

Appropriate Steps and Ethical Considerations of Assessment Practices in the University of Bamenda: A Survey of Self-Assessment and Students' Perception Inventories

Valentine Banfegha Ngalim & Yoh Edelquin Nabi

DOI: 10.31364/SCIRJ/v6.i12.2018.P1218590

<http://dx.doi.org/10.31364/SCIRJ/v6.i12.2018.P1218590>

Abstract: Teaching and assessment are two sides of the same coin. Assessment is an integral part of the teaching profession and the results of any assessment are a measuring rod to evaluate the teacher's output. To ensure the validity of the tests used by teachers, appropriate test construction, administration and scoring procedures should be used. The main question addressed in this study is whether teachers use appropriate construction procedures to enhance the validity of the results obtained from testing. In executing this study, a survey approach was used and the population of the study comprised lecturers and students in some selected faculties of the University of Bamenda. From the population, a sample of 53 teachers and 238 students were selected. Two questionnaires focusing on steps in constructing a valid achievement test, administering it with the use of ethical issues were designed. The reliability of the scores from the instrument was determined using Cronbach alpha and it yielded a value of .702 for students and 0.823 for lecturers. Samples of 30 tests papers were examined. The data collected from questionnaires were analyzed using means and percentages, t-test and factor analysis. The results indicated that teachers exhibited good test developments. They however did not use test blue print, no item analysis and no peer review of items. It was recommended that teacher development seminars be organized regularly to fill the gaps noticed in this study.

Keywords: Assessment practices; University of Bamenda; Student's perception inventories.

INTRODUCTION

University teaching and evaluation practices may not be as organized as in the primary or secondary schools where the pupils or students write final examinations for certification such as the First School Leaving Certificate and the General Certificate of Education. These are standardized summative examinations whose test items have been validated and their reliability established by experts. At the university level, the students are certified after three years of studies based on the cumulative scores obtained from the teacher made test and examinations for these years. This has been referred to as "consolidated scores" (Farrant 1980, p.154). The poor performance of University students in these tests may be attributed to their attitude towards studies or the teaching method employed by the teacher. Most often than not reference is not made to the appropriateness of the teachers' test items.

Many studies have been done on the question of developing instruments for evaluating teacher competencies in test construction or assessment practices (Afemikhe&Yakube, 2015; Agu Onyekuba&Anyichie, 2013; Zhang & Burry-Stock 2003). Most of these studies have been carried out elsewhere, and so, there is need for a similar study in Cameroon. This study is also appalling because lecturers in the University are recruited based upon their acquisition of a Master's Degree and proof of being admitted to a Doctorate program, with little or no professional training as a teacher or in test development. At the same time, lecturers' design of tests and certification of students graduating are dependent on the results obtained. In most cases, there are no external examinations, which are written for the award of a Bachelor's degree. Most assessment practices in the University are written works, assignments, projects, internships and examinations. This study places particular emphasis on written exercises for continuous assessment. The teachers set objective tests as well as essay questions. No matter the format, the procedures for test construction are usually

important for the scores obtained to be valid and relied upon. Therefore, certain practices are important to ensure the acceptability of results from assessments used in schools.

Lecturers in the University are expected to meet the accountability standards, share professional knowledge with their colleagues, care deeply about students and their success, and hold distinctive qualities that characterize their effectiveness. For assessment, the following competencies are needed, as laid out by The American Federation of Teachers, National Council on Measurement in Education and the National Education Association, 1990, (Afemikhe & Yakube, 2015). According to these standards, teachers should be skilled in choosing and developing assessment methods, administering and scoring tests, interpreting and communicating assessment results, grading and meeting ethical standards in assessment. The assessment literature and some of the seven standards form the theoretical framework for investigating teachers' competencies in improving test validity in this study.

Assessment of students is an integral part of the teaching-learning process. At the university level, tests are designed by the lecturers and are used to certify the students at the end of the study period. As such, the tests have to be constructed, administered and graded taking into consideration the psychometric properties of validity, reliability and some ethical issues for the results to be relied upon. They ought to be fair and free from bias, so that the marks students get be a true reflection of their work (Ethical Standards in Testing, 2001; The University of Texas at Austin Faculty Innovation center, 2017). This is to ensure that Universities do not produce students whose knowledge and skills do not reflect the certificates they possess.

Statement of the Problem

The University of Bamenda created in December 2010 hosts some lecturers/ professors who were recruited with academic degrees irrespective of their professional qualities. Some of these teachers have never had any training in the appropriateness of test construction and the ethical considerations associated with test development. In most cases, the recruitment of teachers in Higher Education is principally based upon the acquisition of a Doctorate Degree with little or no teaching experience nor professional training as a test developer. Most teachers believe that they have adequate knowledge of testing and attributed that knowledge to experience and University coursework. They evaluate students based on the past experiences learned from their own teachers in the University. Findings of some researchers reported by Anja (2009) revealed that classroom tests, especially in higher education in Cameroon, are sometimes biased as some teachers give easy test to procure good performance, while others think a difficult test is a reflection of their own intelligence. Some give tests just to obtain marks for the results. Since students' results are a reflection of the teachers' work, some teachers choose a particular test format to suit them and ease their work while others moderate their scoring to minimize failure. This refers to the fact that at times the marks students score, are not a valid representation of their knowledge in the course concerned. Can the problem be attributed to the inappropriateness of test construction? Is it in the unethical considerations of assessment or both?

From one's experiences as a student in some higher education institutions in Cameroon, the quality of some tests given to students is questionable. This is because one does not seem to establish the criteria or standards followed for classroom test development. Some teachers set tests out of the content taught, some ask students to read volumes of pages but they sample just a single domain to be tested. Others give almost the same mark to students no matter how they write. Also, some teachers set questions that do not reflect objectives provided at the initial part of the course. At times, this may lead to arbitrary allocation of marks to meet deadlines.

Objectives of the study

The main purpose of this study is to investigate whether University teachers follow the appropriate steps and ethical considerations in test construction to improve on the validity of the test. Specifically, this study sets;

- To inquire if University teachers use appropriate steps of test construction that ensure the validity of their test.
- To find out the ethical considerations of test construction used by University teachers to improve the validity of their test.

Research questions

The main research question is: which steps and ethical considerations do teachers in the University of Bamenda follow in test construction to improve on the validity of the test?

Specifically, the study is set to answer these questions:

- Which appropriate steps of assessment do University teachers use to ensure the validity of their test?
- What are the ethical considerations of test construction used by University teachers to improve the validity of their test?

Significance of the Study

This study intends to enlighten Heads of Departments, Deans, Directors and the University central administration on useful information and reasons to always plan and organize training sessions on test construction skills for teachers in the University. Heads of Departments can also exploit the findings of this study to diagnose and design professional development courses for teachers in test development. Finally, this study hopes to provide school administrators with data that can be used to evaluate school climate, track progress, and contribute to summative evaluations of a selected group of teachers.

Definition of Key Terms

Assessment practices refer to all the steps and procedures taken by teachers through-out the testing exercise. It includes preparation, administering, grading, recording and reporting of the assessment information. In this case, assessment is the process of collecting data to measure the performance or capabilities of a student or group. This can be paper-and-pencil tests, homework completion, projects and observation.

Teacher-made-test refers to the questions prepared and administered by the classroom teacher for the purpose of measuring the achievement of his/her own pupils. These are tests prepared by the classroom teacher, also known as classroom assessment (Nworgu, 2015). Some of these in the University are written works, assignments, projects, internship, and examinations. The tests are given within the semester period and also include the semester examination set by the teachers.

Competence is “*the ability of a person or organisation to achieve particular levels of performance*” (Mulder 2001, in Roelofs & Sandars, 2007). This refers to teachers’ knowledge, experience and abilities or skills required to perform a particular task effectively. Competency in testing refers to the ability of a teacher to exhibit on the job skills and knowledge gained as a result of training in designing, administering, grading and interpreting students’ results.

An Inventory is an instrument or questionnaire used to collect valuable information through a survey showing the necessary skills or competences needed.

Continuous assessments (C.A) indicate formative tests given during the course of study to monitor the students' progress. They are given in form of assignments, tests, projects group works and the marks are summed up to give one mark known as the C.A mark. In the University of Bamenda, class attendance is considered in the C.A. mark to make up 30% of the exams.

Validity refers to the level of confidence with which an examinee's test score could be used to infer the ability under measurement possessed by the examinee (Shimbery 1990, in Anyanwu, & Onwuakpa, 2015). Validity here, simply means how accurate the scores represent what it is supposed to measure.

A student survey or student rating is a process of using students to evaluate the teacher by getting their impressions of the teacher's practices, with the use of a questionnaire or checklist.

Review of Related Literature

In this study, we are going to exploit the theoretical and the empirical frameworks in order to justify our findings. In the first place, we consider the two theories relevant to our two variables. These include; classical test theory and the Item response theory. The Classical Test Theory (CTT) is defined as a body of related psychometric theory that predicts outcomes of psychological testing such as the difficulty of items or the ability of test takers. The idea was born in 1904 by Charles Spearman and has been modified by George Udny Yule, Truman Lee Kelley, Louis Guttman, and, most recently, Melvin Novick (Demars in Irwing et al., 2018; Wu et al., 2017).

This theory set out to understand and improve the reliability of psychological tests. Classical test theory assumes that each person has a true score that would be obtained if there were no errors in measurement. Because instruments used for measurement are most often imperfect, the score that is observed for each person is often different from the person's true ability or characteristics. The theory concludes that the difference between the true score and the observed score is the result of measurement error. This theory is summarized in an equation as follows:

$$X=T+E$$

Where X is the observed score that is seen on the student's test paper graded by the teacher.

T is the score that the student should have had out of true ability.

E is error arising from the test conditions or that of the student.

The error in measurement may come from the test content, nature of test items, topics sampled, how the test is administered and how it is scored, together with the examinee factors such as ill health, and ethical issues (Demars in Irwing et al., 2018; Wu et al., 2017).

In the second place, the Item Response Theory is a general statistical theory about examinee item and test performance. Also, it deals with how performance relates to the abilities that are measured by the items in the test. This theory was propounded by Ferguson (1942), which states that during the process of testing, there is an encounter between the testee and the test item. The result of such an encounter, if analysed, reveals the ability level of the testee with regard to the item in question. This ability level is an innate unobservable characteristic of the testee, which causes consistent performance on the test and influences the testee's responses to test items. Just as everybody has his/her own ability level, a test item also has inherent characteristics that affect the exhibition of the testee's ability. According to Warm (1978), when a testee enters a testing room, he comes along with a certain ability level whose position in the ability scale is measured by the test items that he confronts. This is to say that some items discriminate between low and high achievers while others that are too difficult or too easy, do not, and are to be discarded from the test to improve on its

validity(Demars in Irwing et al., 2018). Therefore, a teacher requires the competence to know the appropriate steps and the ethical considerations to employ in order to discard some items and establish the validity and reliability of the test.

Research Methodology

This study used the descriptive survey research design. We evaluated the teachers’ competences through the use of teachers’ self -assessment questionnaire and another one for the students’ perception of the teachers’ practices, to support the teachers’ responses. Information was gotten from students to support the teachers’ assessment practices, as well as a sample of some teacher-made-tests examined for content validity. This study was carried out in the University of Bamenda, which is a state University in the Republic of Cameroon. It is found in the North West Region of Cameroon. The faculties in this University include; Arts, Law and Political Science, Science, Health Sciences, Economics, Education and Management Science. The University is also a host to the following professional schools:Higher Teachers’ Training College Bambili,The Higher Technical Teachers’ Training College,Bambili,The College of Technology, The Higher Institute of Commerce and Management,The Higher Institute of Transport and logistics, and The Higher National Polytechnique Institute of Bamenda (The University of Bamenda, 2016).

The target population of this study comprised the university lecturers and students drawn from the faculties of the University of Bamenda, excluding the professional schools.The accessible population was made up of three faculties namely; Faculty of Arts, Science, Law and Political sciences.

Table 1: The accessible population

Faculty	N° of permanent teachers	N° of students
Arts	24	1419
Law and Political science	10	806
science	36	1868
TOTAL	70	4093

Also, this study sampled seventy lecturers and three hundred students using clustersampling technique. Finally a total of 53 lecturers and 238 students answered the questionnaires. The sample size was determined from the Krejcie&Morgan table (Amin, 2005). First of all, the faculties represented the clusters and the respondents from the three faculties were selected using random sampling technique known as ballot. The names of the faculties were written on pieces of papers, folded and thrown on the floor and a boy was asked to pick three papers which represented the faculties to be sampled. The lecturers and students were further selected using convenient sampling due to time constraints. The sample comprised both male and female teachers, and students. In an attempt to answer the above question, the data obtained were analysed and presented on the table below with the use of SPSS V 20-30bt.

Presentation of Findings

The findings of this study have been presented following the two variables as expressed in the research objectives. These include the objective on the appropriate steps in test construction and the ethical considerations involved in test construction. First, findings on the appropriate steps in test construction follow the logic of the teacher and the students' responses of the questionnaire.

Table 2. Teachers' responses

S/ N	item	A	O	R	S	N	Mean	Remark
1	Deciding on test purpose	37	9	7			4.5660	.A
2	Outline content	42	8	3			4.7358	A
3	Use of test blue print	1	15	11	11	15	2.5472	R
4	Giving clear instructions	40	10	3			4.6981	A
5	Organising items in logical manner	14	14	17		1	3.2792	A
6	Conducting item analysis	12	19	13	3	6	2.4717	R
7	Keeping an item bank	24	19	6	3	1	4.1698	.A
8	Writing items a day ahead of test	7	3	12	19	12	2.5094	R
9	Allocating marks per item	39	7	6	1		4.5849	A
10	Considering objectives when setting tests	34	14	4			4.5849	A
11	Editing items for grammatical errors	29	14	7	1	2	4.2642	A
12	Constructing a marking scheme	31	11	9	2		4.3396	A
13	Ensuring clear graphics and diagrams	31	12	3	1	6	4.1509	A
14	Writing items for both low and high achievers	41	10	1	1		4.3396	A
15	Items do not cover all topics	5	11	14	13	10	2.3019	R
16	Seeking opinion of colleagues	6	11	14	17	5	2.9245	R
17	Questions do not provide clues for the next item	16	13	14	6	4	3.5849	A
18	Giving optional questions	13	7	16	8	9	3.1321	A
19	Using one long essay than shorter ones	35	12	4	1	1	2.1509	R
20	Setting easy tests to make students pass			7	11	35	2.3019	A

21	Reviewing items two days before item assembling	14	16	7	14	1	3.5660	A
22	Writing essay to elicit creative thinking	17	16	16	4		3.8679	A
23	Ensuring that language is understood	37	13	3			4.6415	A
24	Using long phrases in items	7	8	24	12		2.2553	R
25	Ensuring conducive environment	21	9	9	9	5	3.6038	A
26	Informing students on areas to read	13	12	16	9	3	3.4340	A
27	Informing students on materials needed	24	10	15	2	2	3.6038	A
28	Ensuring adequate time	43	7	3			4.7547	.A
29	Administering different test formats	14	10	15	5	9	3.2830	A
30	Having vigilance to detect cheating	32	16	5			3.2830	A
31	Marking following the marking guide	11	18	20	4		3.6792	.A
32	The marks reflect the students work	29	21	3			4.4906	.A
33	Making comments when marking essay	27	16	9	1		4.3019	.A
34	All learners are given equal chance	41	10	1	1		4.7170	A
35	Avoiding bias and unfair items	29	12	8	2	2	4.2075	A
36	Ensuring equity in testing	48	5				4.9057	A
37	Administering test whenever I wish	3	4	3	20	23	1.9434	R
38	Avoiding using test as a punitive measure	27	9	3	1	13	3.6792	A
39	Informing students on the date of test	35	12	4	1	1	4.4906	A
40	Releasing items prior to administration			7	11	35	1.5849	R
	Totals in %	42.5	21.1	17.2	9.5	9.7		

The instrument had 40 items covering test construction, administration and ethical considerations in testing. The items were rated on a five point likert scale with Always=5, Often=4, Sometimes=3, Rarely=2 and Never=1. Positive items were scored from 5 to 1, while the negative items were reversed before scoring. Items 8,15,19,20, 37,40 were negative items. The mean score of 3 and above shows acceptance with the statement in the item while mean score of less than 3 shows rejection. The overall analysis from the table above reveals that a majority of the teachers have good assessment practices with 42.5% recorded for ‘always’ meaning ‘very good’,21.1% for good, 17.2% for average ,9.5% for fair and 9.7% for poor practices.

From the table above, concerning the steps taken in test construction, rejection of item 3 showed that teachers do not use the test blue-print or table of specification while constructing the tests. Also, rejection of item 6 indicated that they do not conduct item

analysis, although some said they determine the item difficulty by intuition. Rejection of item 21 signified that they do not give the items for peer review and moderation. Rejection of item 8 on writing items a day ahead of test, acceptance of item 7 on item banking and item 21 on review of items before assembling them show good practices. The rejection of item 15 reveals that the tests items capture all or most of the topics treated. This is supported by acceptance of item 2 on outlining the content before writing items. Item 10 covering most objectives showed that teachers take into cognizance the content coverage.

A comparison was done using the student responses for some selected items from the teacher questionnaire. We shall see on the table below.

Table 3: Showing the responses of teachers and students compared on selected items

Item	response	A	O	S	R	N	Mean	Remark
Questions represent content taught	T	42	8	3			4.7358	A
	S	90	73	66	4	5	4.0042	A
Clear instructions	T	40	10	3			4.6981	A
	S	139	36	38	12	13	4.1597	A
Organise items in order of difficulty	T	14	17	14	7	1	3.6792	A
	S	37	39	65	54	42	3.1134	A
Marks allocation	T	39	7	6	1		4.5849	A
	S	105	37	54	29	13	3.8067	A
Edit grammatical error	T	29	14	7	1	2	4.2642	A
	S	12	62	95	59	10	3.0294	A
Making graphics visible	T	31	12	3	1	6	4.1509	A
	S	59	69	79	23	8	3.6218	A
Write items for both levels to understand	T	27	19	5	2		4.3396	A
	S	23	63	95	41	16	3.1513	A
Giving optional questions	T	13	7	16	8	9	3.1321	A
	S	73	54	76	24	10	3.8739	A
long essay not short ones	T	4	1	12	18	18	2.1509	R
	S	54	49	100	23	12	3.4622	A
	T	17	16	16	4		3.8679	A

Essay eliciting creativity	S	65	55	84	25	9	3.5966	A
Language appropriate	T	37	13	3			4.6415	A
	S	65	63	75	28	7	3.6345	A
Using long phrases	T	7	8	24	12		2.2553	R
	S	20	46	96	60	16	2.9748	R
Total in %	T	47.2	20.8	17.8	8.5	5.7		
	S	26	22.6	32.4	13.4	5.6		

T is for teachers, S for students

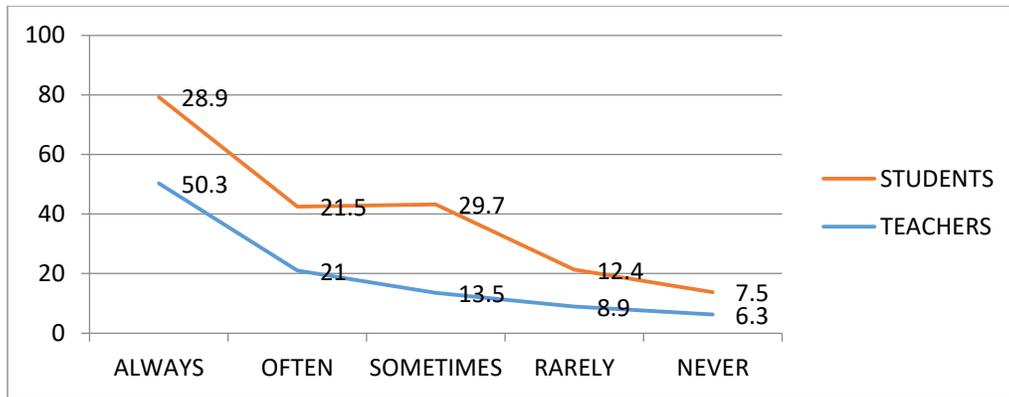


Figure 1. Teacher and student responses compared

The results show that the students confirmed the teachers’ practices especially for what they do always, often, rarely and never. Controversial views were recorded as to what happens sometimes. For example, the item on teachers having high vigilance to detect cheating was refuted by the students as they sometimes cheat in the examination. Also, most of the teachers said they always organize items in order of difficulty; and inform the students on areas or topics to read for the test, but the claims were not accepted by the students. Another point concerned favoritism for which the students claimed that teachers favour some students when marking. Despite all these, the students still accepted that the marks they have truly represent what they deserved.

A further test was conducted to find the differences between the responses. An independent T-test was conducted for the two groups of respondents, that is, 53 lecturers and 238 students to match with. The results showed that though there were some little differences, these were not significant in the mean ratings of the students and teachers. As the p-value (sig=0.332) was greater than 0.05 or 0.1. This is given by the following values, $t(94.6)=-.291, p>0.1, df\ 92$ (where t=t-test, p=significance value and df=degrees of freedom.)

T-Test

Table 4 Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
scores	lecturers	53	91.8868	9.32462	1.28083
	students	238	92.3193	11.60128	.75200

Table 5 on Independent Samples Test

Levene's Test for Equality of Variances		t-test for Equality of Means						
F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
.946	.332	-.254	289	.800	-.43254	1.70504	-3.78841	2.92334
		-.291	91.639	.772	-.43254	1.48527	-3.38257	2.51750

Results from Observation Checklist Used

The observation checklist used to assess 30 samples of different test papers set by the teachers also revealed that the teachers have good practices as shown by the graph below. The papers were rated on the main headings: test format, directions, test length language level, graphics, grammatical and typographical errors, and thinking levels according to Bloom.

Table 5 showing results from the observation checklist on test samples

Test papers	Test format		Directions			Grammatical or typing errors	Clear diagrams	Test length	Thinking level		
	essay	objective	instruction	Time allocation	Mark allocation				K C	A A	S E
total	10	20	27	18	25	2	2	>30, >3	Level covered		
30	34%	66%	90%	60%	83%	6.6%	100%	good			

The teachers used both essay and objective test formats. The essay were mostly short essays, with a few that used long questions only. The objective tests took the form of multiple choice questions (MCQ), filling the blank spaces, and structured questions from a passage or calculations. Clear instructions were written on majority (90%) of the papers although time and mark allocation were not indicated on some of the papers. The test lengths were good, equal or above 30 for objective test and 3 and above

for essay. Some papers had four items only but with less mark allocation which showed that it lacked content coverage. Little grammatical errors were made, diagrams were clear and visible. Most questions asked covered the first three lower thinking levels of bloom ranging from knowledge, comprehension to application. A few questions went up to analysis. The verbs used were –give, state, cite, define, draw, compare, discuss, calculate, describe, comment, briefly explain and analyse.

Discussion

The discussion of the findings has been carried out following the two sub variables of the study. These include, the appropriate steps and the ethical considerations in test construction. The discussion centres on setting test items based on a purpose, organisation of items from lower levels of difficulty to the higher, taking record of what is taught on a daily basis, use of test blueprint, item analysis, giving test items for moderation and lastly revelation of test items prior to the examination period.

Following the analysis, ‘always’ and ‘often’ have been considered as good practices, ‘sometimes’ for average, while ‘rarely’ and ‘never’ stand for bad practices except for reverse items. It was revealed that the teachers had overall good assessment practices (63.6%). They followed most of the steps in item writing, administration and scoring of tests and also considered ethical issues in testing to avoid taking responsibility for the students’ failure. Almost all the teachers decide on the purpose of the test prior to the setting of the items. They outline the content taught, allocate marks per item, organise items to cover the objectives and lastly they ensure that the language used is comprehensible. Anyanwu, and Onwuakpa (2015) hold that testing is meant to provide specific and valid interpretations of the students’ knowledge and skills following the scores attributed. Some lecturers argue that at the tertiary level of education, the purpose of their teaching is research and not fact finding.

Also, most of the lecturers (84.9%) organise items in a logical manner beginning from easy to difficult ones although (13.1)% of the teachers saw no need structuring their tests that way. Ukwuije (2015) advocates item re-arrangement to enhance test validity. This is precisely because if the most difficult items are presented first, this will undoubtedly slow the speed of performance, and provoke un conducive psychological environment within an examinee that is not able to answer the first set of questions.

In addition, 81.1% teachers do not write out items a day before the test. Farrant (1980) and Olubodun (2007) hold that constructing good tests is not an easy task. It requires much thinking, rethinking and reading before constructing test items. This is because a good test does not just exist, but it is made to exist. Test items should be written as the lessons are taught and put in item banks. This may be an exercise book, a folder in the computer although some teachers claim to store them in their minds. Further, some teachers said they develop resource item banks as they teach the lessons. Also, they review the items before assembling them to form the test. They edit the items for grammatical errors and make sure graphics are clear. About 90% give shorter essay questions than one long one.

Moreover, about 70% of the lecturers said their tests cover all topics while 30% did not. They explained through an informal interview that at times they tests topics that are not well understood to see where remediation can be done leaving out sections that have been mastered. The problem here is how do they determine what the students' have mastered? Also, the teachers said some topics are too broad to be covered in a test. This confirms the findings of the report of Newell (2002), who asserts that teacher-made tests usually measure only a limited part of a subject area. They do not cover a broad range of abilities and they rely too heavily on memorized facts and procedures. For some teachers, their questions do not provide clues for the next item but for other they do provide because in problem-solving approach, the answers given to prior questions constitute prerequisites for completing the whole problem.

Besides, the findings revealed that 78% lecturers give optional questions while 22% do not. This gives students the choice of questions and there cannot be a comparison of their knowledge and skills in a course because they do not answer the same test items. According to Farrant (1980), optional questions should be avoided because they tend to favour the students by offering them choices. They also render the task of the teacher more difficult because he/ she is expected to determine and identify the options of comparable difficulty. The teacher may moderate his marking guide without being sensitive to these prerequisites thereby favouring some students as opposed to the case of their counterparts. A good test paper should control the freedom and require the candidates to do similar or the same task so as to render the results of the test more objective and easy to compare (Madsen, 1983 in Zhu & Han, 2011). A given test item should mean the same to all students. Although all the teachers set questions taking into consideration the ability level of the students, that is, for both the high and low achievers to understand, they hardly set easy tests just to make the students pass, but simply to award marks the students deserve. The test should include items at all levels of difficulty, that is, the test items should vary uniformly in difficulty from the most difficult one, to the one which will be answered correctly.

For the problems identified, findings revealed that most teachers (70%) do not use the test blue print in constructing tests. A well written test allows the teacher to accurately and consistently measure students' mastery of specific contents taught in class. Poorly designed test items can lead to inaccurate measurements of learning and provide false information regarding student performance. To construct good items that cover greater content and thinking levels, a test blue print is indispensable. According to Ukwuije (2015), a table of specification allows teachers to clearly define the scope and emphasis of the test, to relate the objectives to the content, and to construct a balanced test. In addition, it provides guidelines in obtaining a representative sample of test tasks in terms of emphasis on content. It is thus recommended that teachers use test specification tables when constructing assessment tasks especially those of the objective-test type.

The findings also proved that most teachers do not carry out item analysis. From the responses, some teachers do so by inductive reasoning. This helps them to eliminate difficult and easy items. Item analysis is a technique used to determine how difficult

and easy the test items are. Item difficulty is obtained by determining the percentage of those that get an answer correct in comparison with those that get it wrong. When most scores cluster at the top or middle or at the bottom, then they lackdiscriminality. Item analysis is helpful to test planners to ascertain if a test possesses the qualities/characteristics it ought to demonstrate in order to serve as a good test item (Olubodun, 2007).

Also, the responses revealed that most teachers do not give their tests for review. Kinyua and Okunya, (2014) found out that many teachers do not subject their exams to moderation. This attitude affects the level of reliability and validity of teacher made tests. Generally, the teachers also take care to administer and score the test well. They ensure that adequate time is allocated, students are informed on the date of test without releasing questions prior to administration. This ensured that the marks students get reflect their work. However, some students observed that they still exploit the vigilance of teachers by cheating during exams. Test items have to be kept in confidentiality. If candidates are able to see questions prior totaking a test, then most of them will seek to memorise rather than reflect, think, create and be critical. The ability and validity of the results of such students will be questionable. Anja opines that “it is unethical to divulgate your intended items prior toadministration” (2009 p.178). What was evident is that all the teachers were conscious of the ethical issues to consider during administering a test. These include; giving equal chances to all candidates, avoiding biased items withoutgiving special preferences to a student or a group of students, and awarding marks that the students deserve.

In a nutshell, all the responses and the papers sampled revealed that the lecturers in the Faculties in the University have good test development practices, (but not the best, only 63.6%) and that they are capable of producing results that are valid and reliable. However, they need further training and capacity building avenues to improve on these practices. Owing to all these, this study recommends thatteachers should use the self-assessment questionnaire to verify the appropriateness of their test items. Testing is a systematic practice and not a game of chance. Heads of Departments are expected to exploit the findings of this study to ensure quality in teaching and evaluation practices. Seminars/ workshops and training sessions should be frequent to help teachers develop appropriate and ethical test practices.

References

- Afemikhe O. A & Yakube N. D. (2015). *Nigerian teachers' utilization of test construction procedures for validity improvement of achievement tests*. University of Benin, Benin City, Nigeria
- Agu N. et al. (2013). *Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skills*; University, Awka, Nigeria. <http://www.academicjournals.org/ERR> accessed 13/04/2017.
- Anja S.N. (2009). *Teaching practice and principles made simple*. Bamenda: Patron Publishing House.
- Anyanwu, I. E., & Onwuakpa, F.I (2015). Improving Validity of Tests through Improved Test Development Procedures, A Paper Presented at the Annual International Association for Educational Assessment (IAEA) Conference Held at the University of Kansas,

Lawrence, Kansas, USA from October 11 to 15, 2015, Retrieved from Dalumpines L. (N.D). *Effective Teaching: General Principles In Test Construction*

Demars, C. E. (2018). "Classical Test Theory and Item Response Theory" in Irwing, P. et al., (eds.), *The wiley handbook of psychometric testing: a multidisciplinary reference on survey, scale and test development* in <https://online.library.wiley.com/doi/book/10.1002/9781118489772>. Accessed 27/08/2018 at 3:02pm.

Ethical Standards in Testing: Test Preparation and Administration. (2001). *Washington Educational Research Association White Paper*: Wera Professional Publications Volume 1.

Farrant J.S. (1980). *Principles and practice of education*. New edition. Singapore: Longman Publishers.

Kinyua K. & Okunya O.L. (2014). *Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya Department of Psychology*. University of Nairobi, Kenya.

Nworgu (2015). *Measurement And Evaluation, Theory and Practice*, University printing press, Nsukka.

Olubodun, O. J (2007). *Test Construction Techniques and Principles: Study Guide*. The University of Lagos, in <https://www.researchgate.net/publication/265085817>. Accessed 25th September 2018.

Roelofs E., & Sandars P. (2007). Towards a framework for assessing teacher competence. *European Journal of vocational training*, N^o 40- 2007/1. Retrieved from <http://www>

The University of Bamenda (2016). In the *University of Bamenda wiki*, retrieved on June 23, 2017 from [http:// University of Bamenda Wikipedia. Html](http://University of Bamenda Wikipedia. Html)

The University of Texas at Austin Faculty Innovation Center. (2017). *Ethical Considerations* Retrieved on June 23, 2017 from http://www.Ethical Considerations _ Faculty Innovation Center.html

Ukwuije P.A. (2015). "Educational Assessment: A Sine Qua Non For Quality Education". A paper presented at the 83rd edition of Inaugural Lecture Series university of Port Harcourt, Nigeria.

Wu, M. et al., (2017). "Classical Test Theory" in *Educational Measurement for Applied Researchers*. 73-90 in https://link.springer.com/chapter/10.1007/978-981-10-3302-5_5. Accessed 27/08/2018 at 3:02pm.

Zhang Z., & Burry-Stock J.A, (2003). Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied Measurement in Education*, 16(4), 323-342.

Zhu J. & Han L. (2011). Analysis on the Main Factors Affecting the Reliability of Test Papers, *Journal of Language Teaching and Research*, Vol. 2, No. 1, pp. 236-238.