

Instructors' Test Construction Competences and their Implications on Teaching and Learning in Tanzanian Universities

Asia M. Rubeba

Department of Educational Psychology and Curriculum Studies

College of Education, The University of Dodoma

Email: rubeba2016@gmail.com,

&

Francis William

Department of Educational Psychology and Curriculum Studies

College of Education, The University of Dodoma

Email: kyambo20001@yahoo.com

Abstract

This preliminary study was conducted in order to examine instructors' competences in assessing teaching and learning outcomes. Specifically, the study focused on examining the current status of test construction practices among university instructors in Tanzania and challenges facing them during test construction practices. The study employed cross sectional survey and case study designs in order to examine the magnitude of the problem prior to initiation of the actual study. Simple random, purposive and convenience sampling techniques were employed in the study. The study employed the sample size of 40 instructors and 120 students who were used as a supplement.

Findings revealed that instructors were properly practiced the following domains: they ensured that their test items had content validity; they set tests in such a way that both low and higher achievers could understand the concepts and they outlined the contents before setting the test items. On the other side, some of the instructors were neither aware about the blueprint and Bloom's taxonomy nor they were using the two. Again, it was noted that feedback was not given to students on time. The findings revealed several challenges facing instructors during test construction practices including large class size, lack of training on test construction preparation and lack of enough resources. Based on study findings, it is recommended that the number of enrolled students should be proportional to the number of available instructors. Instructors should be encouraged to attend seminars/workshops on test construction.

Keywords: Assessment, instructors' competence, test construction practices, test items.

1.0 Background to the Research Problem and its Context

The current education system demands that University instructors should have command of different forms of classroom assessments. Specifically, instructors need to be able to align teaching, learning and assessment that are consistent with the twenty first century (21st) educational aspirations (Omari, 2011). A number of scholars have, in common, agreed that a good constructed test is significant to students' learning (Adodo, 2014: Ahmed *et al.*, 2014: Perumanathan, 2014: Sales, 2013) and has notable significant impact on students' performance (Desforges, 2013: Sadler, 2013: Taras, 2007). Providing a highly particularized example, Gronlund and Waugh (2009) echoed that a well-constructed test provides valid and reliable results that serve as tools for making decisions at different levels. The results could be used by policy makers to change or improve their programmes, employers could use results to hire individuals and results could be used by universities to examine their programmes whether they suit the job markets or not. Therefore, all assessments should be designed to be of high quality so as for them to measure the intended learning outcomes, provide useful and accurate information as well as meet technical and psychometric standards (Sanga, 2016; Gronlund & Waugh, 2009; Popham, 2008).

In relation to this, Sarita (2005) noted that good tests make instructors avoid all undue pressure, promote self-confidence and thinking skills to students and therefore, stimulate academic achievement (Akiri, & Ugborugbo, 2009). In contrast, when a test is perceived as poor, the underlying effect is that of the poor test construction, which does not embrace the idea of valid and reliable test. This perhaps happens when an instructor is incompetent. It is the instructor who determines what, who, when and how to measure (Mpapalika, 2013). This notion is also congruent to the idea by Agu, Onyekuba and Anyichie (2013) who reiterated that without a competent instructor on assessment, construction of quality test cannot be attained at its fullest. In this context, a competent instructor is expected to use highly authentic assessment approaches that motivate students to take more responsibility for their learning and stimulate their abilities to apply learning to a wider range of knowledge (Ogula & Rubeba, 2013) as opposed to traditional paper-pencil techniques, which encourage memorization and less basic skills development (Binde, 2012; Ndalichako, 2004).

Indeed, literatures inform that assessment of students' achievement, worldwide, came into use for the purpose of providing grades and certificates. Experience from a number of countries such as United States of America (USA), United Kingdom (UK), Malaysia, South Africa, Uganda and Tanzania indicates that previously, assessment was used for grading and providing certificates, something that is termed as high stakes examinations in which scholars view it as a traditional way of assessing students' activities (Matuvo & Zubairi, 2015; Lines & Mason, 2005; Msila, 2014; Omari, 2011). The strategy used mostly was paper and pencil that was perceived to measure the large part of the content coverage and easy of scoring and recording results (Gronlund & Waugh, 2009). Within the context of Tanzanian education system, several attempts have been made to boost the education standards, the latest of which is implementation of paradigm shift from content-based testing to performance and competence-based assessment (Mhando, 2012). This innovation places several obligations to the university instructors by requiring them to be competent in using authentic assessment methods such as portfolio, experiments, case studies, projects, observation checklists, self and peer to measure students' learning outcomes (Mwebaza, 2010; Mpapalika, 2013; Mbuja, 2014; Gichuhi, 2014). However, evidence from the literatures suggest that concrete corresponding changes in assessment practices have been lacking thereby creating a little doubt if instructors are competent enough in assessing the students and whether or not the methods currently used really measure what is supposed to measure (Matuvo and Zubairi, 2015; Agu, Onyekuba and Anyichie, 2013; Osadebe; 2015). Sanga (2016: 18) proposes that for effective teaching and assessment, "it is essential to assess what students have achieved, but of critical relevance is to assess how they are learning."

Above and beyond, students in the universities are expected to acquire and master higher knowledge and skills, attitude and values that may be life-long useful in different careers for different purposes [Tanzania Commission for Universities (TCU), 2016]. Berry and Adamson (2011) believe that knowledge and skills shape the future lives of learners and instructors play a significant role in improving teaching and learning. Bryan and Clegg (2006) state that, assessment needs to be an instrument for individuals' liberation. Any change in it needs not compromise the standards but it is in light of this understanding that deliberate and immediate initiatives are urgently needed to reform assessment practices in universities in order to suit into the current move that is competence-based rather than content-based.

Against this background, it suffices to conclude that instructors' competence in assessing teaching and learning activities in universities in Tanzania is questionable and hence, further investigation was needed to explore the real existing state of the art. Therefore, this study aimed at examining instructors' test construction practices and their implications on teaching and learning in Tanzanian Universities.

1.1 Objectives

The study had the following specific objectives:

- a. To assess the current test construction practices by university instructors for measuring learning outcomes; and
- b. To examine challenges facing instructors in constructing test items.

2.0 Methodology

This study was conducted in Dodoma and Mbeya regions in three (3) selected higher learning institutions, namely, the Saint John's University of Tanzania (SJUT), Mbeya University of Science and Technology (MUST) and the University of Dodoma (UDOM). Selection of the three institutions was appropriate in that among other assessment methods, which measure teaching and learning outcomes, paper and pencil were used in these universities, like in many other universities in the world. The study employed cross-sectional survey design using quantitative research approach to generate and analyse data from the field. The study expected to gather data from instructors with different educational backgrounds and teaching experiences at one point in time. In addition, cross-sectional survey design was used in the study to understand the problem as it looked then without referring to its background. Furthermore, case study strategy was used for generating qualitative data. The choice of the case study strategy was due to the fact that the researchers were unaware of the current status of test construction practices among university instructors and its influence on the teaching and learning process. The target population comprised of all university instructors and students who were found at the said three respective institutions.

Simple random, purposive and convenience sampling techniques were used to select 160 respondents in categories of 120 students and 40 instructors. Questionnaire with close-ended and open-ended questions together with interview guide were used to obtain information from the respondents.

Data were analysed by using descriptive statistics and themes. The mean scores were set at 4.00 (cut-off points) whereby three categories of mean scores were formed. Cronbach's Alpha test was used to determine internal consistency of the questionnaires and 0.76 was obtained. In terms of validity, Pearson's Product Moment correlation was performed in which obtained results accounted for 0.79. Both data indicated positive and high correlation.

3.0 Results and Discussion

Recall, findings from this study were sought based on the following two objectives: (1) to assess the current test construction practices by university instructors for measuring learning outcomes and (2) to examine challenges facing instructors in constructing the test items.

3.1 Demographic Characteristics of Respondents

Table 1: Instructors' demographic characteristics (n=40)

Demographic characteristics		Frequency (f)	Percentages (%)
Sex of the respondents	Male	26	65
	Female	14	35
Professional development workshop/ seminar in test construction	Attended	16	40
	Not attended	24	60
Teaching Profession	Yes	12	30
	No	28	70

Source: Field data (September to December, 2017).

Table 1 indicates that out of 40 instructors, there were 26 (65%) males and 14 (35%) females who participated in the study. This is an indicator that universities provide an equal opportunity to both genders - male and female instructors. The study done by Agu, Onyekuba and Anyichie (2013) found that gender had no influence on instructors' competence in assessment. Again, 16 (40%) of the respondents had attended professional development, while 24 (60%) had not attended any training on test construction (Table 1). That might be due to the fact that there is neither mandatory attendance for training nor a policy that forces instructors to attend training. Matuvo and Zubairi (2015) in their study, observed that academic members of staff

in Ugandan universities lacked appropriate assessment skills in evaluating their students. From the study, it was recommended that formal assessment training programmes should be made mandatory to all academic members of staff in universities in Uganda in order to improve their assessment skills to ensure quality in the way they assess students (*ibid.*). Furthermore, Table 1 indicates that 12 (30%) instructors were teachers by profession and 28 (70%) of instructors were non-teachers. From this finding, it is clearly shown that most of the university instructors were not teachers by profession. It might be due to the modalities of their employment whereby among other factors, higher Grade Point Average (GPA) is a factor for being hired in the universities in Tanzania (Tanzania Commission for Universities, 2014, 2019). This is why most of them agreed with the notion that they were outlining the content covered for the semester before setting test from them and they were not preparing blueprints (Table of specifications) as guides during test construction. On the other side, instructors who are teachers by profession were aware about the blueprint (Popham, 2008), for they were oriented during their course of study. However, scholars indicate that most of them do not use it in setting test items (Chung, 2008, 2017; Popham, 2008).

Table 2: Students' Demographic Characteristics (n=120)

Demographic characteristics	Frequency (f)	Percentages (%)
Sex of the respondents		
Male	65	54
Female	55	46

Source: Field data (September to December, 2017)

Table 2, the findings show that, the study employed 65(54%) of male students and 55 (46%) of female students in three universities respectively to represent other university students. This shows that universities in Tanzania offer equal based opportunity to both genders that is male –female students in terms of higher learning education. This is an indicator of meeting sustainable development goals (2030 agenda), specifically goal number 4 which stress on *inclusive and equitable quality education, and promote life learning opportunities for all* [United Nations Educational, Scientific and Cultural Organization (UNESCO), 2017].

3.2 Existing Status of Test Construction Practices among University Instructors

This section gives a summary of findings on the existing status of test construction practices. The Tables are presented based on the extent to which instructors practiced a given domain either above, moderate, or below the cut-off point. The three categories are expressed by using mean scores as explained in the methodology part.

Table 3: Existing status of test construction practices among university instructors with mean scores above the cut-off point ($\bar{x} = 4.00$)

Domains measured	Mean (\bar{x})	Std. Deviation (sd)
I ensure tests have content validity.	4.35	0.662
I write test so that both high and low achievers can understand.	4.28	0.847
I outline the content covered for the semester before setting test from them.	4.20	0.687
I organize test items in a logical manner.	4.15	0.580
I assign scores for each test item.	4.08	0.730
I analyze test results to identify areas of difficulty.	4.08	0.829
I give clear instructions to guide the test takers.	4.08	0.616

Source: Field data (September to December, 2017)

Table 3 indicates findings of the current status of test construction practices among university instructors in the selected universities. The findings revealed to have domains with mean scores above the cut-off point as follows:

“I ensure tests have content validity (Mean(\bar{x}) = 4.35 & standard deviation (sd) = 0.662), I write test so that both high and low achievers can understand (Mean(\bar{x}) = 4.28 & standard deviation (sd) = 0.847), I outline the content covered for the semester before setting test from them (Mean (\bar{x}) = 4.20 & standard deviation (sd) = 0.687), I organize test items in a logical manner (Mean(\bar{x}) = 4.15 & Standard deviation (sd) = 0.580), I give clear instructions to guide the test takers (Mean (\bar{x}) = 4.08 & Standard deviation (sd) = 0.616), I assign scores for each test item (Mean(\bar{x}) = 4.08 & Standard deviation (sd) = 0.730) and I analyze test results to identify areas of difficulty (Mean(\bar{x}) = 4.08 & Standard deviation (sd) = 0.829)” [Table 3].

Therefore, from Table 3, the findings imply that instructors were paying much attention by ensuring that their tests had content validity and they wrote their items in such a way that both high and low achievers could understand them. Content validity answers the question, ‘how adequately does the sample of assessment tasks represent the domain of tasks to be measured’ (Gronlund & Waugh, 2009). However, a good test paper ought to consider, among other important aspects, mixing of test items, which test lower and higher levels of understanding and also, coverage of course contents that students learned in the class (External examiner, 2017).

These findings were contrary to findings obtained from students on the same matter See Table 4.

Table 4: Students’ responses on instructors competence in test construction practices (n=120)

Statement	Agree		Disagree	
	F	%	F	%
Instructors use difficult words in constructing test items	78	65	42	35
Instructors set test with ambiguous items	69	57.5	51	42.5
Instructors set test items that covers small portion of the contents taught in the class	65	54.1	55	45.9

Source: Field data (September-December, 2017)

The questionnaire administered to 120 students (Table 4) revealed that 78 (65%) agreed with the notion that some of the instructors used difficult words in constructing test items. In contrast, 69 (57.5%) said that some of the test and examination items contained ambiguous items and 65 (54.1%) agreed with the notion that some of the instructors assessed only a small portion of content covered in a semester. On this regard, one student said that,

“Many students fail their tests and examinations not because they do not know the content taught but because of the nature of questions asked by the instructors. Some of the questions contained more than one meaning and others had very difficult concepts, which one cannot understand them easily” (Student 1, 2017).

Another student had this to say,

“It is okay that we do tests, but there are very few questions set in test papers. There is no agreement between the materials covered in class and the number of questions set. We take time to peruse the materials. I find myself scoring low marks. Really, it is very discouraging” (Student 2, 2017).

Another student commented on the nature of items by saying the following:

“Some questions are very simple to understand by all students, regardless of their understanding abilities, though there are some tough and challenging questions. If one is not keen enough, he/she may interpret the question in a different way and end up losing marks” (Student3, 2017).

In due regard, from views, it shows that the constructions of test items in sampled universities did not follow the principles of test construction. In such situation, there is likelihood of compromising the quality of tests that might lead to misinterpretation of students’ abilities. In fact, a good test should be able to discriminate between the lower from the higher achievers. However, the power of the test to discriminate among students should not be attributed to students’ failure to understand the question set but rather, to the nature of students’ abilities. Notably, instructors need to engage students in learning by giving them challenging questions but in a supportive environment (Khan, 2012).

Table 5: Existing status of test construction practices among university instructors with moderate mean scores

Domains measured	Mean (\bar{x})	Std. Deviation (sd)
I set tests with due regard to the time available for testing.	4.00	0.906

Table 5 indicates the domain with moderate mean scores (4.00 cut-off points): ‘I set tests with due regard to the time available for testing (Mean (\bar{x}) =4.00 & Standard deviation (sd)=0.906).’

Table 6: Existing status of test construction practices among university instructors with domains below the mean scores (cut-off point)

Domains measured	Mean (\bar{x})	Std. Deviation (sd)
I prepare a test blueprint as a guide in the test construction.	2.38	1.005
I use Bloom's taxonomy in setting test items	2.35	1.145
I subject test items for item analysis.	1.83	0.844
I keep a resource bank of questions that can be referred to when setting tests.	3.43	1.174
I add enough test items to cover all the requisite levels of domains	3.68	1.185
I ensure that the items are measuring the determined objectives.	3.63	0.897
I set essay items that elicit creative and imaginative answers from the students.	2.55	1.218
I prepare a marking guide while constructing the test.	3.15	1.210
I add sufficient items to cover the appropriate instructional units.	3.78	1.097
I submit test items for moderation to the Head of department.	1.98	0.920
I avoid overlapping alternatives in writing objective tests.	3.38	0.897
I limit essay tests to high level objectives.	3.70	1.137
I use items that measure knowledge.	3.65	1.406
I avoid repetition of the questions during test items construction.	3.83	0.813
I provide feedback to students about their performance promptly.	2.90	0.911
I comment on students' script to guide learning.	2.95	0.959
I ensure tests have face validity.	3.73	0.933

Besides, Table 6 shows domains with low mean scores below cut-off points, which include:

“I avoid repetition of the questions during test items construction (Mean(\bar{x}) =3.83 & Standard deviation (sd)=0.813), I add sufficient items to cover the appropriate instructional units (Mean(\bar{x}) =3.78 & Standard deviation (sd)=1.097), I limit essay tests to high level objectives (Mean(\bar{x}) =3.70 & Standard deviation (sd)=1.137), I comment on students’ script to guide learning (Mean(\bar{x}) =2.95 & Standard deviation (sd)=0.959), I promptly give feedback to students about their performance (Mean(\bar{x}) =2.90 & Standard deviation (sd)=0.911), I prepare a test blueprint as a guide in the test construction (Mean(\bar{x}) =2.38 & Standard deviation (sd)=1.005, I use Bloom’s taxonomy in setting test items (Mean(\bar{x}) =2.35 & Standard deviation (sd)=1.14), I submit test items for moderation to the Head of department (Mean(\bar{x}) =1.98 & Standard deviation (sd)=0.920) and I subject test items for item analysis (Mean(\bar{x}) =1.83 & Standard deviation (sd)=0.844).”

On the other side, instructors in the selected sampled universities showed little practices on the following domains as indicated in Table 6:

“I comment on students’ script to guide learning, I give feedback to students about their performance, I prepare a test blueprint as a guide in the test construction, I use Bloom’s taxonomy in setting test items, I submit test items for moderation to the Head of department and I subject test items for item analysis.”

Findings on provision of feedback concur with results from the study by Sales (2013) who observed that instructors were not providing feedback to students. With respect to findings from this study, one underscores the importance of timely giving feedback to students. The feedback students receive whether oral or written has an impact on their learning (Ahmad *et al.*, 2014). Students need to know about their progress (Ogula, 2009; Odili, 2014) so that they could make efforts throughout their school life time.

Furthermore, in terms of preparation of test blueprint, instructors should prepare it in order to have an adequate sample of items set in test paper. The test blueprint indicates the content covered, weight and domains to be measured (Popham, 2008; Omari, 2011a & b). Lack of preparation of Table of specifications (blueprint) may cause mismatch between the content covered and the instructional objectives (Ogula, 2009). Furthermore, use of Bloom’s taxonomy during setting questions signifies the extent to which the learning outcomes were covered and mastered by the students (Anderson &

Krathwohl, 2001). Besides, the taxonomy calls for assessment to be done not only on content part but also on competence level in order to determine the level of students' understanding and application of the learned materials in a different environment (*ibid.*). With respect to this, higher order and lower order thinking skills have to be measured (*ibid.*). Therefore, failure to use it could accelerate lack of balance among expected learning outcomes (*ibid.*).

Furthermore, the findings revealed that instructors hardly submitted test items for moderation purposes. From this finding, it shows that prepared test items in the sampled universities lacked legitimacy could compromise the standards. The External Examiner in his report on moderation of university examinations observed that lack of serious internal moderation could result into invalid test and finally, students could perform poorly, an aspect of which is injustice (External Examiner, 2017). Tests act as a central role of the teaching and learning process. They help in measuring sample behaviour (Babygeya, 1998). Thus, moderation of test papers, whether internal or external, makes an important part of the quality of the examination standards. Therefore, efforts should be made to make sure that all test items are submitted to the respective authorities for moderation purposes prior to commencement of tests.

Likewise, it was noted that analysis of test items was also given less attention than analysis of test results (see Table 3 & Table 6). However, it should be borne in mind that it is the role of the instructor to determine the extent to which the tool used to measure learning outcome suffices and in turn, brings meaningful learning (Adodo, 2014). Ndalichako (2015: 2) in her study stated that, "teachers are expected to play a dual-role of facilitating students' learning and of implementing classroom assessment in a manner that will enhance meaningful learning outcomes." Besides, Kurebwa (2012) explains that teachers are required to have a clear understanding of their learners' abilities. Therefore, without a clear understanding of learners' abilities, teachers have no definite idea whether or not learners are meeting standards (Kurebwa, 2012: 45). In due regard, instructors need to analyse test items in order to determine the level of difficultness and discrimination prior to actual commencement of the test. The results obtained from the analysis could be used to give feedback to both instructor and students and where the concepts are misconceived, they need to be corrected.

4.3 Challenges Faced by Instructors during Construction of Test Items

Findings revealed that instructors faced several challenges including, but not limited, to large class size, lack of training on test construction and lack of resources. Table 7 gives explanations on those challenges.

Theme	Sub-themes	Explanations	Suggestions
Challenges experienced by university instructors during construction of test items.	Large class size	Instructors explained that most of their classes had big number of students in that case they were limiting the number of questions in the test papers to make it easy to mark.	Classes to have proportional number of manageable students with the number of instructors.
	Lack of resources	About 75% of respondents were concerned about lack of resources in terms of man power and materials.	To increase budget that to be used to buy the materials and the government to employ more instructors.
	Lack of training on test item preparation	It was contended by instructors that, there are limited trainings on test construction. This is why the construction of test is done as business as usual and some instructors were simply using their own experiences. Again, sometimes the training was given during high peak time that is a few days before commencing examination. In that way, few instructors were attending while others were busy finalizing the course works and it was not obligatory to attend.	Universities to continue offering professional development/ in service training on various issues including preparation on test construction. There is a need of finding proper time to offer workshops/ seminars in order to make majority of instructors attend. Again, instructors should be encouraged to attend.

4.3.1 Large class size

Large class size has been observed as a big challenge during preparations of constructing test items. During interviews, most of the university instructors mentioned about it. Showing his concern, one instructor from University J had this to say,

“I have two different courses to teach with total number of students 950. With such number of students, it is not easy for me to measure students’ understanding in different learning outcomes, either to provide several activities. Also, it is not easy to give many essay questions. Rather, you give only one, which makes it easy for you to mark” (Interview with instructor 1, University J, 2017).

The quote suggests that due to large number of students in the selected universities, instructors were having insufficient time to follow the recommended test construction principles during test construction. Such situation was making construction of tests invalid as well as unreliable test items. Matuvo and Zubairi (2015) affirm that the increase in enrolment rate, which is not proportional to the number of available instructors has an effect on assessment practices.

4.3.2 Lack of resources

Instructors demonstrated lack of teaching and learning resources such as photocopy machines, printers and projectors as barriers towards their efforts in improving teaching and assessment as well. In showing her concern, one instructor had this to say,

“The nature of test item we construct is also affected by available facilities. For example, in this department, only the head of department has access to the printer, computer and photocopy machine for administrative duties. Others don’t have such privilege. So before I think of how I will test my students, I have first to think of how much cost will I incur in the process and this will have implication on the type of test items to be set. In fact, we compromise good tests due to our inability to meet the costs involved. This is normal practice here” (Interview with Instructor 3, University I, 2017).

From this finding, it is clearly observed that university instructors face serious problems in terms of shortage of facilities. This finding is in agreement with Mapolisa and Tshabalala (2013) who found that most of the public

universities in Zimbabwe had adequate teaching and learning facilities including printers, bond papers, paper punches, and files. From their study, it was recommended that the Government of Zimbabwe needs to provide adequate funding for public universities that could be used for buying teaching and learning equipment including stationeries (*ibid.*). UNESCO (2012) expresses that Africa faces many obstacles in its efforts to improve education, such as lack of resources, inequitable access to services, and absence of mechanisms to ensure quality in training and provision of education.

4.3.3 Lack of training on test construction

It was observed from the study that some universities neither provided professional development on teaching nor training on test construction. Besides, the universities whereby seminars were conducted, few instructors were able to attend. During interviews, one instructor explained that,

“We have rarely test/examination items construction orientation or seminars for instructors here in this university. We could expect to have one or two each academic year because these instructors are only subject experts, not teachers by profession. So they only use their experience to construct test items. That is why you find they know nothing about test construction procedures or standards of good test items” (Interview with Instructor 1, University H, 2017).

Professional development is very important in teaching effectiveness. Testing is an integral part of the teaching and learning process. Instructors need training on test construction in order to set tests that adhere to principles of test construction and be able to provide quality feedback to their students. Furthermore, instructors need to be well informed in order to make valuable decisions towards students' attainments. Its adequacy could bring injustice in judging students' performance. Again, the items may lack validity and reliability. Matuvo and Zubairi (2015) observed that academic members of staff in the Ugandan universities lacked appropriate assessment skills in evaluating their students. Babalola (2009) found that most state-owned universities were not adequately productive due to inadequate staff development programmes. Ibbad and Fawad (2015) claim that higher training to the staff leads to the better output of staff members. According to them (*ibid.*), training should be provided on a continuous basis after regular intervals.

5.0 Conclusion

Based on the research findings, the following conclusions are made:

- a. Most of the instructors in the sampled universities showed to construct tests by ensuring that their tests had content validity;
- b. It was observed from the study that instructors were writing the items in such a way that both high and low achievers could understand the items. However, from students' views, it was clearly shown that some questions had ambiguous statements, which could mislead them;
- c. Moreover, the findings showed that instructors were able to outline the content covered for the semester before setting tests from them and failed to prepare table of specification (blue print), which indicates content covered, instructional behaviour and distribution of the weights per domain as well as learning outcomes;
- d. Also, instructors minimally practiced the following variables: commenting on students' script to guide learning, provision of prompt feedback to students about their performance, preparation of test blueprint as a guide in the test construction, uses of Bloom's taxonomy in setting test items, submission of test items for moderation to the Head of Department and subject test items for item analysis ; and
- e. Furthermore, instructors faced various challenges including large class size, lack of training on test construction and lack of resources.

5.0 Recommendations

The study recommends the following:

- a. The number of enrolled students should be proportional with the number of instructors;
- b. Universities should employ additional instructors who could serve the purposes;
- c. The fund budgeted for buying teaching and learning materials should be increased; and
- d. There should be continuous professional development or seminars given in the universities and instructors should be encouraged to attend seminars/ workshops on test construction.

6.0 References

- Adodo, S.O. (2014). "An evaluation of secondary school teachers' competency in evaluating students' cognitive and psychomotor achievement in basic science and technology (BST)- Ondo State, Nigeria." *Journal of Emerging Trends in Educational Research and Policy Studies (JETERAPS)* 5(3): 48-532014 (ISSN: 2141-6990). Retrieved from <http://jeteraps.scholarlinkresearch.com/articles>.
- Agu, N., Onyekuba, C., & Anyichie, C. (2013). "Measuring teachers' competences in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory." *Academic Journals. Educational Research and Reviews* Vol. 8(8): 431-439. Retrieved from <http://www.academicjournals.org/ERR>.
- Ahmad, B., Zubairi, M., Ibrahim, B., Othman, J., Rahman, A., Rahman, A., Nor, M. (2014). *Assessment for learning practices and competences among Malaysian university lecturers*. University of Cumbria. Vol. 8 (1). Pages 14-31. Retrieved from <http://194.81.189.19/ojs/index.php/prhe/article/view>.
- Akiri, A., & Ugborugbo, M. (2009). "Teachers' Effectiveness and Students' Performance in Public Secondary Schools in Delta State, Nigeria." *Study Home Comm Sci*, 3(2):107-113.
- Anderson, W., & Krathwohl, R. (2001). *Taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman. Retrieved from <http://thesecondprinciple.com/teaching>.
- Binde, A. (2004). "Making sense of assessment: the same from outside different from inside." Proceedings of a workshop held from 28th March to 3rd April 2012 at the University of Dodoma. Teaching and learning improvement in higher education. Volume 1. Interpress. The University of Dodoma.
- Chung, R. R. (2008). "Beyond assessment: Performance assessments in teacher education." *Teacher Education Quarterly*, 35(1): 7-28
- Desforges, M. (2013). "Quality education in Tanzania: Issues and strategies." *In Mwenge University Journal of Academic Studies*. Vol.1. issue (1): 11-27. Mwenge University College of Education. Moshi.

- External Examiner. (2017). "Report on moderation of university examinations. Semester 1, 2016/2017." The University of Dodoma.
- Gichuhi, C. (2014). Teachers 'competence in tests construction within Blooms taxonomy for effective learning assessment: Published master thesis. Kikuyu-Kiambu County. Retrieved from <http://erepository.uonbi.ac.ke/bitstream/handle/11295/76048.pdf>
- Ibbad, A & Fawad, A. (2015). "Reasons for Low Performance of Teachers: A Study of Government Schools operating in Bahawalpur City, Pakistan." *International Journal of Academic Research in Progressive Education and Development*. Vol. 4, No. 2 ISSN: 2226-6348. <http://dx.doi.org/10.6007/IJARPED/v4-i2/1764>.
- Khan, B. (2012). "Relationship between assessment and students' learning." *International Journal of Social Sciences and Education*, 2(1): 576-588. Retrieved from: http://ecommons.aku.edu/pakistan_ied_pdcn/1.
- Kitta, S. (2017). TCU Quality Assurance Training Workshop, 10th February, 2017. Principles and Procedures of Test Construction. Tanzania Institute of Education (TIE): Dar es Salaam.
- Kurebwa, M. (2012). "Assessment problem in Zimbabwe's primary schools with special reference to Gweru district schools." (Published PhD Thesis). Zimbabwe Open University.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. (8thed.). New York: Macmillan.
- Mapolisa and Tshabalala (2013). "An evaluation of the impact of inadequate teaching and learning resources in public institutions of higher learning in Zimbabwe." *International Journal of Advanced Research*. Volume 1, Issue 10: 739-745. ISSN 2320-5407 <http://www.journalijar.com>.
- Masozzi, N. (2016). *Will new university entry qualifications improve quality graduate?* Retrieved from <http://m.ippmedia.com/en/features/will-new-university-entry-qualifications-improve-quality-graduates>.
- Matuvo, M., & Zubairi, A. M. (2015). "Assessment practices in the Developing World: Predictors of assessment practices in Ugandan Institutions of Higher Learning." *IJUM Journal of Educational Studies*, 3 (2): 75-112. ISSN 2289-8085. Retrieved from <http://journals.iium.edu.my/ijes/index.php/iejs/article/view/96>.

- Mbua, F. (2014). "The relationship between continuous assessment scores and performance in Certificate of Secondary Education (CSEE) Geography in Moshi district." (*Master's dissertation*). Mwenge Catholic University. Tanzania.
- Mhando, E.S. (2012). *Reflective teacher: Essays on education*. Morogoro: Elimu Reflective Networks.
- Mpapalika, K.M. (2013). "Tanzania science teachers' practices and challenges in continuous assessment" (*Master's Thesis, The University of Witwatersrand, Johannesburg*). Retrieved from <http://wiredspace.wits.ac.za>.
- Ndalichako, J. L. (2004). "Towards an understanding of assessment practices of primary school teachers in Tanzania." *Zimbabwe Journal of Education Research*, 16(3): 168-177. Retrieved from <http://dx.doi.org/10.4314/zjer.v16i3.26046>.
- Ndalichako, J. L (2015). "Secondary school teachers' perceptions of assessment in Tanzania." *International Journal of Information and Education Technology*, Vol. 5: 5, May 2015.
- Odili, J. N. (2014). *Assessing secondary school teachers' competences in continuous assessment skills in Delta State, Nigeria*. Vol. 4(1). Pages 46-55. US-China Education Review B, ISSN 2161-6248-David. Retrieved from www.davidpublishing.com/.../upfile/.../2014/2014012401355690.pdf
- Ogula, P. (2009). *Principles and practice of teaching and learning: A handbook for teachers and educators*. Port Victoria: New Kemit Publishers
- Ogula, P. (2009). *Principles and practice of teaching and learning: A handbook for Teachers and Educators*. Port Victoria: New Kemit Publishers.
- Ogula, P., & Rubeba, A. (2013). "Research findings on school effectiveness: implications for curriculum implementation in schools in East Africa." (Volume 1). In *Mwenge Journal of Academic Studies*. Mwenge University College of Education. Moshi.
- Omari, M. (2011a). *Concepts and techniques in educational assessments and evaluation*. (Volumes 1 & 2). Dar es Salaam: Oxford University Press.
- Omari, M. (2011b). *Concepts and Methods in Educational Research: 'A Practical Guide Based on experience.'* Dar es Salaam: Oxford University Press.

- Perumanathan, P.S. (2014). Formative assessment and feedback in the primary classroom: An interplay between teachers' beliefs and practices. Retrieved from <http://hdl.handle.net/10063/3312>
- Popham, J. (2008). Classroom assessment: What teachers need to know. USA: Pearson Education Inc.
- Sales, R. (2013). "Assessment practices in higher education: The experiences of newly appointed academics in professional fields from a phenomenological perspective." Published Thesis (Doctor in Health and Social Care), University of the West of England.
- Sanga, P.L. (2016). "Implications of teacher educators' practices in assessment for student learning in Tanzania." *Makerere Journal of Higher Education*, 8 (1) (2016) 3 - 24 ISSN: 1816-6822; 8 (1) (2016) 3 - 24 DOI: <http://dx.doi.org/10.4314/majoh.v8i1.1>.
- Tanzania Commission for Universities. (2014). Quality assurance general guideline and minimum standards for provision of University education in Tanzania. (2nd edition). Dar es Salaam.
- Tanzania Commission for Universities. (2019). Quality assurance general guideline and standards for provision of University education in Tanzania. (3rd edition). Dar es Salaam.
- Taras, M. (2007). "Assessment for learning: Understanding theory to improve practice." *Journal of Further and higher education*, 31(4): 363-371.
- UNESCO. (2014). UNESCO moving forward the 2030 Agenda for Sustainable Development. Retrieved from <https://en.unesco.org/creativity/sites/creativity.pdf>.
- UNESCO. (2012). *Education for All Global Monitoring Report 2012*. Retrieved from www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/pdf/gmr2012-report-ch1.pdf.