teacher to determine how well a student will perform in future		3.78	1.26
17. Predictive validity does not involve statistical analysis		3.44*	1.46

Table 2 shows the 17 aspects of validity in which the teachers lack know. The mean of their responses ranged from 2.58 to 4.38 which were above the midpoint of the scale (Mean=2.50) between agreeing and disagreeing with different aspects of validity. The standard deviation ranged from 0.94 to 1.66, which shows that their responses were not too variable. The aspects of validity the teachers agreed that they were not knowledgeable in are items 2, 3, 15 and 17; while the other items were the ones they strongly agreed that they do not know as shown in Table 2. Generally, the aspects of validity the teachers do not know were those that deal with criterion-related evidence, concept of face validity and sources of invalidity of test scores.

Table 3Result of t-test analysis of Gender difference in Teachers' knowledge about Validity

Group	n	Mean	SD	df	t _{value}	P _{value}
Male	46	108.20	15.63	87	0.419	.679
Female	43	106.91	13.19			

$p > .05$

Table 3 shows there is no significant gender mean difference in the responses of the teachers about their knowledge of validity of classroom test. Mean differences in their knowledge of aspects of validity of classroom test due to academic disciplines was conducted and the results are presented in Tables 4. Post hoc comparison of means of the five groups was conducted using Scheffe method but there was no significant mean difference between the groups.

Table 4Results of ANOVA for Mean difference in Teachers' Knowledge due to Academic Discipline

Source	Sum of Squares	df	Mean Square	F	P _{value}
Between Groups	292.157	4	73.039	3.40	.850
Within Groups	1804.618	84	214.853		
Total	18339.775	88			

Tables 5 shows no significant mean difference on what the teachers know about validity due to their years of teaching at the university level. Post hoc comparison of the means of the five groups was not significant too.

Table 5Results of ANOVA for Mean difference in Teachers' Knowledge due to Years of Teaching

Source	Sum of Squares	df	Mean Square	F	P _{value}
Between Groups	1529.460	4	382.365	1.911	.116

Within Groups	16810.315	84	200.13
Total	18339.775	88	

p > .05

Tables 6 shows no significant mean difference on what the teachers know about validity due to their academic ranks. Post hoc comparison of the means of the five groups was not significant too.

Table 6 Results of ANOVA for Mean difference in Teachers' Knowledge due to Academic Ranks

Source	Sum of Squares	df	Mean Square	F	<i>p</i> _{value}
Between Groups	338.197	4	84.549	0.395	.812
Within Groups	18001.579	84	214.305		
Total	18339.775	88			

p > .05

The teachers' responses on areas they need capacity building in validating classroom assessments were analysed. The percentage responses of the order of ranks are presented in Table 7. Six areas of need for capacity building were identified. The areas of greatest need were valuating validity of assessment instruments, determining predictive and content validity of classroom assessments.

Table 7 Percentage Responses on Aspects of Validity that the Teachers Need Capacity Building (n=89)

S/N	Aspects of Validity	n	%	Rank
1.	Evaluating validity of assessment instruments	61	68.50	1
2.	Determining predictive validity of assessment instruments	57	64.00	2
3.	Determining content validity of assessment instruments	45	50.60	3
4.	Determining Face validity of assessment instruments	43	48.30	4
5.	Understanding and adherence to item writing rules	41	46.10	5
6.	Writing questions that cover Bloom's taxonomy of educational objectives	36	40.00	6

IV. Discussion

The study found that teachers know some aspects of content-related evidence of validity, procedure for ensuring coverage and adequate sampling of content and objectives, as well as correlating students' scores in two measures for predictive validation. This is not surprising because the teachers participate in internal moderation of examination questions at the Departmental Examination Board meetings, where relevance and content coverage were usually emphasized and where reports of external examiners' reports are read. However, their lack of adequate knowledge about aspects of validity that deal with criterion-related evidence, concept of face validity and sources of invalidity of test scores together with their high ratings on the need for capacity building in the areas of evaluating validity of assessment, predictive and content validity assessment require that the teachers be trained and retrained. This is consistent with the view of Popham (1990) that item relevance and content coverage help in providing evidence from which valid inferences can be drawn. Since it has been argued that the traditional conception of validity is fragmented and incomplete because it fails to take into account evidence of meaning of scores as a basis of action and the social consequences of the use of score (Messick, 1996) retraining of teachers through workshops and seminars becomes imperative. Messick's modern concept of validity as a unified concept, which lays more emphasis on the use of a test should provide framework for such capacity building not only for education teachers but also those who are not professional teachers. Furthermore, the teaching and learning of educational measurement and evaluation courses should emphasize validity of classroom assessments. Efforts should be made by teachers to demystify the teaching and learning of the course in order to enable student teachers acquire the necessary knowledge and competencies for conducting valid and reliable classroom assessments in schools in Nigeria.

The findings of the study also show that there are no significant mean differences in the responses of the teachers in what they know and what they do not know about aspects of validity. These findings imply that all the teachers need capacity building to update their knowledge and competencies in development and validation of assessment instruments. It is also necessary in the light of the observation (Dosumu, 2002; Adeola & Fajonyimi, 1999; Imo, 2012; Agu, Onyekuba & Anyichie, 2013) that teachers lack test construction skills. This is not surprising because most teachers receive little or no training or support after certification. Teachers therefore need some basic knowledge on how to develop and validate their classroom tests to enable them use results of their classroom assessment for taking relevant decisions about the students. The situation may be worse in other faculties in the university where most lecturers do not have any formal training in educational assessment. Efforts should therefore be directed to assessment of teachers' competence in test construction and building necessary competence for effective teaching and learning in the system.

V. Conclusion

The focus of this study was to provide evidence about what teachers know and do not know about validity of classroom tests. The purpose was to provide baseline for capacity building in ensuring validity of classroom tests in order to take appropriate decisions about students' learning and development. The findings of the study have shown that the teachers have some knowledge about aspects of content-related evidence of validity, procedure for ensuring coverage and adequate sampling of content and objectives, as well as correlating students' scores in two measures for predictive validation. It was also found that teachers lack adequate knowledge about aspects of validity that deal with criterion-related evidence, concept of face validity and sources of invalidity of test scores. Their high ratings on the need for capacity building in the areas of evaluating validity of assessment, predictive and content validity assessment require that the teachers be trained and retrained through regular in-house seminars and workshops to update their knowledge of validity of students' scores and how to interpret them for improving students' learning. One limitation of this study was the requirement of opinions of teachers in only one faculty in a university in Nigeria without triangulation of data sources. Nonetheless, the findings have implications for building capacity of teachers in the most important component of testing, which is validity. This study has therefore made a useful contribution in providing empirical evidence that can be used as baseline for addressing the problems associated with validity of classroom tests at the university.

References

- [1]. Adeola, A. O., & Fajonyomi, A. A. (1999). Adult learners' performance on standardized and nonstandardized tests as a function of sex and location. *Nigeria Journal of Counselling Development*, 4(1), 32-36.
- [2]. Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). *Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory*. Retrieved from http://www.academicjournals.org/article/article1379700583_Agu%20et%20al.pdf
- [3]. Dosumu, C. T. (2002). *Issues in teacher-made tests*. Ibadan: Olatunji and Sons Publishers.
- [4]. Imo, G. C. (2012). *Effects of training in test construction on quality of teacher-made tests and students' physics achievement in secondary schools in Plateau State, Nigeria*. An unpublished Ph. D Thesis, University of Jos, Jos, Nigeria.
- [5]. Killen, R. (2003). Validity in outcomes-based assessment. *Perspectives in Education*, 21 (1). Retrieved from [http://repository.up.ac.za/bitstream/handle/2263/4913/killenvalidity.\(2003\).pdf?sequence=1](http://repository.up.ac.za/bitstream/handle/2263/4913/killenvalidity.(2003).pdf?sequence=1)
- [6]. Messick, S. (1996). *Validity and Washback in language testing*. Prentice, NJ: Educational Testing Service.
- [7]. National Universities Commission (2002). *Quality assurance in Nigerian Universities vol.1. Ranking academic programmes in 1999 and 2000 accreditation exercise*. Abuja: NUC.
- [8]. Popham, W. (1990). *Modern educational measurement: A practical perspective* (2nd ed.). Englewood Cliffs: Prentice-Hall.
- [9]. Professional Testing Inc. (2006). Test validity. Retrieved from http://www.proftesting.com/test_topics/pdfs/test_quality_validity.pdf
- [10]. Rudner, L. M., & Schafer, W. D. (2002). *What teachers need to know about assessment*. Washington, DC: National Education Association.
- [11]. Author (2008). *Fundamentals of educational measurement and evaluation*. Jos: Fab Anieh Ltd.
- [12]. Author (2005). Strategies for quality assurance in educational assessment at the university level. *Nigerian Journal of Curriculum Studies*, 12 (3), 10-17.