

# Test Construction Competence among Primary School Teachers in Dar es Salaam Region, Tanzania

Bernadetha B. Kapinga  
Institute of Adult Education  
Email: [benadetakapinga@yahoo.com](mailto:benadetakapinga@yahoo.com)

&

Anathe R. Kimaro  
Institute of Adult Education  
Email: [rabeltrandumi@gmail.com](mailto:rabeltrandumi@gmail.com)

## Abstract

Test construction is an essential part of teachers' responsibility. Therefore, teachers are supposed to craft well-functioning test items in ensuring effective teaching and learning. In this view, a teacher becomes a crucial and important part of the learning process. In due regard, this study sought to examine primary school teachers' test items construction practice competences in Ilala municipality in Dar es Salaam region, Tanzania. Specifically, the study sought to assess teachers' ability in test items construction and the extent to which teachers apply principles of test items construction in teaching and learning. The study used cross-section design with the sample of 72 primary school teachers. Findings from this study revealed that primary school teachers are weak in test construction. Even though some principles were done right, most of the critical issues related to test construction principles, test reliability and validity were overlooked. This questions validity of results, which would be awarded to pupils. The study recommended that although Tanzania, as a country, does not have a state-wide standard in testing, it is important for the Ministry of Education, Science and Technology (MoEST) to train teachers in assessment (especially, testing competence practices). Also, it is highly recommended that head teachers should take up the challenge of inviting instructors from teacher training institutions/universities to organise workshops for teachers on a regular basis so as to sharpen their skills on effective test construction practice.

**Keywords:** Assessment, action verbs, Blooms taxonomy, test construction, test items.

## 1.0 Introduction

Competency in test construction is an essential tool needed by every teacher if learning and instructional objectives are to be effectively attained. The importance of tests in the education system is enormous. Test provides a platform by which any significant education objectives can be achieved (Lal, 2005; Rudner, Schafer, 2002). Effectiveness of learning goals, entrenched in the curricula of a school, continues to be the most fundamental sign pole for institutional superiority, educational development and individual goals. Therefore, teachers are required to have adequate knowledge in achieving learning objectives in an accurate and precise manner. Thus, teachers must have the capability in science and arts in test construction (Ebinye, 2001).

A well-written test allows teachers to accurately and consistently measure students' mastery of specific contents taught in class. Likewise, poorly constructed test items can lead to inaccurate measurements of learning and provide false information regarding students' performance as well as instructional effectiveness (Baker, 2003; Ebinye, 2001; Dosumu, 2002). Such inaccurate measurements or errors occur in three categories: the first errors inherent in the instrument, errors in use of the instrument and errors emanating from responses of test takers (Kratwohl, 2002).

Test construction ability and quality are fundamental tools required by any teacher if teaching and learning goals are to be achieved. It cannot be possible without teachers themselves being competent in the arts and science of test construction (Baker, 2003; Chidolue, 1999; Rudner & Schafer, 2002). For example, Alele-Williams (2002) delineates that teachers should follow three major steps in classroom test construction: (i) planning the test, (ii) item writing and (iii) item analysis. Moreover, Izard (2005) cautioned that test constructors who plunge directly into item writing are likely to produce a lopsided test. It means that some areas will be over-represented while others may remain untouched. The test constructed without a blueprint is likely to be overloaded with impermanent and less important material.

In Tanzania, like its East African counterparts, teachers' assessment practice is characterized by use of poorly focused questions, predominance of questions that require short answers involving factual knowledge and evocation of responses that involve repetition rather than reflection plus lack of procedures designed to develop higher-order cognitive skills (Banda, 2011; Kahwa, 2009; Kimaro, 2019; Shemweleka, 2008).

Such phenomenon observably brings to the fore questions on whether or not primary school teachers have ability to construct test items and whether or not teachers follow test item construction principles.

### 1.1 Competencies in Test Construction

The quality of a test given by a teacher is closely linked with its ability to provide the kind of information needed regarding students' performances. A well written test allows the teacher to accurately and consistently measure students' mastery of specific contents taught in class. Results of such tests allow teachers to measure, to some degree, how effective their instruction has been. Conversely, poorly designed test items can lead to inaccurate measurements of learning and provide false information regarding students' performance as well as instructional effectiveness (Nitko, 2001). Any characteristics of a test item, which distracts the examinee from the major point or focus reduces effectiveness of that item (Frey, 2007). Any item answered correctly or incorrectly because of extraneous factors in the item results in misleading feedback to both examinee and examiner (Frey, 2007). Koksai (2004) outlined factors that are inherent in poorly designed tests, which if well-handled, will lead to quality classroom-based tests. They include the following:

- a. Non-specification of the target audience, about skill or area of ability the test intended to measure, extent of time allocated for each test item as well as about points test-takers would get for each correct response;
- b. Separate sections not clearly stated;
- c. Test items having more than one possible answer because they were not conceptualized;
- d. Not stating time allocated for each task on papers but only total time available to perform all tasks was given;
- e. Non-consideration students' level in test construction;
- f. Unclear instructions;
- g. Tasks students are expected to perform not being in concert with tasks they are told to do during classroom instructions;
- h. Lack of representativeness of items concerning what the teacher intended to test; and
- i. Collage-production of some test items (*ibid.*).

The teacher must have wondered what actually went wrong. Was it that she/he did not teach well or that pupils did not understand what was taught? The problem stems from neither the teaching nor the pupils' learning but from the way the test item was constructed. The tasks of for pupils required to perform in the test were not perfectly clear and thus, gave room for more than one possible correct response.

Therefore, to construct good test items, classroom teachers should possess competencies in the following areas as outlined by Chidolue (1999: 36):

“...determining the purpose of each testing exercise; stating specific, measurable educational objectives; making good content outline; preparing test plan which will guide item construction; choosing appropriate test item formats; constructing clear, precise and unambiguous items; constructing items that focus the attention of a group of students, often with widely varying background experiences, on a single idea; constructing items with appropriate difficulty and discriminative indices; developing marking guide suited for the test; performing item analysis of their test items; developing tests that are economical in time and money; giving clear directions on how the test should be administered and taken; reviewing the test in order to correct any errors made during item construction.”

Adding to the outline of competencies, Koksai (2004) noted that in test construction, it is essential that the teacher should ask the following questions: Is the task perfectly clear? Is there more than one possible correct answer? Can test-takers arrive at the correct response without having the skill supposedly being tested? Do test-takers have enough time to perform the task(s)?

## **1.2 Purpose of the Study**

Specifically, the study sought to assess ability of primary school teachers in test construction and whether or not teachers apply principles of test construction in selected primary schools.

## **2.0 Methodology**

This study used cross-sectional design and it was conducted in Ilala municipality, Dar es Salaam region. Proportionate stratified random sampling procedure was used to ensure that teachers from different public primary schools found in Ilala municipality were selected in the same proportion as they existed in the population. To do this, 50 percent of teachers in each of the three-sample primary schools in Ilala municipality were sampled (28 from the first school, 19 from the second school and 25 from the third school). The sample comprised 31 male and female teachers 41. Three data collection methods were used in this study. They comprised documentary review, interview and questionnaire. Documentary review primarily, focused on the teacher's ability to construct test items. In this endeavour, both primary and secondary documents were consulted. Consulted documents included teachers' weekly as well as monthly tests and term examination. Also, interview was used to solicit information from teachers on their ability to construct test item, their ability to use table of specification and Bloom's taxonomy. Likewise, questionnaires were administered to teachers for assessing their ability to construct test items and the extent to which teachers applied principles of test construction. The questionnaires comprised both close-ended and open-ended questions. For many questions in the questionnaires, the sub-program "frequencies" from the Statistical Package for Social Sciences (SPSS) was used in the process. Content analysis was employed to identify, analyze and interpret open-ended questions from questionnaire and interview. They were analyzed in relation to research questions of the study.

## **3.0 Research Findings**

### **3.1 Primary School Teachers' Ability in Test Items Construction**

To obtain relevant information on teachers' ability in test item construction, the questionnaire was administered to 72 primary school teachers. Table 1 presents summary of the study findings.

**Table 1: Teachers' views on their ability in crafting test items**

Statement	Teachers (n=72)					
	Always		Sometime		Never	
	f	%	F	%	F	%
I face difficulty to use Bloom's taxonomy to construct test item	49	68	17	23.6	5	6.9
I face challenges to prepare marking scheme two or more days after constructing the test	10	13.3	34	47.2	18	25
I face difficulties in constructing a table of specifications	64	88.8	2	2.7	6	8.2
I face difficulties in selecting questions from topics I think students have understood	2	2.7	4	5.5	66	91.6
I face challenges to match objectives from the syllabus with the test items	9	12.5	46	63.8	17	23.6
I face difficulties to check for clarity of crafted items	15	20.8	41	56.9	6	8.3
I face difficulties in constructing multiple choice item type	9	12.5	40	55.5	13	18
I always face difficulties in arrange questions into sections based on their nature or type	5	6.9	26	36.1	41	56.9
I face difficult in construct new test item in every test hence I rely on past questions	29	40.2	33	45.8	11	15.2
I face challenges in checking for the item difficulty and discrimination after the test	42	58.3	17	23.6	13	18
I fa challenges to compute for the item difficulty and index discrimination of the test scores.	49	68	07	9.7	16	22.2
I face challenges to craft more items than actually needed	28	38.8	39	54.1	5	6.9
I have difficult to craft good test item hence I have to use my textbook to write test items	43	59.7	25	34.7	4	5.5

Table 1 shows that most teachers face challenges in many areas of test construction. For example, teachers face difficulty in constructing a table of specifications, use Bloom's taxonomy to construct test items, compute for the item difficulty, index discrimination of the test scores as well as in crafting good test items such that they have to use their textbooks to write test items (Table 1). This shows that primary school teachers have inadequate skills in test construction.

Furthermore, the researcher used assessment criteria such as sequential order, layout of question papers, attractiveness as well as clarity of printing, clarity of language and teachers' application of test construction principles in order to triangulate questionnaire findings. Based on the aforesaid criteria, reviewed test/examination papers showed that most of the reviewed test/examination papers had incomplete and unclear instructions, some statements were very long, they had unclear instructions and some teachers extracted statements directly from textbooks. Such findings were supported by results from interviews as one experienced teacher confirmed that,

"It is difficult to construct multiple choice items. Thus, what I do is to buy ready-made tests from people providing private tuition services to primary school pupils in Kariakoo....to me it is easy to buy rather than construct them" (School D, 2019).

Another teacher who assumed to be experienced and required to construct weekly, monthly and term test items for his class of seven pupils said that,

"Because I have taught standard seven for more than 7 years now, there is no problem. I have a bank of questions in my office and others are at home... So I just select 50 items to make one test. I am doing this because I know they are good test items and I have been using them for some years now and my students are doing very good" (School A, 2019).

From the foregoing quote, it is obvious that such teachers did not have required test construction skills. Teachers had difficulties in construction of test items. Hence, they decided to copy past paper items without considering quality and relevance to their subject and competencies to be measured.

Furthermore, after careful evaluation of test instruments, it was evident that the test developer who was the subject teacher failed to sample adequately to cover all content areas listed in the scheme of work for the relevant term. Analysis from papers revealed that content of papers focused on a few of

the areas taught. It was evident that items on the instruments (tests) did not adequately sample the content taught. It implies that a pupil who attained 70 percent or higher cannot be addressed as having adequate mastery of the content taught since he/she was not assessed in all taught areas. Likewise, a pupil who obtained 20 percent or less cannot be referred to as lacking mastery of the taught content. This is because the higher scoring pupil might have specialized in sampled areas, while the lower scoring pupil did not. It is possible that the higher scoring pupil lacked mastery over the content areas, which assessment instruments did not cover. Results from such assessment instruments can only be interpreted in terms of identifying pupils' strength(s) and weakness(es) or how much pupils knew in the few content areas assessed. Thus, the instruments lacked some degree of content validity.

### 3.1.1 Application of principles of test construction by teachers

This study noticed that despite the situation that very few teachers were found to be more familiar with table of specification concept than it was expected. Its uses remained a challenge to many of them. Many teachers did not use table of specification in their test item construction. For, example, it was noted that out of 72 teachers who were given questionnaire to fill, only 7(9.8%) teachers said sometimes they used table of specification to prepare test item, whereas 64(88.9%) had never use it. Their responses ranged from, "... sometimes to never have used table of specification in construct test item since I have employed in this school." In such a case, it is apparent that there was a mismatch between what they claimed to know and what they were practicing in their daily routines.

**Table 3: Teachers' responses on using table of specification in test item construction**

	Frequency	Percent	Valid percent	Cumulative percent
Never	64	88.9	88.9	88.9
Sometimes	7	9.8	9.8	98.7
Always	1	1.3	1.3	100
Total	72	100		

Source: Field data, 2019

Again, it was noted that only 1(1.3%) primary school teacher who filled in the questionnaire was always able to use table of specification in writing test items. This was similarly reported in interview findings as one experienced science subject teacher confirmed that,

“I remember learning table of specification when I was doing my Grade A studies many years ago. I remember properly it was taught in subject called *upimaji na tathimini* ...Yes, yes....but since then, ten years now, frankly speaking I have never used ....I don't remember even how it looks like sir” (interviews with science subject teacher from school B, 06/02/2020).

During interviews, another teacher, but in Kiswahili subject, gave the reason that there was no need to bother to remember and she insisted that,

“I don't remember how to use it again but I just need a little orientation to remember and be able to use it. I remember it helps the teacher to know how many items he/she will include in the test or so but I am not very much sure. You know, in primary school, we don't use it. Thus, it is a reason many of us don't remember it. I think we need training for that aspect for sure” (Interviews with Kiswahili subject teacher from school C, 09/02/2020).

From the findings, it is obvious that primary school teachers do not use table of specification, a pattern, which leads to poorly crafted questions. Interview findings corroborate with findings from documentary analysis as most of the reviewed items only required pupils to just recall facts. Few of the items measured comprehension and application. Short answer items covered items on knowledge and comprehension. Thorough account of the assessment tasks suggests that almost all items were “recall” type of questions. This implies that pupils who engage in rote learning are those who will perform well and not necessarily those who have mastery over the material taught. Thus, the assessment instruments did not comprehensively assess different types of thinking skill.

A follow-up question was whether or not there was use of Bloom's taxonomy on test items construction. The study revealed that out of 72 teachers who responded to the question, only 23(31.9%) teachers said sometimes they used Bloom's taxonomy to prepare test items, whereas 45(62.5%) they never had used it (see Table 3).

**Table 3: Teachers' responses on using table of specification in test item construction**

	Frequency	Percent	Valid percent	Cumulative percent
Never	64	88.9	88.9	88.9
Sometimes	7	9.8	9.8	98.7
Always	1	1.3	1.3	100
Total	72	100		

Source: Field data (2019).

Furthermore, teachers were given a list of action verbs and they were asked to indicate types of action verbs they mostly preferred in their test construction. Table 4 presents summary of the study findings.

**Table 4: Teachers' responses on using the following action verbs in their test item construction**

	Action verbs	Teachers (n=72)					
		Always		Sometimes		Never	
		F	%	F	%	f	%
1	List, define, name, arrange, outline	67	93	5	6.9	0	0
2	Describe, explain, identify, classify	59	81.9	13	18	0	0
3	sketch, illustrate, prepare, demonstrate	32	44.4	27	37.5	13	18
4	Differentiate, examine, compare,	15	20.8	46	63.8	11	15.2
5	Evaluate, predict, defend	9	12.5	18	25	45	62.5
6	Compose, prepare, organize, create, design	7	9.7	34	47.2	31	43

Table 4 shows that majority of teachers always use list, define, name, arrange, outline, describe, explain, identify and classify action verbs while constructing tests for their students. Nevertheless, only few teachers said always they use action verbs such as evaluate, predict as well as defend and compose, prepare, organize, create or design, respectively, in test item construction (Table 4). It leads to suggest that teachers' test items concentrated more on the lower levels of thinking with high percentages on very often and often scales for verbs on the lower levels of cognitive abilities. On the other hand, rarely and very rarely scales got higher percentages for higher order cognitive abilities verbs.

**Table 5: Action verbs usage by primary school teachers**

Subject	Correct action verb used in line with Blooms taxonomy	Total number of items	Percent
Kiswahili	61	150	40.6
English	54	150	36
Geography	29	150	19.3
History	18	150	12
Mathematics	7	150	4.6
Science	21	150	14
Total	190	900	21.1

Table 5 reveals that most of the items constructed by teachers did not employ action verbs. None of the subject teachers employed action verbs on even half of the test items constructed (Table 5). Only in 190 items that action verbs were used out of the total 900 test items in the three subjects, accounted for only 21.1 percent (Table 5). The syllabus states objectives in action verbs but framing of the questions did not reflect achievement of instructional objectives.

#### 4.0. Discussion

The study findings revealed that primary school teachers faced difficulties in constructing table of specifications, use of Bloom's taxonomy to construct test items, compute for the item difficulty, index discrimination of test scores and craft good test item. Hence, they used their textbooks to write test items. This shows that primary school teachers have inadequate skills in test construction. This finding corroborates with other previous research findings (for example, Kita & Tilya, 2010; Tilya & Mafumiko, 2010; Timothy, 2011; Kimaro, 2019). For example, Shemweleka (2008) and Timothy (2011) anchored that there was a tendency for some teachers to claim to have enough competency in selecting appropriate assessment items, however, in practice, they turned out to be poor in selecting relevant assessment tools for each particular competence. Similarly, Njabili (1999); Nitko and Brookhart (2007); and Njabili and colleagues (2005) found out that most of the tests

used for continuous assessments and the end of the term examinations in the primary schools contained ambiguous as well as misleading questions. This was said to be among key reasons some pupils failed in tests.

After careful evaluation of test instruments, it was evident that the test developer who is the subject teacher failed to sample adequately to cover all content areas listed in the scheme of work for the relevant term. Analysis from papers revealed that content of papers focused on a few of the areas taught. It was evident that items on the instruments (tests) did not adequately sample the content taught. The results from such assessment instruments can only be interpreted in terms of identifying pupils' strength and the weakness or how much pupils knew in the few content areas assessed. Thus, instruments lacked some degree of content validity. An assessment task, which lacks content validity is likely not to reflect the important content, skills and learning outcomes specified in the school's or district's curriculum framework and content standards (Nitko, 2001). This is reflected in test papers where emphasis was placed on fewer content areas. This is because the test reflected the learning outcome of those few content areas taught.

Again, some of the reviewed test items measured trivial knowledge. Such questions demanded the lowest form of thinking such that the pupils who did not have any knowledge of the material could have answered the questions correctly. Nitko (2001) argues that most worthwhile learning involves students' using a combination of skills and content rather than using isolated skills or bits of content. This suggests that assessment instruments do not measure worthwhile learning, to some extent. It is evident that assessment instruments lacked, to some extent, content representativeness and relevance.

The thorough account of assessment tasks suggests that almost most of the items were "recall" type of questions. This implies that examinees who engage in rote learning are those who will perform well and not necessarily those who have mastery over the material taught. Thus, the assessment instruments did not comprehensively assess different types of thinking skills. For an assessment result to be valid, the tasks should assess a student's ability to use strategies and processes that reflect how scholars in the discipline think (Nitko, 2001). The examined assessment instruments deviated from Nitko's assertion. That is, only one lower-level cognitive process was greatly emphasized.

## 5.0 Conclusion and Recommendations

The evaluations of test items revealed that investigated primary school teachers are weak in test construction. Even though some principles were done right, most of the critical issues, which are related to test construction principles, test reliability and validity were overlooked. This questions the validity of results, which would be awarded to students. It is important for classroom teachers to be aware of the fact that measurement of psychological constructs like academic achievement is a difficult aspect to do. This is due to the complex and dynamic nature of human beings. However, there is need for teachers to gather some information about pupils for decision-making on curriculum, pupils and education policy. This information is needed not only for teachers but also for parents, schools' management and policymakers. Because the information collected is used for decision-making, it must be as accurate as possible. If tests with low validity and reliability are mostly used, then, inappropriate decisions are likely to be made.

The accuracy of classroom assessment results is very important but difficult to achieve. The complex nature of examinees, examination conditions, problems with test instruments and other factors reduces the validity of classroom assessment results. However, through careful planning of the test as well as adherence to principles in test construction, test assembling, test administration, scoring and result interpretation can help teachers to gather valid as well as reliable information about pupils. However, it appears that some teachers do not have much knowledge in testing practices. Although Tanzania, as a country, does not have a state-wide standard in testing, it is important for the Ministry of Education, Science and Technology (MoEST) to train teachers in assessment (especially, testing practices). Also, it is highly recommended that head teachers should take up the challenge of inviting instructors from teacher training institutions/universities to organise workshops for teachers on a regular basis so as to sharpen their skills on effective test construction practice.

## 6.0 References

- Alele-Williams G. (2002). "Measurement and evaluation in mathematics: The way forward." *Journal of Basic Science*. 1(1):1-7.
- Baker J.O. (2003). *Testing in modern classrooms*. London: George Allen and Urwin Ltd.
- Banda, S. (2011). "Application of constructivist approach in competence-based curriculum in secondary schools in Tanzania: The case of chemistry subject in Songea municipality." Unpublished dissertation. University of Dar es Salaam
- Bloom, B., M. Englehart, E. Furst, W. Hill & Krathwohl D, (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans.
- Chidolue M. (1999). *Introduction to tests and measurements*. Unpublished mimeograph on measurement and evaluation. Awka, Nnamdi Azikiwe University.
- Childs R.A. (1989). Constructing classroom achievement tests. ERIC Digest. ED 315426.
- Dosumu C. T. (2002). *Issues in teacher-made tests*. Ibadan: Olatunji and Sons Publishers.
- Ebinye P. O. (2001). "Problems of testing under the continuous assessment programme." *Journal of Quality Education*, 4(1): 12-19
- Frey B.B. (2007). "An introduction to quality test construction." *Journal of Education*, 2(12) 23-37.
- Izard J. (2005). *Overview of test construction*. Paris: UNESCO.
- Kahwa, J. (2009). "Identification of the gaps between intentions and practices in the implementation of the 2005 revised biology curriculum." Unpublished Master degree Dissertation. University of Dar es Salaam.
- Kimaro. A. R. (2019). "Influence of self-efficacy and competencies on continuous and comprehensive evaluation practices among primary school teachers." Dodoma University. Thesis.
- Koksal D. (2004). "Assessing teachers' testing skills in Elt and enhancing their professional development through distance learning on the net." *Turkish Journal of Distance Education* 5(1), 12-26.
- Krathwohl, D. R. (2002). "A Revision of Bloom's taxonomy: An Overview." *Theory into Practice*, 41 (4): 212-218. DOI:

- Lal J. P. (2006). *Educational measurement and evaluation*. New Delhi: Anmol Publications, Pvt, Ltd.
- Nitko, J. A. (2001). *Educational assessment of students*. New Jersey: Prentice Hall
- Rudner L. M. & Schafer WD. (2002). *What teachers need to know about assessment*. Washington, DC: National Education Association.
- Shemwelekwa, R. (2008). "The effectiveness of adoption of competence-based education for teaching." *Journal of Education and Practice*, 2(4), 121-138
- Silker R. T. (2003). *Teachers and tests*. London: Basil Blackwell.