International Journal of Evaluation and Research in Education (IJERE)
Vol. 14, No. 3, June 2025, pp. 2051~2060
ISSN: 2252-8822, DOI: 10.11591/ijere.v14i3.30825

Primary school teachers' competences in constructing quality multiple-choice questions in Tanzania

Florence Kyaruzi[1], Kingtin Kisamo[2]
[1]Department of Educational Psychology and Curriculum Studies, University of Dar es Salaam, Dar es Salaam, Tanzania
[2]National Examinations Council of Tanzania, Directorate of Examinations Administration, Dar es Salaam, Tanzania

Article Info

Article history:

Received Apr 10, 2024
Revised Oct 17, 2024
Accepted Oct 30, 2024

Keywords:

Multiple-choice questions
Primary school
Quality test
Teacher competence
Test construction

ABSTRACT

Multiple-choice questions (MCQs) are highly used in assessing students' learning outcomes due to their versatility in promoting test content validity. Nevertheless, teachers experience psychometric challenges in constructing quality MCQs. This study investigated primary school teachers' competence in constructing quality MCQs in Dar es Salaam Region, Tanzania. Specifically, we examined teachers' competence in constructing quality MCQs, analyzed the types of item-writing flaws (IWFs) found in teacher-made MCQs, and examined the challenges that teachers encounter in constructing quality MCQs. The study was guided the by classical test theory (CTT) and it employed a mixed-method research approach. In particular, data collection and analysis were conducted using a concurrent triangulation design. A total of 213 MCQs constructed by 20 teachers from ten primary schools were collected and analyzed. The results indicated that primary school teachers were less competent in constructing quality MCQs as 156 (73.2%) of the teacher-made MCQs assessed lower levels of Bloom's cognitive domain. Similarly, most of the MCQs had IWFs, with 274 (69.2%) having test wiseness flaws and 122 (30.8%) having irrelevant difficulty flaws. The findings call for capacity building to promote teachers' competence in constructing quality test items.

Corresponding Author:

Florence Kyaruzi
Department of Educational Psychology and Curriculum Studies, University of Dar es Salaam
Dar es Salaam, Tanzania
Email: sakyaruzi@gmail.com

1. INTRODUCTION

Multiple-choice questions (MCQs) are highly used to assess students' learning outcomes due to their effectiveness and efficiency if well-constructed [1]. MCQs are efficient, easy to administer and score, have high content validity-ability to assess a broad range of course materials in a relatively short time [2], and enable teachers to provide timely feedback. In addition, quality MCQs produce valid and reliable results, are free from marking subjectivity and allow educators to conduct item analysis to ascertain the quality of the test items [3], [4]. However, in order to be effective, MCQs need to comprise items that measure various levels of Bloom's taxonomy [5], [6]. Nevertheless literature shows that teachers experience challenges in constructing meaningful stems and plausible distractors that are free from item writing flaws (IWFs) [7]. Also, some studies show that MCQs are not perceived as good enough to reflect students' learning ability to use analytical thinking to identify the correct answer from the given alternatives [5], [8]. Although a lot of criticism has been mounted against the use of MCQs for a long time, Oc and Hassen [8] argue that when used in an online environment, they promote self-regulated learning among students. However, quality MCQs

require teachers who are conversant in preparing them. This is the central concern of this study which seeks to ascertain teachers' competence in constructing quality MCQs.

Various circulars and policies emphasize the importance of assessment competences among teachers, such as the ability to link assessment to instructional purposes by constructing quality assessment tools such as MCQs [9], [10]. Effective MCQs should conform to the predetermined guidelines that explain how to construct quality MCQs which can effectively measure the intended learning outcomes [3]. The most common indicators of quality MCQs, include: i) the ability to measure learning outcomes at various levels of cognition as per Bloom's taxonomy; ii) being error-free; iii) having plausible distractors and appropriate stems; and iv) containing a correct answer [11]. Teachers' mastery of the assessment principles of test construction is a prerequisite to construction of quality test items that limit guessing and promote self-regulated learning among students.

MCQs are commonly used to measure students' learning in Tanzanian primary schools. In the primary school leaving examination (PSLE) administered by the National Examinations Council of Tanzania (NECTA), a considerable proportion of questions are MCQs of the single-best-answer type with five homogeneous choices except for mathematics subject [12]. Nonetheless, some teachers and other educational stakeholders hold a misconception that MCQs are less useful in educational assessment [6]. Such a claim is attributed to examination malpractices such as guessing [13] and IWFs resulting from teachers' incompetence in constructing quality MCQs [14], [15]. According to Downing [16], inexperienced constructors of test items prepare poor-quality, faulty, and low cognitive-level test items. In this context, there is a growing need in Tanzania for an inquiry into primary school teachers' competences in constructing quality MCQs given that they constitute a greater proportion of the test items in thigh-stakes national examinations. An investigation into the quality of teacher-constructed MCQs is important to get a theoretical understanding of teachers' competence in constructing MCQs.

We employed the classical test theory (CTT) to conceptualize teachers' competences in creating quality MCQs. This theory calls for adherence to the guidelines for constructing quality test items. According to the theory, a genuine score can be correctly obtained if there are no measurement error [17]. Thus, the sum of the true score and error score on a psychological measure (test) equals the observed score. Therefore, the number of correctly answered items is taken into account when evaluating the examinees' true level of performance or competence [10]. So, teachers' proficiency in creating quality MCQs was assessed based on CTT by checking their MCQs for compliance with the standards of MCQs. As per CTT, the number of flawless MCQs that the teachers had created served as a measure of their competence in creating such questions [17]. MCQs constructed without adhering to the guidelines for constructing quality items were considered having IWFs, which have the potential to adversely affect the way students comprehend and respond to MCQs [18]. In most cases, poorly constructed MCQs might embody test wiseness IWFs and irrelevant difficulty IWFs. Test wiseness IWFs give clue to the correct answer while irrelevant difficulty IWFs prevent students from demonstrating their true ability [19].

Although MCQs are highly used in assessing students' learning outcomes in primary schools in Tanzania [1], [14], empirical studies that have analyzed the actual MCQs made by primary school teachers to ascertain their competences in constructing them are scarce. Most of the studies conducted in the country measure teachers' competence in constructing MCQs based on self-report measures [13]. Even those which were carried out elsewhere concentrated on comparing students' performance on MCQs and short-answer items to ascertain teachers' proficiency in creating high-quality MCQs [20], [21]. Also, many studies have examined the tests constructed by secondary school teachers [3], [22] but not by primary school teachers who comparatively use MCQs to a greater extent. Examining Tanzanian primary school teachers' competences in constructing quality MCQs by directly analyzing the quality of teacher-made MCQs is necessary considering the usefulness of MCQs in assessing student learning. This study investigated primary school teachers' competences in constructing quality MCQs. Specifically, this study was intended to answer three research questions:

i) To what extent are teachers competent in constructing quality multiple-choice test questions?
ii) What type of IWFs are found in teacher-made MCQs?
iii) What challenges do teachers encounter in constructing MCQs and how could they be addressed?

2. RESEARCH METHOD

2.1. Participants and research design

We employed a mixed-methods research approach to investigate primary school teachers' competences in constructing quality MCQs. Particularly, we used a concurrent triangulation research design. The mixed-methods research approach enabled researchers to collect rich data for a better understanding of the research problem as the methods complemented each other [23], [24]. Qualitative data were gathered by

conducting a documentary review to identify IWFs in the MCQs constructed by teachers while quantitative data were collected using a questionnaire. The use of mixed methods was in line with Wolf [24], who argues that competence is best measured using multiple test instruments. Furthermore, the concurrent triangulation research design was appropriate because it facilitated triangulation of methods to ensure corroboration and achieve an in-depth understanding of the phenomenon.

Data was collected from 10 primary schools in Dar es Salaam, Tanzania. Dar es Salaam was purposefully selected because it had emerged as the first-ranked region out of 26 regions in PSLE performance for five consecutive years (2017-2021) [25], which gave an impression that teachers in this administrative region are conversant with the construction of quality MCQs. Ten primary schools were selected using the stratified sampling technique based on their performance ranking in the 2021 PSLE national examinations including four highly-ranked, three moderately-ranked and four lowly-ranked schools. Data were purposefully collected from 20 (11% female) experienced teachers of science subjects. The teachers of science subjects were sampled because science has a number of concepts that can be suitably assessed using MCQs. The study was conducted after attaining research permits from the respective university, regional and school authorities. During data collection, teachers voluntarily participated in the study after being informed of the purpose of the study and having signed an active informed consent form. The sampled respondents were professional teachers whose qualifications ranged from certificate (40%), diploma (50%) to bachelor degree (10%) in education.

2.2. Instruments

Data were collected through a combination of methods including a workshop for constructing MCQs, an IWF checklist, and questionnaires. A one-day workshop was conducted at one teachers' resource center (TRC) in the Dar es Salaam Region to engage teachers in the construction of MCQs. A total of 213 MCQs were constructed by the teachers in that workshop whereby each teacher constructed at least 10 MCQs measuring specific levels of Bloom's taxonomy. The teachers were provided with all resources, including syllabuses, textbooks [10] and the PSLE examination format which consisted of a sample table of specifications [12]. Then they were asked to construct MCQs from various topics taught in grade VI and class VII. As supported by Ørngreen and Levinsen [26], the workshop enabled the researchers to get hands-on information about teachers' competences in constructing MCQs. The constructed MCQs formed an adequate sample because, according to the CTT, a sample of over 200 test items is ideal to guarantee the reliability and validity of teachers' competences in constructing MCQs [27]. After the workshop on constructing MCQs, a presentation on the principles of constructing quality MCQs was done to consolidate teachers' competences in constructing quality MCQs.

A total of 15 IWFs identified by Tarrant and Ware [28] based on the recommendations made by other researchers [29] served as the basis in designing the checklist used to define IWFs. Additionally, based on Bloom's taxonomy, the cognitive level of each MCQ was incorporated into the tool. To make the tool easier to refer to, the researchers added rows identifying the various forms of IWFs. Based on the results of the pilot research that was conducted in one school, it was determined that the instrument was valid for use in data collection because it had a high level of inter-rater reliability and an overall agreement of 87%. Table 1 indicates the sample checklist for evaluating IWFs in teacher-constructed test MCQs. Besides, a questionnaire was used to collect data about teachers' competences in constructing quality MCQs. This questionnaire consisted of 26 items clustered into two constructs for measuring competence in constructing quality MCQs concerning competences for constructing the stem (Cronbach's alpha=.71), and competence for preparing the items' alternatives (Cronbach's alpha=.70). The questionnaire items were measured on a five-point Likert scale: strongly disagree (1), disagree (2), partially agree (3), agree (4), and strongly agree (5). The questionnaire enabled the researchers to determine whether the findings from teachers' appraisal of MCQs reflected their actual competences in constructing MCQs. Teachers responded to the questionnaire after handing over the MCQs that they had constructed to avoid data contamination as the questionnaire had clues to the guidelines for constructing quality MCQs.

2.3. Data analysis

Content analysis procedures were employed in analyzing the MCQs constructed by teachers. We opted for the content analysis technique since it is very useful in checking the qualities of an assessment tool (i.e., compliance with MCQs construction criteria) [2]. Four specialists in educational evaluation were engaged in analyzing the MCQs that the teachers had created. To achieve reliability of the analysis, the specialists assessed the kinds of IWFs present in the teacher-made materials [30] independently but in pairs by referring to the rules of creating quality MCQs. The experts achieved analysis reliability by calculating the agreement percentage of the total number of coded MCQs [30]. The obtained percentage of agreement was 93%, which is greater based on the conventions threshold 80% [31].

The degree to which the items addressed different levels of Bloom's cognitive domain and followed the rules of creating MCQs was evaluated. A tally sheet was utilized to ascertain the frequency of IWFs in teacher-constructed MCQs using the chosen category to produce quantitative data. Then, an IWF checklist bearing the rules of writing MCQs and Bloom's cognitive domain levels was used to analyze the data. In order to identify relevant variables, some quantitative data were analyzed in terms of descriptive statistics with the help of in SPSS version 25.

Table 1. Checklist on IWFs in teacher-made MCQs

Question code	Frequency and types of IWFs (coded 1-10)										Total IWFs per MCQ
	1	2	3	4	5	6	7	8	9	10	
1	1	0	0	0	1	0	0	1	0	0	3
2	0	0	0	0	1	0	0	1	0	1	3
3	0	0	0	0	1	0	0	0	1	0	2
4	0	0	0	0	1	0	0	0	0	0	1
5	0	0	0	1	1	0	0	1	0	0	3
.	.	.	.	.	.	.	.	.	.	.	.
213	1	0	0	0	0	0	0	0	0	0	1
Total	104	11	7	22	124	3	2	102	12	10	398

Note. Key for interpreting types of IWFs:
1. Lost sequence in data – Options not arranged in chronological or numerical order
2. Grammatical clues – Using words that give hints on the correct response
3. Logical clues – The stem and alternatives are arranged in predictable patterns
4. Greater details incorrect option – The correct answer is longer with more details
5. Implausible distractors – Some distractors have implausible content
6. Use of absolute terms – Use of extreme terms such as always, never, only
7. Use of 'all of the above' – MCQs consist of all of the above as the alternatives
8. Poor stem structure – The stem does not represent a clear problem independent of the alternatives
9. More than one correct answer – MCQs with more than one correct answer
10. Use of negative stem – Alternatives use not true or except as an option

3. RESULTS AND DISCUSSION

3.1. Teachers' competences in constructing quality MCQs items

The first research question examined the extent to which teachers were competent in constructing quality multiple-choice items. Consistent with what Creswell [32] recommends, several methods were used to assess teachers' competences in constructing quality MCQs such as self-reports, and content analysis of teacher-constructed MCQs. Data for answering this question were gathered using questionnaires, in which teachers reflected on their strengths in constructing MCQs and identified areas in which they needed to improve. Table 2 presents the findings on teachers' competences in constructing quality MCQs indicating the mean, standard deviations (SD), and frequency for each item.

Table 2. Teachers' competences in constructing quality MCQs-stem and alternatives

Statements	Mean	SD.	Min.	Max.
When constructing the stem of a multiple-choice item, I:				
1) Place the main idea in the stem	4.10	0.85	1	5
2) State the stem in simple and clear language	4.10	1.07	1	5
3) Write the stem with clear directions	3.90	1.12	1	5
4) Apply specific Bloom's verbs	4.05	0.78	1	5
5) Present a single clear objective in the stem	4.20	0.83	1	5
6) State the stem in a positive form whenever possible	4.25	0.85	1	5
7) Emphasize on bold or Italics whenever negative wording is used	3.15	1.53	1	5
When constructing alternatives of multiple-choice items, I:				
8) Make all distractors plausible	3.80	1.15	1	5
9) Present alternatives in similar patterns	2.65	1.27	1	5
10) Make the alternatives grammatically consistent with the stem	3.15	1.14	1	5
11) Avoid using all of the above and none of the above	3.75	1.29	1	5
12) Vary the position of the correct answer randomly	3.60	1.54	1	5
13) Avoid the use of specific determiners such as never, always, and only	3.75	0.91	1	5
14) Make alternatives to have only one correct alternative	4.45	1.05	1	5

Note. SD.=standard deviation, Min.=minimum, Max.=maximum

The findings in Table 2 indicate that most of the teachers were conversant with the principles of constructing MCQ stem, especially stating the stem in a positive form, stating the stem in simple and clear language, applying specific verbs as per Bloom's taxonomy, and placing the main idea in the stem. However,

half (50%) of them rarely bolded or italicized the negatively phrased words whenever they used them. Further, the findings in Table 2 indicate that most teachers were conversant with the principles of constructing quality alternatives. Specifically, the teachers evaluated themselves as conversant with including only one correct answer in the alternatives, avoiding the use of specific determiners, and constructing plausible distractors. Nevertheless, the findings indicated that teachers struggled with making the alternatives grammatically consistent with the stem and presenting alternatives in a similar pattern. Generally, the results imply that most of the teachers evaluated themselves as competent in constructing MCQ stems and alternatives.

The data obtained from teachers' self-reports was supplemented with data collected through documentary analysis of teacher-constructed MCQs to ascertain teachers' competences in constructing quality MCQs. The 213 teacher-constructed MCQs measured various levels of Bloom's taxonomy, namely 77 (36.1%) remembering, 79 (37.1%) understanding, 29 (12.2%) applying, 21 (9.9%) analyzing, nine (4.2%) evaluating, and one (0.5%) creating. Analysis of the findings showed that 156 (73.2%) MCQs measured remembering and understanding levels of Bloom's cognitive domain. The remaining items were developed at moderate (22.1%) levels of applying and analyzing while higher thinking skills of evaluating and creating were assessed by less than 5% of the items.

Table 3 presents a sample MCQ on each level of Bloom's taxonomy. The findings indicate that most of the teacher-constructed MCQs assessed lower thinking skills, which might imply that the teachers were less competent in constructing MCQs that assess higher-order learning outcomes. While the current curriculum emphasizes competency-based assessments that measure higher-order learning outcomes, this was not reflected in the teacher-constructed MCQs as most of the developed items assessed lower-order skills. Brown and Abdulnabi [11] insist that test items that are well-designed to assess higher levels of cognition are indicators of teachers' competence in test construction. Furthermore, the findings concur with previous studies [2], [18] that showed that most of the teacher-made tests and examinations assessed lower levels of Bloom's cognitive domain. In particular, the findings are consistent with Kyaruzi [2] who found that secondary school Mathematics teachers constructed tests that mostly assessed remembering (37.4%), understanding (27.4%), and applying (26.4%). The presence of lower-level MCQs signals a lack of competence in constructing quality items among teachers [1], [11], which calls for capacity building on assessing higher-order learning outcomes.

Table 3. Sample MCQs measuring various levels of Bloom's taxonomy

S/N	Level of Bloom's taxonomy	Sample MCQs
1	Remembering (question 16)	What type of safety sign is used in unauthorized access areas such as in the laboratory? A) Black B) Green C) Red D) Yellow
2	Understanding (question 38)	Why are carbon dioxide and hydrogen sulphide gases removed during the production of electricity from biogas? A) To purity methane and prevent corrosion B) To increase the energy from the biogas C) To increase the efficiency of biogas D) To have the property of biogas as fuel
3	Applying (question 39)	Find the mechanical advantage efficiency of a system consisting of five pulleys if an effort of 12 kilograms is used to lift 48 kilograms of load? A) 80% B) 90% C) 60% D) 40%
4	Analyzing (question 120)	In an experiment on osmosis water travels from a weak solution to a heavy solution. When we boil a potato and find that some part of it was not boiled, what causes it? A) The cells were more stable B) The water froze C) The potato cells were dead D) Water flows more via the peels
5	Evaluating (question 95)	Which of the following groups, if they do not do physical exercise, have a higher chance of getting diseases such as high blood pressure, cancer, diabetes, and bone diseases? A) Elderly people B) Children C) Pregnant women D) People who do light work
6	Creating (question 96)	The respiratory system is made up of the nose, pharynx, trachea, bronchi, lungs, and air sacs. What is the nerve system made of? A) Gallbladder and pancreas B) Brain, motor, and portal nerves C) Kidney, liver, and skin D) Motor nerves and sensory nerves

3.2. IWFs embodied in teacher-made MCQs

The second research question examined the types of IWFs found in the teacher-made MCQs. Data for answering this question were obtained from the document analysis of 213 teacher-constructed MCQs. The findings indicated that teacher-constructed MCQs had a plenty of test wiseness IFWs 274 (69.2%) and irrelevant difficulty IWFs 122 (30.8%). Table 4 summarizes the IWFs that were found in the 213 teacher-made MCQs.

The analysis of IWFs in Table 4 indicates that most items were flawed, where there were 274 (69.2%) IWFs related to test wiseness and 122 (30.8%) related to irrelevant difficulty. Among the MCQs having test wiseness IWFs, there were 124 (31%) implausible distractors and 106 (27%) alternatives not arranged in a logical order. Moreover, the MCQs with IWFs related to irrelevant difficulty included 100 (25%) with poor stem structure particularly having alternatives that do not relate with the stem.

Cumulatively, out of the 213 teacher-constructed MCQs, 197 (91.1%) were ill-constructed, with one up to five IWFs. Findings in Table 4 indicate a high rate of IWFs, which implies that the teachers were less competent in constructing quality MCQs. The findings on teachers' competence in constructing quality multiple-choice stems and alternatives coincide with those of [17], [27], whose studies revealed relatively similar findings. Tariq *et al.* [18] evaluated the cognitive levels of MCQs and item writing with third-year medical students and found that 87 (65.9%) of the MCQs had IWFs. Generally, the presence of IWFs in the teacher-constructed MCQs suggests that the teachers were less competent in constructing quality MCQs consistent with the claim that teachers who are competent in MCQ item writing can prepare high-quality MCQs [16]. Teachers' inability to construct quality MCQs, among other reasons, could be attributed to a lack of formal training on educational assessment, particularly on the principles of MCQs item writing [10]. It follows that a lack of competence in constructing plausible distractors among teachers lowers the quality of MCQs, including their discrimination power [7].

The findings indicated that teachers hardly prepared alternatives that were in concordance with the stem and alternatives that were grammatically correct. Most of the teacher-constructed MCQs contained IWFs related to test wiseness, which provides students with clue to the right response and may result in false-positive results by favoring underprepared (but test-wise) students. In contrast, the questionnaire results suggest that teachers were less conversant with organizing alternatives in a similar pattern, ensuring the alternatives are grammatically consistent with the stem, and emphasizing negative words by bolding or italicizing them. These results contradict the Kissi [22] study in secondary school, which revealed that the majority of instructors had a commensurate degree of proficiency in creating MCQs in terms of quality stems and alternatives. The findings underscore a pressing need to acquaint teachers with the principles of developing quality MCQs that appropriately measure learning outcomes at various levels of Bloom's taxonomy.

Table 4. Types of IWFs in teacher-constructed MCQs

	IWFs	Frequency (%)
IWFs related to test wiseness (N=274)		
1	Lost sequence in data (alternatives not in chronological or numerical order)	106 (27%)
2	Grammatical clues (using a word that gives a hint on the correct response)	10 (3%)
3	Logical clues (answers are arranged in a predictable pattern)	6 (2%)
4	Greater details in the correct option	23 (6%)
5	Implausible distractors	124 (31%)
6	Use of absolute terms (always, never, only)	3 (1%)
7	Use of "all of the above", "none of the above" as the alternative	2 (1%)
IWFs related to irrelevant difficulty (N=122)		
8	Poor stem structure (alternatives not in the stem or ambiguity in the stem)	100 (25%)
9	More than one correct answer or no answer at all	12 (3%)
10	Use of negative stem (not true, except)	10 (3%)
	Total IWFs*	396 (100%)

Note: Some MCQs had more than one IWFs

3.3. Challenges encountered by teachers in constructing quality MCQs

The third research question examined the challenges that teachers encounter in constructing MCQs. Data for this question were collected by using a questionnaire in which the teachers rated the challenges they faced in constructing quality MCQs. Specifically, teachers were presented with a checklist and asked to identify the potential challenges they encountered in constructing quality MCQs as summarized in Table 5.

Table 5. Challenges encountered by teachers in constructing quality MCQs

	Challenge(s) encountered	Frequencies (%)
1.	Adding enough test items to cover all the requisite levels of the cognitive domain	20 (100%)
2.	Conducting item analysis to determine their MCQ characteristics	17 (85%)
3.	Creating plausible distractors in MCQs	17 (85%)
4.	Designing MCQs according to different difficulty levels	13 (65%)
5.	Writing MCQs so that both high and low achievers can understand	12 (60%)
6.	Avoiding the use of clues in MCQs	10 (50%)
7.	Adding sufficient MCQs to cover the appropriate instructional units	9 (45%)
8.	Designing questions that are aligned with the subject objectives and activities	7 (35%)

Table 5 presents findings on the challenges that teachers faced in constructing quality MCQs. Most teachers had difficulty constructing MCQs that measure various levels of cognition 20 (100%), conducting item analysis to determine the characteristics of the items 17 (85%) and constructing plausible distractors 17 (85%). According to the findings, the quality of the prepared MCQs was not up to the expected standards and the majority of items measured lower levels of cognition. Nonetheless, the findings presented in Table 5 provide insights that somewhat contradict the findings from self-reports using a close-ended questionnaire. Thus, while teachers reported in the questionnaire that they were competent in constructing quality MCQs, results from a content analysis of teacher-constructed MCQs showed that teachers experienced challenges in constructing quality MCQs and most items were flawed. The presence of a high rate of IWFs in teacher-constructed MCQs implies that the teachers had limited competence in constructing quality MCQs. Furthermore, the responses to the open-ended items of the questionnaire revealed that nine teachers (45%) had positive perceptions of MCQs but admitted that constructing quality MCQs was a challenging task. Besides, 11 (55%) participants thought that MCQs are ineffective in assessing students' learning outcome. The following excerpt represents the views of some teachers regarding MCQs:

> "*MCQs do not prove to be an effective method of measurement; you won't be able to test many of the skills that are required by using MCQs. They have a limited scope for assessment.*" (Participant 2)
> "*Also, if students do not know anything about the answer, they can pick one alternative that turns out to be correct, I do not think that is truly deserved.*" (Participants 8)

The findings indicate that some teachers view MCQs as ineffective in assessing students' learning outcomes. This aligns with Ludewig *et al.* [7], who found that some educators hold a misconception that MCQs are not useful and it is challenging to construct MCQs that measure higher levels of cognition. However, scholars such as Kanji and Dhamani [5] emphasize that well-constructed MCQs are capable of testing both lower and higher-level thinking skills and promoting self-regulated learning [8]. The findings call for capacity building to improve teachers' competences in constructing quality MCQs. This aligns with the views of Brown and Abdulnabi [11], who maintain that meaningful learning occurs when educators are skilled in the craft and science of creating excellent evaluation instruments.

The findings indicate that nine (45%) teachers experienced challenges in organizing MCQ stems and formatting them clearly to allow students to demonstrate their ability. Similarly, it was noted that some teachers experienced challenges in constructing MCQs with content validity demonstrating how well the test measures students' achievement of several instructional objectives. Regarding the effectiveness of MCQs in assessing students, 12 (67%) teachers stated that MCQs have a limited scope of assessment. In general, they argued that one cannot test many skills by using MCQs. Unlike the commonly reported advantage of MCQs in attaining content validity of the test, teachers expressed a negative view about the use of MCQs in assessing students. Such a perception might be the reason behind the faults observed in teacher-constructed MCQs and teachers' lack of awareness of the potential of attaining content validity of the MCQs by covering a wider part of the course in a single MCQ test [1], [2]. To address the identified challenges, teachers provided different views concerning how to improve their competences in constructing quality MCQs. In particular, 13 (67%) teachers commented that they should be capacitated through seminars and workshops. They also proposed that the course in educational measurement and evaluation should be compulsory for every teacher trainee in all the teacher training institutions. It is argued that teachers' perception that MCQs are not effective in assessing students' learning outcomes is rooted in the challenges they encounter in constructing quality MCQs. Various scholars [6], [11] informed that only a few teachers have formal training in constructing test items which limits their capacity to prepare quality MCQs.

Given that teacher-constructed MCQs were characterized by IWFs and assessed lower levels of the Blooms' taxonomy, they were deemed to be poorly constructed according to CTT. Generally, their measurement errors (such as IWFs) could adversely affect the quality of tests and subsequent test-based inferences. The embodiment of so many IWFs in teacher-constructed MCQs might compromise assessment of learners' achievement and subsequent decisions on teaching and learning. So, there is an urgent need for intervention to improve teachers' competence in constructing quality MCQs. It is noted that constructing quality test items is possible only if teachers have the competence to construct them [4]. Part of the intervention could be to provide teachers with clear guidelines on the principles of constructing quality MCQs [15]. Consistent with CTT [19], the findings imply a call for quality assurance mechanisms to promote school-based assessment tasks.

4. CONCLUSION

This study investigated primary school teachers' competences in constructing quality MCQs. Quite many IWFs were found in the teacher-constructed MCQs. The IWFs were of different kinds, including inconsistencies between alternatives and their stems, implausible distractors, and illogical arrangement of alternatives. These results highlight particular areas on which potential interventions can focus in the attempt to improve teachers' competences in creating quality MCQs. The findings also indicated that most of the teacher-constructed MCQs measured lower levels of Bloom's cognitive domain. This an attestation that most of the teachers were incompetent to construct quality MCQs that assess higher-order thinking as per Bloom's levels of cognition.

To that end, the findings suggest that there is a great deal of variability in the competences of teachers in writing MCQs. Enhancing teachers' ability to create high-quality MCQs is supposed to be a top priority because of the significant role that MCQs have in the high-stakes assessments primary school students in Tanzania. The results indicate that in-service training on creating high-quality MCQs is desperately needed to empower teachers. The training can capitalize on the IWFs that were observed. As this study examined teachers' competences in constructing quality MCQs by using questionnaires and content analysis of teacher-constructed MCQs, we recommend a study on the association between teachers' demographic variables and the quality of test items they construct. Further studies can focus on the effectiveness of in-service training in promoting teachers' competence in constructing MCQs and other assessment tools.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Florence Kyaruzi	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
Kingtin Kisamo	✓	✓	✓	✓		✓	✓	✓	✓	✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the University of Dar es Salaam review board.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author [FK]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] J. Jovanovska, "Designing effective multiple-choice questions for assessing learning outcomes," *Infotheca*, vol. 18, no. 1, pp. 25–42, 2018, doi: 10.18485/infotheca.2018.18.1.2.

[2] F. Kyaruzi, "Effectiveness of mathematics continuous assessment feedback in secondary schools in Tanzania," *Journal of Education, Humanities and Sciences*, vol. 1, no. 2, pp. 51–63, 2012.

[3] F. Tangianu *et al.*, "Are multiple-choice questions a good tool for the assessment of clinical competence in internal medicine?" *Italian Journal of Medicine*, vol. 12, no. 2, pp. 88–96, 2018, doi: 10.4081/itjm.2018.980.

[4] R. M. Furr, *Psychometrics: an introduction*. Thousand Oaks, CA: SAGE Publications, Inc, 2021.

[5] Z. Kanji and K. A. Dhamani, "9-assessing the cognitive domain through MCQs: critical to quality assurance in higher education," *Journal of Higher Education in Africa*, vol. 15, no. 1, pp. 135–142, 2022, doi: 10.57054/jhea.v15i1.1496.

[6] S. Kitta and F. Tilya, "Assessment status of learner-centred learning in Tanzania in the context of the competence-based curriculum," *Papers in Education and Development*, no. 29, pp. 1–14, 2010.

[7] U. Ludewig, J. Schwerter, and N. McElvany, "The features of plausible but incorrect options: distractor plausibility in synonym-based vocabulary tests," *Journal of Psychoeducational Assessment*, vol. 41, no. 7, 2023, doi: 10.1177/07342829231167892.

[8] Y. Oc and H. Hassen, "Comparing the Effectiveness of Multiple-Answer and Single-Answer Multiple-Choice Questions in Assessing Student Learning," *Marketing Education Review*, vol. 35, no. 2, pp. 44–57, 2025, doi: 0.1080/10528008.2024.2417106.

[9] Ministry of Education and Vocational Training Tanzania, "Education and training policy," (in Swahili), 2014. Accessed: Sep. 6, 2024. [Online.] Available: https://www.moe.go.tz/sw/nyaraka/sera-ya-elimu-na-mafunzo-mwaka-2014-toleo-la-2023

[10] Ministry of Education, Science and Technology Tanzania, "National curriculum framework for basic and teacher education," (in Swahili), 2019. Accessed: Jul. 12, 2024. [Online.] Available: https://www.tie.go.tz/publications/4

[11] G. T. L. Brown and H. H. A. Abdulnabi, "Evaluating the quality of higher education instructor-constructed multiple-choice tests: impact on student grades," *Frontiers in Education*, 2017, doi: 10.3389/feduc.2017.00024.

[12] The National Examinations Council of Tanzania (NECTA), "Examination format for primary educations," 2024. Accessed: Jan. 12, 2024. [Online.] Available: https://www.necta.go.tz/publications/all

[13] M. Y. Mazana, C. S. Montero, and R. O. Casmir, "Assessing students' performance in mathematics in Tanzania: the teacher's perspective," *International Electronic Journal of Mathematics Education*, vol. 15, no. 3, 2020, doi: 10.29333/iejme/7994.

[14] P. O. Ajayi and V. O. Idusogie, "Analysis of item writing flaws (IWFs) evident in objective formats examination questions in federal college of education (technical) Asaba, Nigeria," *Journal of Educational Research in Developing Areas*, vol. 1, no. 3, 2021, doi: 10.47434/jereda.1.3.2020.181.

[15] P. Przymuszała, K. Piotrowska, D. Lipski, R. Marciniak, and M. Cerbin-Koczorowska, "Guidelines on writing multiple choice questions: a well-received and effective faculty development intervention," *SAGE Open*, vol. 10, no. 3, p.2158244020947432 2020, doi: 10.1177/2158244020947432.

[16] S. M. Downing, "The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education," *Advances in Health Sciences Education*, vol. 10, no. 2, pp. 133–143, 2005, doi: 10.1007/s10459-004-4019-5.

[17] M. J. Allen and W. M. Yen, *Introduction to measurement theory*. Long Grove, IL: Waveland Press, 2001.

[18] S. Tariq, S. Tariq, S. Maqsood, S. Jawed, and M. Baig, "Evaluation of cognitive levels and item writing flaws in medical pharmacology internal assessment examinations," *Pakistan Journal of Medical Sciences*, vol. 33, no. 4, pp. 866–870, 2017, doi: 10.12669/pjms.334.12887.

[19] H. Pham, J. Besanko, and P. Devitt, "Examining the impact of specific types of item-writing flaws on student performance and psychometric properties of the multiple choice question," *MedEdPublish*, vol. 7, 2018, doi: 10.15694/mep.2018.0000225.1.

[20] B. Mehta, B. Bhandari, P. Sharma, and R. Kaur, "Short answer open-ended versus multiple-choice questions: a comparison of objectivity," *Annals of the National Academy of Medical Sciences (India)*, vol. 52, pp. 173–182, 2020, doi: 10.1055/s-0040-1712619.

[21] F. Farooqui, N. Saeed, S. Aaraj, M. A. Sami, and M. Amir, "A comparison between written assessment methods: multiple-choice and short answer questions in end-of-clerkship examinations for final year medical students," *Cureus*, vol. 10, no. 12, p. e3773, 2018, doi: 10.7759/cureus.3773.

[22] P. Kissi, "Multiple-choice construction competencies and items quality: evidence from selected senior high school subject teachers in Kwahu-South District," Ph.D. dissertation, University of Cape Coast, Cape Coast, Ghana, 2020.

[23] R. B. Johnson and A. J. Onwuegbuzie, "Toward a definition of mixed methods research," *Journal of Mixed Methods Research*, vol. 1, no. 2, pp. 112–133, 2007, doi: 10.1177/1558689806298224.

[24] A. Wolf, "Competence based assessment. Does it shift the demarcation lines?" in *Key Qualifications in Work and Education*, W. Nijhof and J. Streumer, Eds., Dordrecht: Springer Netherlands, 1998, pp. 207–220, doi: 10.1007/978-94-011-5204-4_13.

[25] The National Examinations Council of Tanzania (NECTA), "Primary school leaving examination school ranking," Dar es Salaam, 2021. Accessed: Jun. 18, 2024. [Online.] Available: https://onlinesys.necta.go.tz/results/2022/psle/results/reg_02.htm

[26] R. Ørngreen and K. Levinsen, "Workshops as a research methodology," *Electronic Journal of e-Learning*, vol. 15, no. 1, pp. 70–81, 2017.

[27] R. K. Hambleton and R. W. Jones, "An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development items: instructional topics in educational measurement," *Educational Measurement: Issues and Practice*, vol. 12, no. 3, pp. 38–47, 1993.

[28] M. Tarrant and J. Ware, "Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments," *Medical Education*, vol. 42, no. 2, pp. 198–206, 2008, doi: 10.1111/j.1365-2923.2007.02957.x.

[29] J. Breakall, C. Randles, and R. Tasker, "Development and use of a multiple-choice item writing flaws evaluation instrument in the context of general chemistry," *Chemistry Education Research and Practice*, vol. 20, no. 2, pp. 369–382, 2019, doi: 10.1039/c8rp00262b.

[30] M. D. White and E. E. Marsh, "Content analysis: a flexible methodology," *Library Trends*, vol. 55, no. 1, pp. 22–45, 2006, doi: 10.1353/lib.2006.0053.

[31] L. A. Gottschalk, M. K. Stein, and D. H. Shapiro, "The application of computerized content analysis of speech to the diagnostic process in a psychiatric outpatient clinic," *Journal of Clinical Psychology*, vol. 53, no. 5, pp. 427–441, 1997, doi: 10.1002/(SICI)1097-4679(199708)53:5<427::AID-JCLP3>3.0.CO;2-K.

[32] J. W. Creswell and J. D. Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: SAGE Publications, Inc, 2017.

BIOGRAPHIES OF AUTHORS

Florence Kyaruzi is the senior lecturer in Psychometry and Mathematics Education, Department of Educational Psychology and Curriculum Studies at the University of Dar es Salaam. He holds a Ph.D. in the Learning Sciences from the Lüdwig-Maximilians-University of Munich (LMU), Germany. His research expertise includes student and teacher conceptions of assessment, mathematics education, gender issues in education, teacher education, feedback, and mentorship. He can be contacted at email: sakyaruzi@gmail.com; florence.kyaruzi@duce.ac.tz.

Kingtin Kisamo is the principal examination officer in the Directorate of Examinations Administration, National Examinations Council of Tanzania (NECTA). His research areas include assessment and evaluation, test construction, and science education. Drawing from his practical experience, Kisamo has published several articles on the assessment and evaluation of student learning at basic education levels. He can be contacted at email: kingtink@gmail.com.