

Nigerian teachers' utilization of test construction procedures for validity improvement of achievement tests

Omaze Anthony Afemikhe & Sylvanus Yakube Imobekhai

Institute of Education, University of Benin, Benin City, Nigeria

Teachers are very important stakeholders in the education system as they drive whatever takes place within the classrooms of the school system. Apart from teaching, they generate assessments which are used to evaluate teaching efficacy and learning achievement within their classrooms. The validity of these assessments is of essence if we are to have confidence in the interpretations and uses to which the assessments are put. To ensure the validity of the tests used by teachers, appropriate test construction procedures should be utilized by them. The main question addressed in this study is whether teachers use appropriate construction procedures which confer on tests the requisite validity. In executing this study a survey approach was applied and the population of the study was composed of teachers in primary and secondary schools in Benin metropolis in Nigeria. From the population a sample of five hundred teachers made of two hundred and fifty each from primary and secondary school levels were selected. A questionnaire focusing on steps in constructing a valid achievement test was designed. The response option was a three-point scale of 'all the time', 'sometimes' and 'not at all'. The validity-evidence of the questionnaire was established using a juror of experts in measurement and evaluation and they were to determine the adequacy, comprehensiveness and suitability of the items. The reliability of the scores from the instrument was determined using Cronbach alpha and it yielded a value of .751. The data collected would be analyzed using means and standard deviation, an interpretative norm, t-test and ANOVA. The results indicated that teachers utilise approved procedures in constructing achievement test. They however did not use procedures to enhance content-validity evidence, and qualitative analysis. Results were also affected by experience. It was recommended that teacher development programmes be mounted to fill the gaps noticed in this study.

Introduction

Assessment is a pervasive term in education as it is at the core of what happens and in particular drives teaching and learning within the classroom. According to Smith (2001), it is a set of processes through which inferences are made about learners' learning process, skills, knowledge and achievement. It is however more encompassing as it is a process of collecting information for making decisions within the educational system. These decisions could be about the students, school, curriculum and even educational policies. Within the school the functions of assessment can be looked at from two perspectives; that of the learners and the authority.

From the learners' angle three issues are germane; these are Choose, Learn and Qualify. Assessment information can be used to choose students into programmes; while assessments can also highlight the strengths and weaknesses of students in the learning process; and yet assessments could also be used to determine those who qualify and therefore worthy of certificates. These assessments include those for improvement of learning and those for certification which have been technically referred to as formative and summative assessments respectively.

From the perspective of the authorities, assessment serves the functions which have been referred to as Select, Monitor and Hold Accountable. An important aspect of education is who gets admitted or selected into an educational programme. Schools use assessment information composed of test scores and other affective measures for selecting potential students. Assessment information can equally be used to track the functioning of the components of the educational system which is a monitoring function. In this case the questions asked include how are the students performing, what are the problems hindering optimum performance of the students and the school? The schools belong to the community and it invests in them. As a result the community could be interested in how well the schools are doing the job assigned to them. Thus assessment information is useful in holding accountable those responsible for the different components of the school system if it is to achieve its goals. The managers must of necessity show that they have used judiciously money budgeted and allocated to the system.

The use of assessment as it relates to the students is very much emphasized because it is used in teaching and learning; some have even seen assessment and teaching and learning as two sides of the same coin. Assessment has been described as an instructional tool (Bailey, 2004) and important in teaching-learning process (Al-Shara'h, 2011). This use however is greatly dependent on teachers' assessment practices (Cumming, 2001; Mertle, 2005). They can use it to direct the learning process. Assessment practice is the totality of the steps and procedures taken by teachers throughout an assessment; it includes preparation, administering, grading, recording and reporting of the assessment information. The traditional testing procedures used in assessment include multiple choice, matching, true/false, short answer and essay questions. Alternative assessments such as observation, conferences, portfolios, peer and group assessments techniques have equally been used. The preparation of each of these assessments is important in ensuring the validity of the information generated.

The traditional assessment processes, particularly achievement tests, still dominate the applications within most Nigerian schools. In particular, achievement tests are tools which dominate the assessment terrain in most Nigerian schools. The procedures for its construction are usually important if the validity of the scores obtained from it is to be relied upon. Thus certain practices are important in ensuring the acceptability of results from assessments used in schools. For example the test blue print is important to improve the validity of teachers' evaluations based on tests constructed for classroom use (Fives & DiDonato-Barnes, 2013). This is because it is used to align objectives, instruction and assessment (Notar, Zuelke, Wilson & Yunker, 2004). It also

ensures there is no lopsidedness in the topics or subject matter coverage, thus providing high content-validity evidence. The cornerstone of classroom assessment practices is validity evidence of the judgments about students' learning and knowledge (Wolming & Wilkshom, 2010). Accurate and valid information about student achievement is essential for effective instruction through feedback and its use in adapting instruction to students' needs and abilities (Martinez, Stecher & Borko, 2009).

Item analysis of test responses is equally important in ensuring the validity of tests. The results make teachers confident of their abilities to construct test items and they benefit from such items in the future. This is through the improvement of the items and consequently making them valid, practical and reliable (Al-Younes, 2006). The validity of the information is also important as the information can be used to improve performance (Stiggins & Chappuis, 2005). Teachers, because of their long interaction with students throughout the school year, gain a good understanding far richer and multidimensional than what is obtainable from standardized tests (Meisels, Bickel, Nicholson, Xue & Atkins-Burnett, 2001). It is in this regard that construction of a valid achievement test is relevant. The ratings of teachers may however reflect some biases. The quality of the assessment by teachers is related to their teaching experience and pedagogical development (Rodriquez, 2004) and features of the classroom (Llosa, 2004).

The foregoing shows the importance of good quality assessments in schools. It is no wonder that assessment competencies have been outlined which teachers should imbibe to be able to execute their assessment functions. These competencies include:

- choosing assessment methods appropriate for instructional decisions;
- developing assessment methods appropriate for instructional decisions;
- administering, scoring, and interpreting the results of both externally- produced and teacher-produced assessment methods;
- using assessment results when making decisions about individual students, planning teaching, developing curriculum and making recommendations for school improvement;
- developing valid grading procedures which use pupils assessment;
- communicating assessment results to students, parents, other lay audiences and other educators; and
- recognizing unethical, illegal and otherwise inappropriate assessment methods and uses of assessment information (The American Federation of Teachers, National Council on Measurement in Education and the National Education Association, 1990).

According to Omo-Egbekuse, Afemikhe and Imobekhai (2012) teachers need to be skilled and competent on all these standards to be able to assess their students efficiently and effectively; the results of their study showed teachers' indicating high level of competence in all areas with the exception of recognizing unethical, illegal and otherwise inappropriate assessment methods and uses of assessment information. The actual elements of the evaluation schemas that teachers institute have received less attention in Nigeria. Agu, Onyekuba & Anyichie (2013) attempted a construction of a test construction inventory which they found to be valid and reliable. As teachers use achievement tests very often in their classroom, it is necessary to examine what precautionary steps they take to ensure that the tests are valid. This is the focus of this study. This is with the anticipation that if the procedures adopted fall short of expectations, teacher development programmes can be organized to enhance the practices exhibited by them.

Consequently, this study attempted to answer the following questions:

1. Do teachers use appropriate construction procedures which confer on achievement tests the requisite validity?
2. Is there variability in test constructions procedures used by primary and secondary school teachers?
3. Are there variability in test construction procedures along the lines of teaching experience?

Methodology

In executing this study a cross-sectional survey approach was applied. The population of the study was composed of teachers in primary and secondary schools in Benin metropolis in Nigeria. From the population, a sample of five hundred teachers made of two hundred and fifty each from primary and secondary school levels were selected. The selection involved sampling of schools and all teachers in the sampled school were eligible to take part in the study. The teachers were not compelled but were approached to complete the questionnaire after the details of the study had been explained.

A questionnaire focusing on steps in constructing a valid achievement test was designed by the researchers. The questionnaire titled 'Achievement Test Construction Procedures Questionnaire' was composed of two sections. Section A asked respondents to supply some demographic information such as sex, school ownership, type of school (primary or secondary), highest educational qualification, experience and training background in education. Section B itemized some activities used in constructing a valid achievement test. The respondents were to indicate on a three-point scale of 'all the time', 'sometimes' and 'not at all' how often each of the activities was carried out. The validity-evidence of the questionnaire was established using 3 jurors of experts in measurement and evaluation and they were to determine the adequacy, comprehensiveness and suitability of the items. Based on the comments and observations of the experts the questionnaire was corrected and copies produced for the determination of the reliability of the scores. The responses provided were scored as 'all the time' = 3, 'sometimes' = 2 and 'not at all' = 1. The reliability of the scores from the instrument was determined using Cronbach alpha and it yielded a value of 0.751.

The emerging data from the full study were analysed using means and standard deviation. An interpretative norm for the items was set as 2 the mean of the response categories. A mean of ≥ 2.50 was taken to mean that the activity was carried out.

Results and Discussion

Out of the 500 questionnaires administered four hundred and seventeen usable ones were retrieved giving a return rate of 83.4%. Among the teachers there were 189 (45.3%) primary and 211 (50.6%) secondary school teachers. The distribution of the teachers show that 120 (28.8%), 111(26.6%), 76 (18.2%), 25 (6.0%) and 61 (16.5%) were 0-4 years, 5-9 years, 10-14 years, 15-19 years and 20 years and above experience respectively.

From Table 1, using the interpretative norm setup, it is found among test construction practices listed, that the teachers generally do not 'generate a table of specification or test blue print', 'use only objective test items', 'find out how difficult test items are for the examinees',

‘seek opinion of other teachers of the subject on quality of test items’, ‘use only essay questions’, and ‘make sure that information in one question does not provide a clue to another question’.

Table 1: Mean and standard deviation of procedures

Procedure	N	Mean	Std. Deviation
Decide on purpose of test.	411	<u>2.54</u>	0.54
Outline the content test would cover.	402	<u>2.61</u>	0.55
Specify objectives to be tested.	405	<u>2.61</u>	0.55
Generate a table of specification or test blue print.	395	2.17	0.70
Write out items well ahead of date for test.	407	<u>2.52</u>	0.61
Use only objective test items.	407	2.25	0.65
Find out how difficult test items are for the examinees.	403	2.41	0.63
Find out if the constructed test items are of good quality.	410	<u>2.67</u>	0.57
Seek opinion of other teachers of the subject on quality of test items.	409	2.28	0.63
Edit items for grammatical accuracy.	409	<u>2.63</u>	0.58
Use both essay and objective test items.	403	<u>2.64</u>	0.54
Decide on whether item can discriminate between high and low scoring candidates.	408	2.27	0.71
Ensure constructed items match the table of specification.	399	<u>2.52</u>	0.67
Write items a day or few days before it is to be used.	404	<u>2.57</u>	0.57
Take a decision on how test would be scored.	402	<u>2.76</u>	0.48
Consider the ability level of the students.	400	<u>2.74</u>	0.47
Make the instructions very clear to the students.	400	<u>2.86</u>	0.37
Make sure that visual or graphical materials are clear.	395	<u>2.76</u>	0.48

Table 1 Cont'd

Procedure	N	Mean	Std. Deviation
------------------	----------	-------------	-----------------------

Ensure language of test items is easily understood by all examinees.	406	<u>2.77</u>	0.49
Ensure that some students are not advantaged or disadvantaged because of the cultural, religious or social group they belong.	397	<u>2.52</u>	0.65
Use only essay questions.	398	2.02	0.59
Construct test items that are really important in learning the subject.	401	<u>2.75</u>	0.47
Ensure the questions set are within the curriculum taught?	403	<u>2.85</u>	0.40
Make sure the test items are measuring what you (the teacher) intends they should?	402	<u>2.84</u>	0.43
Make sure that information in one question does not provide a clue to another question?	397	2.38	0.65

In Table 2, significant differences were noticed between primary and secondary school teachers in practices such as ‘seek opinion of other teachers of the subject on quality of test items’ (primary = 2.87, secondary = 2.20, $t = 2.713$, $df = 393$, $p = 0.007$), ‘take a decision on how test would be scored’ (primary = 2.71, secondary = 2.82, $t = -2.235$, $df = 389$, $p = 0.26$), ‘ensure language of test items is easily understood by all examinees (primary = 2.84, secondary = 2.71, $t = 2.526$, $df = 393$, $p = 0.012$).

Table 2: t-test of difference between means of procedures used between primary and secondary school teachers

Procedure	Type of school	N	Mean	Std. dev.	t	df	Sig.(2-tailed)
Decide on purpose of test.	Primary	189	2.53	0.56	-.740	396	.460
	Secondary	209	2.57	0.52			
Outline the content test would cover.	Primary	184	2.59	0.55	-.970	388	.333
	Secondary	206	2.64	0.55			
Specify objectives to be tested.	Primary	185	2.61	0.50	-.187	389	.852
	Secondary	206	2.62	0.60			
Generate a table of specification or test blue print.	Primary	176	2.13	0.72	-1.524	379	.128
	Secondary	205	2.23	0.67			

Table 2 Cont'd

Procedure	Type of school	N	Mean	Std. dev.	t	df	Sig.(2-tailed)
Write out items well ahead of date for test.	Primary	186	2.53	0.63	.635	391	.526
	Secondary	207	2.49	0.61			
Use only objective test items.	Primary	185	2.30	0.65	1.439	391	.151
	Secondary	208	2.20	0.66			
Find out how difficult test items are for the examinees.	Primary	183	2.44	0.64	.461	387	.645
	Secondary	206	2.41	0.62			
Find out if the constructed test items are of good quality.	Primary	188	2.62	0.61	-1.814	394	.070
	Secondary	208	2.72	0.53			
Seek opinion of other teachers of the subject on quality of test items.	Primary	185	2.37	0.61	2.713	393	.007
	Secondary	210	2.20	0.65			
Edit items for grammatical accuracy.	Primary	186	2.66	0.54	.740	393	.460
	Secondary	209	2.61	0.62			
Use both essay and objective test items.	Primary	183	2.69	0.54	1.429	389	.154
	Secondary	208	2.61	0.54			
Decide on whether item can discriminate between high and low scoring candidates.	Primary	184	2.30	0.68	.581	392	.562
	Secondary	210	2.26	0.74			
Ensure constructed items match the table of specification.	Primary	179	2.55	0.65	1.020	386	.308
	Secondary	209	2.48	0.69			
Write items a day or few days before it is to be used.	Primary	186	2.59	0.56	.457	391	.648
	Secondary	207	2.57	0.57			
Take a decision on how test would be scored.	Primary	183	2.71	0.49	-2.235	389	.026
	Secondary	208	2.82	0.46			

Table 2 Cont'd

Procedure	Type of school	N	Mean	Std. dev.	t	df	Sig.(2-tailed)
Consider the ability level of the students.	Primary	184	2.72	0.48	-.601	387	.546
	Secondary	205	2.75	0.47			
Make the instructions very clear to the students.	Primary	181	2.85	0.41	-.929	388	.354
	Secondary	209	2.88	0.34			
Make sure that visual or graphical materials are clear.	Primary	183	2.81	0.42	1.598	382	.111
	Secondary	201	2.73	0.52			
Ensure language of test items is easily understood by all examinees.	Primary	186	2.84	0.40	2.526	393	.012
	Secondary	209	2.71	0.57			
Ensure that some students are not advantaged or disadvantaged because of the cultural, religious or social group they belong.	Primary	184	2.57	0.62	1.353	384	.177
	Secondary	202	2.48	0.69			
Use only essay questions.	Primary	184	2.03	0.56	.127	386	.899
	Secondary	204	2.02	0.61			
Construct test items that are really important in learning the subject.	Primary	184	2.76	0.44	.357	389	.721
	Secondary	207	2.74	0.49			
Ensure the questions set are within the curriculum taught.	Primary	185	2.85	0.37	-.058	391	.953
	Secondary	208	2.85	0.41			
Make sure the test items are measuring what you (the teacher) intends they should.	Primary	185	2.84	0.42	-.064	390	.949
	Secondary	207	2.84	0.43			
Make sure that information in one question does not provide a clue to another question.	Primary	184	2.33	0.65	-1.646	385	.101
	Secondary	203	2.43	0.63			

The ANOVA summary in Table 3 shows significant differences in ‘decide on purpose of test’ ($F=3.374$, $df=4$, 394 , $p=0.10$), ‘Specify objectives to be tested’ ($F=2.473$, $df=4$, 387 , $p=0.044$).

The differences were between teachers with 0-4 years and those with 10-14 years teaching experience.

Table 3: ANOVA summary table for differences among means of teaching experience

Procedure applied		Sum of Squares	df	Mean Square	F	Sig.
Decide on purpose of test.	Between Groups	3.877	4	.969	3.374	.010
	Within Groups	113.191	394	.287		
	Total	117.068	398			
Outline the content test would cover.	Between Groups	1.081	4	.270	.898	.465
	Within Groups	115.525	384	.301		
	Total	116.607	388			
Specify objectives to be tested.	Between Groups	3.011	4	.753	2.473	.044
	Within Groups	117.823	387	.304		
	Total	120.834	391			
Generate a table of specification or test blue print?	Between Groups	2.006	4	.501	1.017	.398
	Within Groups	185.934	377	.493		
	Total	187.940	381			
Write out items well ahead of date for test?	Between Groups	.632	4	.158	.427	.789
	Within Groups	143.744	389	.370		
	Total	144.376	393			
Use only objective test items?	Between Groups	1.484	4	.371	.879	.476
	Within Groups	164.140	389	.422		
	Total	165.624	393			

Table 3 Cont'd

Procedure applied		Sum of Squares	df	Mean Square	F	Sig.
Find out how difficult test items are for the examinees?	Between Groups	1.663	4	.416	1.044	.384
	Within Groups	153.212	385	.398		
	Total	154.874	389			
Find out if the constructed test items are of good quality?	Between Groups	1.189	4	.297	.906	.460
	Within Groups	128.584	392	.328		
	Total	129.773	396			
Seek opinion of other teachers of the subject on quality of test items?	Between Groups	1.921	4	.480	1.191	.314
	Within Groups	157.706	391	.403		
	Total	159.626	395			
Edit items for grammatical accuracy?	Between Groups	3.056	4	.764	2.273	.061
	Within Groups	131.376	391	.336		
	Total	134.432	395			
Use both essay and objective test items?	Between Groups	.898	4	.224	.749	.559
	Within Groups	115.936	387	.300		
	Total	116.834	391			
Decide on whether item can discriminate between high and low scoring candidates?	Between Groups	4.133	4	1.033	2.030	.090
	Within Groups	198.485	390	.509		
	Total	202.618	394			
Ensure constructed items match the table of specification?	Between Groups	.240	4	.060	.130	.971
	Within Groups	176.737	383	.461		
	Total	176.977	387			

Table 3 Cont'd

Procedure applied		Sum of Squares	df	Mean Square	F	Sig.
Write items a day or few days before it is to be used?	Between Groups	1.303	4	.326	.978	.419
	Within Groups	129.292	388	.333		
	Total	130.595	392			
Take a decision on how test would be scored?	Between Groups	.245	4	.061	.257	.905
	Within Groups	92.244	387	.238		
	Total	92.490	391			
Consider the ability level of the students?	Between Groups	.900	4	.225	1.012	.401
	Within Groups	85.296	384	.222		
	Total	86.195	388			
Make the instructions very clear to the students?	Between Groups	.706	4	.176	1.260	.285
	Within Groups	53.798	384	.140		
	Total	54.504	388			
Make sure that visual or graphical materials are clear?	Between Groups	1.064	4	.266	1.125	.344
	Within Groups	90.045	381	.236		
	Total	91.109	385			
Ensure language of test items is easily understood by all examinees?	Between Groups	.304	4	.076	.304	.875
	Within Groups	97.731	390	.251		
	Total	98.035	394			
Ensure that some students are not advantaged or disadvantaged because of the cultural, religious or social group they belong?	Between Groups	1.128	4	.282	.650	.627
	Within Groups	165.278	381	.434		
	Total	166.407	385			

Table 3 Cont'd

Use only essay questions?	Between Groups	1.516	4	.379	1.113	.350
	Within Groups	130.391	383	.340		
	Total	131.907	387			
Construct test items that are really important in learning the subject?	Between Groups	.776	4	.194	.890	.470
	Within Groups	84.160	386	.218		
	Total	84.936	390			
Ensure the questions set are within the curriculum taught?	Between Groups	.377	4	.094	.583	.675
	Within Groups	62.440	387	.161		
	Total	62.816	391			
Make sure the test items are measuring what you (the teacher) intends they should?	Between Groups	.239	4	.060	.315	.868
	Within Groups	73.285	386	.190		
	Total	73.524	390			
Make sure that information in one question does not provide a clue to another question?	Between Groups	1.340	4	.335	.806	.522
	Within Groups	158.370	381	.416		
	Total	159.710	385			

The fact that teachers do not generate a table of specification or test blue print is worrisome as one cannot then guarantee the content-validity evidence of these tests constructed by the teachers. Thus the teachers may not have aligned objectives, instruction and assessment, which is a benefit as indicated by Notar, Zuelke, Wilson & Yunker (2004), The teachers not using 'only objective test items' or 'use only essay questions' is a good practice as through them the teacher can assess all levels of cognitive functioning. The difficulty of test items is important as items have to be tailored to the ability of the examinees. If too difficult or too easy, validity can be jeopardized. It is therefore a concern that teachers do not 'find out how difficult test items are for the examinees'. The consequence would be that teachers cannot improve the items and consequently their validity, practicality and reliability would be questionable (Al-Younes, 2006) Seek opinion of other teachers of the subject on quality of test items is a qualitative analysis issue that can enhance validity evidence. When questions serve as cue to other items, teachers would not be testing what

they intend. These teachers not ensuring that information in one question does not provide a clue to another question is equally of concern.

The significant differences noticed between primary and secondary school teachers in practices such as 'seek opinion of other teachers of the subject on quality of test items' (primary = 2.87, secondary = 2.20, $t = 2.713$, $df = 393$, $p = 0.007$) is not unexpected. It is in favour of primary school teachers. As noted by Omo-Egbekuse, Afemikhe and Imobekhai (2012), many primary school teachers were initially employed with the teachers' grade II certificate and this programme used a lot of clinical approaches in training would-be teachers. A majority of these teachers have now been upgraded to NCE holders and even other higher degrees and the initial training still has effect on them. As indicated by (Rodriquez, 2004) teaching experience influences assessment implementation and differences between experienced teachers and beginners on 'decide on purpose of test' in favour of experienced teachers are therefore not surprising.

Conclusion and recommendations

This study showed that teachers ignore some aspects of achievement construction procedures germane to ensuring validity of the tests. They however apply some procedures recommended for test construction. Variations on some few aspects of the procedures were however noticed between primary and secondary school teachers and between beginning and experienced teachers. It is therefore recommended that teacher development programmes be put in place based on the gaps noticed in this study. It is also advisable that teacher retention be encouraged in schools as experience is important in implementing procedures that can enhance validity-evidence of achievement tests.

References

- Al-Younes, Y.A. (2006). The extent of the University of Jordan faculty members' practices of test preparation, correction, analysis, interpretation and discussion of results with students. *Assyout University Journal of the Faculty of Education*, 22(1), 153-193.
- Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). Measuring teachers' competence in construction classroom-based tests in Nigerian secondary schools: Need for a test construction inventory. *Educational Research and Reviews*, 8(8), 431-439.
- Bailey, K.M. (2004). *Learning about language assessment: Dilemmas, decisions and directions*. Boston, MA: Heinle & Heinle Publishers.
- Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: Specific purposes or general purposes? *Language Testing*, 18, 207-224
- Fives, H. & Didonato-Barnes, N. (2013). Classroom test construction: The power of a table of specification. *Practical Assessment, Research & Evaluation*, 18(3) Available online: <http://pareonline.net/getvn.asp?v=18&n=3>. Accessed July 18, 2015.
- Leung, C. Y. & Andrews, S. (2012). The mediating role of textbooks in high-stakes assessment

- reform. *ELT Journal*, 66(3), 356-365.
- Llosa, L. (2008). Building and supporting a validity argument for a students'-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issue and Practice*, 27(3), 32-42.
- Martinez, J.F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments and student achievement in Mathematics; Evidence from ECCS. *Educational Assessment*, 14, 78-102.
- Malone, M.E. (2013). The essentials of assessment literacy: contrast between testers and users. *Language Testing*, 30(3), 329-344.
- Meisels, S.J., Bickel, D.D., Nicholson, J., Xue, Y., & Atkins-Burnett (2001). Trusting teachers' judgments: a validity study of a curriculum-embedded performance assessment in Kindergarten-Grade 3. *American Educational Research Journal*, 38(1), 73-95.
- Notar, C.E., Zuelke, D.C., Wilson, J.D. & Yunker, B.D. (2004). The table of specification: Insuring accountability in teacher made tests. *Journal of Instructional Psychology*, 31, 115-129.
- Omo-Egbekuse, J., Afemikhe, O.A., & Imobekhai, S.Y. (2012). Nigerian teachers' ratings of their competence in standards for educational assessment of students. *Benin Journal of Educational Studies*, Volume 21, Nos 1 & 2, 2012, 1- 11
- Rodriguez, M.C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, 17(1), 1-24.
- Stiggins, R.J., & Chappius, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, 44(1), 11-18.