JRIIE Journal of Research Innovation and Implications in Education
Center for Research Implications & Practice

Website:www.jriiejournal.com ISSN 2520-7504 (Online) Vol.9, Iss.2, 2025 (pp. 950 - 963)

Challenges in Constructing Multiple-Choice (MC) Matching Item (MI) and True-False (TF) Test Formats among Educators Who Are Not Teachers by Profession in Tanzanian Higher Learning Institutions

Yusuph Maulid Kambuga,
College of Business Education, Tanzania
Email: Kambuga2008@yahoo.com

Abstract: *A test plays a significant role in education since it determines the success of instruction and sorts students' grades. However, constructing a high-order quality test for educators who are not teacher professionals is not simple. The research employs a descriptive design, utilising an online questionnaire distributed to 31 educators without formal teaching training who managed to respond from learning institutions in Dodoma City. The study's objectives include assessing the impact of the lack of formal training on the quality and effectiveness of test items, examining the alignment of test questions with learning objectives, and exploring the factors influencing the implementation of Bloom's Taxonomy in test construction. Findings indicate that most educators (74.2%) struggle to formulate unambiguous questions, while 67.7% find it difficult to develop competing test distractors. However, some educators feel confident in designing effective assessments without formal training. Based on the study findings, the study concludes that there is a need for professional development initiatives to enhance educators' competencies in test construction.*

Keywords: *Non-professional educators, Objective test formats, Test construction, Bloom's Taxonomy, Tanzania.*

How to cite this work (APA):

Kambuga, Y. M. (2025). Challenges in Constructing Multiple-Choice (MC) Matching Item (MI) and True-False (TF) Test Formats among Educators Who Are Not Teachers by Profession in Tanzanian Higher Learning Institutions. *Journal of Research Innovation and Implications in Education, 9*(2), 950 – 963. **https://doi.org/10.59765/srhr45.**

1. Introduction

Competency in test construction is an important skill all higher learning educators require to evaluate learning and instructional objectives effectively. Test construction skills for teachers play a vital role in students' assessment, focusing on factual knowledge, comprehension of concepts, computational proficiency, proper application of methodologies, and practical skills (Apostol et al., 2023; Hamafyelto et al., 2015; Iqbal et al., 2022; Kissi, 2020; Monica, 2021). Other studies show that test construction ability and quality are fundamental tools any teacher requires if teaching and learning goals are to be achieved. However, it is impossible for teachers who are not competent in the arts and science of test construction to construct quality tests (Amani et al., 2021; Kapinga & Kimaro, 2019). The assessment process, which uses various techniques and tools to measure students' learning across different activities, serves the primary purpose of providing feedback to both the teacher and the student and grading the student (Bharti, 2024; Fuentealba, 2011;

Hamafyelto et al., 2015; Yambi & Yambi, 2020). An assessment for learning denotes an evaluative process designed to furnish teachers with information conducive to modifying learning activities, empowering them to shape the learning process (Abubakar et al., 2022; Munna & Kalam, 2021; Nwani et al., 2022).

Research shows that in any learning setting, tests are primary instruments of evaluation to aid in the measurement of students' learning since they aid teachers in evaluating students' comprehension and grasp of a subject (Ehigbor & Obaze Agbonluae, 2023; Quansah & Amoako, 2018; Ragma & Valdez, 2023). A test is an assessment tool a teacher uses to gather students' feedback and use the data to establish students' learning issues and prepare for subsequent teaching (Ehigbor & Obaze Agbonluae, 2023; 2015; Gichuhi, 2014; Kissi, 2020; Sibo, 2019). Research further indicates that tests give important feedback to pupils, teachers, parents, and communities, helping to evaluate the quality of education and performance of students (Kellenberg & Mobarak, 2011; Kissi et al., 2023; Mamoon-Al-Bashir et al., 2016; Obilor, 2019). Ehigbor & Obaze Agbonluae (2023) and Opie et al. (2021) also suggest that the test evaluates an individual's performance in a particular acquired area or ability. Studies observed the significant impact of test information on instructional and managerial decision-making that calls for teachers' mastery of the principles and techniques employed in test construction (Adebunmi, 2020; Mpuangnan, 2024; Ragma & Valdez, 2023).

College instructors should be capable of crafting test questions to prompt students to solve problems for others and themselves, according to the established system of Bloom's taxonomies of education (Opie et al., 2021; Quansah et al., 2019; Quansah & Amoako, 2018; Setiyana & Muna, 2019). According to Hoque (2017) and Murphy et al. (2023), Bloom's taxonomy model covers cognitive, psychomotor, and affective domains. Teachers should incorporate questions across all three areas as they create items for tests, employing action verbs that test lower and higher order of thinking, capability, and attitude (Coleman, 2017; Mutakin & Rahman Hakim, 2020; O'Neill & Murphy, 2010). Bloom's Taxonomy framework is the standard for the design of test and assessment items, and educators are required to develop test items based on its principles (Chandio et al., 2021; Mutakin & Rahman Hakim, 2020; O'Neill & Murphy, 2010). Benjamin Bloom developed three primary learning domains to enhance learners' competencies (Bloom et al., 1984; Opie et al., 2021; Quansah & Amoako, 2018). Other studies update that the cognitive domain emphasises knowledge and mental skills and consists of six levels: remembering, understanding, applying, analysing, evaluating, and creating (Chandio et al., 2021; Gichuhi, 2014; Sibo, 2019), as the affective domain focuses on emotion, attitudes, and feelings, encouraging teachers to pose questions that stimulate the learners' problem-solving abilities for themselves and others (Chandio et al., 2021; Coleman, 2017). Besides, the psychomotor domain highlights the ability to perform tasks using body muscles and encompasses levels such as perception (awareness), set, guided response, mechanism, complex overt response, adaptation, and origination (Gichuhi, 2014; Quansah & Amoako, 2018; Sibo, 2019). According to Kiss (2020) and Sibo (2019), teacher-made tests require a good understanding of Bloom's taxonomy to ensure comprehensive scheme coverage during testing.

Studies conducted in Nigeria and Ghana on the teacher's competence in test constructions show that teachers by profession tend to construct effective evaluative instruments more than non-professional teachers (Hamafyelto et al., 2015; Quansah et al., 2019; Sibo, 2019). Professional instructors typically utilise various assessment methods effectively, a practice that is less common among non-professional instructors (Sibo, 2019). A study by Hamafyelto et al. (2015) found that some instructors in Nigerian schools create inadequate tests, while others rely on copied test items due to insufficient expertise in test design. Other studies indicate that test construction is a significant source of poor testing among many school teachers, especially less experienced teachers (Ehigbor & Obaze Agbonluae, 2023; Quansah et al., 2019). On the other hand, Capan Melser et al. (2020), Dibattista and Kurzawa (2011) and Sibo (2019) claim that poorly designed tests would result in low-quality classroom-based examination performance. A study by Gichuhi (2014) found that out of 38 teachers, 30 (78.9%) in Kiambu County in Kenya were unsure if they had the necessary skills and competencies to design tests in their subjects. Gichuhi's study further reveals that although most teachers had heard about Bloom's Taxonomy, only 10.5% incorporated it into their testing methods (Gichuhi, 2014).

Studies in Ghana have determined that teachers fail to uphold testing principles and, therefore, score low in testing practice (Asamoah-Gyimah, 2022; Quansah et al., 2019). Adebunmi (2020) and Hamafyelto et al. (2015) confirm that there exists a general concern regarding the shortage of test construction skills among educators in various study areas and levels of study in most developing nations. Adebunmi (2020) identifies that low-quality test items are a significant issue in that the application of poor items to gauge accomplishment can lead to mistakes in reporting students' accomplishment. To protect the accomplishment of learning by students and the success of the assessment process, this study investigated non-professional teachers' challenges in crafting multiple-choice, multiple-true-false, and true-false test tasks in Tanzanian institutions of higher learning. Specifically, the study seeks to; identify the General issues faced by non-

professional instructors in constructing Multiple-Choice (MC), Matching Item (MI), and True-False (TF) Test Formats to the clarity of questions, relevance, and congruence with learning objectives; identify how the lack of formal instruction in test construction affects the quality and efficiency of the test items in terms of the factors of question validity, reliability, and potential biases; and determine the factors influencing the use of Bloom's Taxonomy for teacher tests in their areas of study.

2. Literature Review

This study draws its foundation from Bloom's Revised Taxonomy framework by Anderson and Krathwohl (2001) and Biggs' constructive alignment theory (1996). Anderson and Krathwohl's taxonomy emphasise that instructors need to design assessment activities that are congruent with cognitive levels, ranging from recalling to creating. Constructive alignment, however, demands congruence among learning outcomes, teaching strategies, and assessment activities. These experts suggested that pedagogically unprepared teachers could develop low-level test items, which could contribute to affecting student learning outcomes. Haladyna et al. (2022) opine that test development requires measurement and assessment professionals trained to develop reliable, valid, and face-valid tests. Rubeba (2024) also adds that educators who are developing objective test formats need to be proficient in the skill of constructing high-quality ordered items to measure a wide range of cognitive levels.

Research indicates that instructors who lack pedagogical training lack the competence required to create high-quality ordered tests that demonstrate reliability, validity, and face validity (Rubeba & William, 2019; Fields, 2019; Quileste et al., 2020). A study conducted by Amani et al. (2021) indicates that even professional teachers have challenges designing high-quality ordered tests in line with the test grid (table of specifications), particularly those without pedagogical training. Rubeba (2024) suggests that the teachers reflect on whether an objective or subjective presentation will be most appropriate for evaluating students' progress in learning, as well as providing feedback to teachers and students. Therefore, in formulating such test tasks, teachers should abide by the test construction rules by having a test grid (table of specifications) apportion questions according to the provided learning domains. Amani et al. (2021) also add that instructors are most likely not skilled at developing high-quality classroom tests, particularly in using tables of specification and test-item analysis. During a study by Kembo (2020), Mwakamele (2018), and Rubeba (2024), it was discovered that Tanzanian higher institutions' lecturers are inclined to use poorly developed objective test items due to the fact that they have minimal knowledge of Bloom's Taxonomy.

Biku et al. (2018) and Fialho et al. (2023) argue that inexperienced teachers rely more on instinct than on following systematic testing templates. Sanga (2016) suggested that schools implement tailor-made programs to introduce and familiarise educators with effective instructional approaches and assessment methods to assist them in creating high-quality test tasks in line with expected learning objectives. Despite considerable research efforts in Tanzania in the area of test construction incompetence in secondary schools, there has been an enormous gap in the literature concerning higher learning institutions, particularly for instructors not formally pedagogically trained. This research aims to bridge that gap.

3. Methodology

3.1. Design

Research design is the procedure involved in the research process: data collection, analysis, and report writing (Chali et al., 2022; Creswell, 2014). The research design in this study is descriptive. According to Creswell (2013), a descriptive design is a detailed method that systematically explains and describes the research object. Descriptive research describes and interprets things, such as the condition of a thing or relationship, opinions that develop, an effect that occurs, and ongoing trends in the community (Cresswell, 2014). This design provides a detailed and accurate picture of the challenges faced by non-professional teacher lecturers in constructing MC, MI, and TF test formats.

3.2. Population and Sampling

The study was conducted in Dodoma City, the capital of Tanzania, using higher learning institutions within the city. The target population was all educators who did not have formal training as teachers. Online questionnaires were sent to institutions through institutional websites and to colleagues working in those institutions. As described by Alessi (2010) and Tanner (2018), convenience sampling was the best technique to use when a survey is posted on a website, and all visitors are invited to respond or when an invitation to participate is circulated. Using an online survey, respondent anonymity is preserved, and responses may be more candid. As Dillman et al. (2009) and Ritter & Sue (2007) describe, using institutional websites tends to make response rates for online surveys relatively low.

3.3. Data Collection Method

The research employed quantitative research design, prudently structured to gather measurable data. This approach was effectively implemented through an online questionnaire survey, which was designed to capture a wide range of responses from participants. The survey included a series of carefully formulated questions aimed at gathering specific information and insights related to the study's objectives. An internet-based survey contained three parts: Part A gathered individual information for demographic analysis, part B dealt with identification of issues faced by teachers in test design using open-ended questions to restrict response, and Part C employed a five-point Likert scale for assessing agreement or otherwise of the teachers' response with statements related to aims of cognitive levels of Bloom's Taxonomy in constructing test items.

3.4. Data Analysis

The research data were analysed using descriptive statistics (e.g., percentages, frequencies, mean score and standard deviations). The inferential statistical tests, such as t-tests, were also employed to find significant differences in what challenges lecturers face based on their academic discipline, teaching experience in years, and the type of test formats. Further, regression analysis was done to find out how variables like knowing Bloom's Taxonomy, institutional help, and not receiving any formal training affect teachers in using the Bloom's Taxonomy framework when designing tests.

4. Results and Discussion

4.1. Social demographic characteristics

The research accounted for some variables in terms of demographic attributes, including gender, level of education, and teaching experience, as indicated in Table 1. This overview gives an account of the respondents' demographic profile and information about the distribution of significant demographic variables in the study population. Regarding gender representation, the data reveal a notable gender disparity, with the majority of respondents being male (74%), while females accounted for a slightly smaller proportion (32.26%). Examining the education level of respondents, the data revealed that most respondents (74%) have a master's level, followed by those with PhD qualifications, accounting for 19.4%. However, a low percentage of PhD holders (19.4%) indicates that not all educators know educational assessment or pedagogy training. The lack of formal training for many educators can lead to challenges in designing impactful multiple-choice (MC), multiple-item (MI), and true-false (TF) tests. In addition, Bachelor's degrees accounted for 12.9% of the total respondents. It was observed that most respondents had five to ten years of teaching experience (64.52%), while 19.35% and 16.13% had one to five-ten years and above. The findings imply that educators with 1 to 5 years of teaching experience may lack comprehensive expertise in high-quality test design compared to their counterparts with 10 or more years of experience in the field.

Table 1: Social demographic characteristics of respondents

Variables		Freq.	Percent
Sex	Male	21	67.74%
	Female	10	32.26%
Educational Level	Bachelor’s degree	4	12.90%
	Master’s degree	21	67.74%
	PhD	6	19.35%
Working Experience	1-5	6	19.35%
	6-10	20	64.52%
	11 and above	5	16.12%

4.2 Common challenges faced by non-professional teachers in constructing objective questions

Competence in constructing an item test becomes an absolute requirement every educator in higher learning institutions must possess, regardless of formal training attained. However, developing great items of objective tests requires significantly more ability, effort, and time. test is not easy work. Tables 2 and 3 underline non-professional educators' prevalent challenges in constructing objective assessments. The data indicate that the majority of teachers (61.29%) administer two tests per semester, while 38.71% conduct three or more tests within

the same period. Regarding the duration allocated for test construction, the findings show that most educators (58.06%) dedicate several days to this task, 9.68% invest a single day, 25.81% spend a few hours, and 6.45% allocate just a few minutes. Furthermore, many educators (74.2%) reported that they sometimes struggle to formulate unambiguous questions, while only 6.5% and 9.7% stated they never or rarely face such difficulties.

Table 2: Common challenges facing non-professional teachers in constructing objective questions

Average tests constructed by Educators per term	Freq.	Percent
Two tests	19	61.29
Three tests and the Above	12	38.71
Time taken by an educator to construct one test in a subject area		
Few minutes	2	6.45
A few hours	8	25.81
A day	3	9.68
A number of days	18	58.06
How educators struggle to formulate clear, unambiguous questions for their MC, MI, and TF tests		
No	10	32.26
Yes	21	67.74

Furthermore, we ran multiple responses on the common issues encountered when creating multiple-choice (MC), multiple-answer (MI), and true-false (TF) test questions. The results showed that 83.9% of respondents were concerned about the Difficulty of creating plausible distractors (for MC). This was followed by ambiguity in question-wording, which 54.8% of respondents noted.

Table 3: Common Issues Encountered when creating MC, MI and TF Questions

Challenges	Freq	Percent	Percent of Cases
Ambiguity in question-wording	17	27.9%	54.8%
Difficulty in creating plausible distractors (for MC)	26	42.6%	83.9%
Ensuring fairness and a lack of bias	12	19.7%	38.7%
Misalignment with learning objectives	6	9.8%	19.4%
Total	61	100.0%	196.8%

Besides, the study performed a descriptive statistic of mean and standard deviation to measure educators' confidence level in preparation for the MC, MI and TF test questions without formal educational measurement and assessment training. The results show that the mean score of all measured variables was above the midpoint (3) with an average standard deviation of 0.7, although there is some variability. This signifies that the majority of educators are comfortable designing effective assessments, while others may feel less equipped, particularly in the absence of formal training, as detailed in Table 4.

Table 4: The absence of formal training in test construction impacts the quality of test items, including their validity, reliability, and biases

Variable	Obs	Mean	Std. Dev.	Min	Max
How confident do educators feel about the validity of their test questions?	31	3.516	.724	2	5
How confident are educators in their test questions to yield consistent results across different student groups and times	31	3.419	.672	2	5
How often do educators review test questions for biases that may disadvantage certain student groups?	31	2.903	1.012	1	5
How does a lack of formal training in test construction affect student performance and learning outcomes?	31	3.452	.888	2	5
How often do educators receive feedback on their test questions, and how do they use this feedback to improve test construction skills	31	2.484	1.122	1	5

4.3 Factors influencing the implementation of Bloom's Taxonomy in teacher tests

The multiple responses were performed on educators' challenges when applying Bloom's Taxonomy to their test questions. The results showed that Time constraints were a common concern for 64.5 % of respondents. This was followed by Difficulty in creating higher order thinking questions, which was noted by 61.3% of respondents.

Table 5: Challenges educators encounter when applying Bloom's Taxonomy to test questions

Challenges Educators Encounter	Freq	Percent	Percent of Cases
Time constraints	20	35.7%	64.5%
Lack of familiarity with the taxonomy	12	21.4%	38.7%
Difficulty in creating higher-order thinking questions	19	33.9%	61.3%
Limited resources or examples	5	8.9%	16.1%
Total	56	100.0%	180.6%

In addition, multiple responses regarding support or resources for improving skills in creating clear, relevant, and aligned questions were also made. The results showed that the majority of respondents, scoring 83.9%, mentioned professional development workshops. However, mentorship from experienced educators and access to a bank of questions and templates were 9.7% and 16.1%, respectively, as detailed in Table 6.

Table 6: Type of support or resources educators need to improve their skills in creating clear, relevant, and aligned test questions

Type of support or resources required	Freq	Percent	Percent of Cases
Professional development workshops	26	46.4%	83.9%
Mentoring from experienced educators	8	14.3%	25.8%
Peer collaboration and review	6	10.7%	19.4%
Online resources or tutorials	8	14.3%	25.8%
Access to question banks and templates	5	8.9%	16.1%
Experienced educators	3	5.4%	9.7%
Total	56	100.0%	180.6%

4.4 Association between Factors Influencing the Implementation of Bloom's Taxonomy and Education Level

The study further investigated the relationship between factors influencing the implementation of Bloom's Taxonomy and education level, as shown in Table 4. A Chi-square test for association was conducted, revealing a p-value of less than 0.05, which indicates a significant association with educational level. The results showed that both familiarity with the different levels of the Cognitive Bloom's Taxonomy Domain (which include Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating) and the frequency of using Bloom's Taxonomy as a framework for designing test questions were significantly associated with education level (p-value less than 0.05). This suggests that as education level increases, familiarity with the various levels of Bloom's Taxonomy and the frequency of its application in test design also increases. The findings suggest that educators with a bachelor's degree often lack familiarity with Bloom's Taxonomy framework, unlike those with a Master's or PhD, who report similar levels of familiarity. This implies that as educators further their education, they encounter various learning and assessment theories, which they apply in their professional practices. On the other hand, other factors did not show significant associations.

Table 7: Association between Factors influencing the implementation of Bloom's Taxonomy and Education Level

Items		Educational level							
		Bachelor		Master's		PhD		Chi-square	Sig
		N	Percent	N	%	N	%		
Ever heard of Bloom's Taxonomy Domain	Yes	4	80.0%	17	85.0%	4	66.7%	0.995	0.608
	No	1	20.0%	3	15.0%	2	33.3%		
Familiarity with the different levels of Bloom's Cognitive Taxonomy Domain (e.g., remembering, understanding, applying, analysing, evaluating, creating)	I am not familiar at all	0	0.0%	2	10.0%	1	16.7%	3.737	0.012
	Slightly familiar	3	60.0%	6	30.0%	3	50.0%		
	Moderately familiar	1	20.0%	7	35.0%	2	33.3%		
	Very familiar	1	20.0%	5	25.0%	0	0.0%		
	Extremely familiar	0	0.0%	0	0.0%	0	0.0%		
Use Bloom's Taxonomy as a framework when designing test questions	Never	0	0.0%	2	10.0%	2	33.3%	5.888	0.040
	Rarely	1	20.0%	5	25.0%	1	16.7%		
	Sometimes	3	60.0%	7	35.0%	3	50.0%		
	Often	0	0.0%	2	10.0%	0	0.0%		
	Always	1	20.0%	4	20.0%	0	0.0%		
Institutional support provides resources or training on using Bloom's Taxonomy for test construction	Not supportive at all	1	20.0%	2	10.0%	1	16.7%	2.497	0.869
	Slightly supportive	2	40.0%	7	35.0%	3	50.0%		
	Moderately supportive	1	20.0%	9	45.0%	2	33.3%		
	Very supportive	1	20.0%	2	10.0%	0	0.0%		
	Extremely supportive	0	0.0%	0	0.0%	0	0.0%		

4.5 Association between Factors influencing the implementation of Bloom's Taxonomy and teaching experience

The study examined the relationship between factors influencing the implementation of Bloom's Taxonomy and teaching experience, as shown in Table 5. A Chi-square test for association was used, revealing a p-value of less than 0.05, which indicates a significant association with teaching experience. The results show that both familiarity with the different levels of the Cognitive Bloom's Taxonomy Domain (e.g., Remembering, Understanding, Applying, Analysing, Evaluating, Creating) and the frequency of using Bloom's Taxonomy as a framework for designing test questions are significantly associated (p-value less than 0.05) with teaching experience. This suggests that as teaching experience increases, so does familiarity with the levels of Bloom's Taxonomy and the frequency of its use in question design.

Table 7: Association between Factors influencing the implementation of Bloom's Taxonomy and teaching experience

Items		Teaching experience						t-test	
		1 - 5 years		5 - 10 years		10 years and above		Chi-square	Sig
		N	%	N	%	N	%		
Ever heard of Bloom's Taxonomy Domain	Yes	4	66.7%	16	80.0%	5	100.0%		
	No	2	33.3%	4	20.0%	0	0.0%	1.956	0.376
Familiarity with the different levels of Bloom's Cognitive Taxonomy Domain (e.g., remembering, understanding, applying, analysing, evaluating, creating)	I am not familiar at all	1	16.7%	2	10.0%	0	0.0%		
	Slightly familiar	4	66.7%	6	30.0%	2	40.0%		
	Moderately familiar	1	16.7%	9	45.0%	0	0.0%		
								11.091	0.046
	Very familiar	0	0.0%	3	15.0%	3	60.0%		
	Extremely familiar	0	0.0%	0	0.0%	0	0.0%		
Use Bloom's Taxonomy as a framework when designing test questions	Never	1	16.7%	2	10.0%	1	20.0%		
	Rarely	3	50.0%	4	20.0%	0	0.0%		
	Sometimes	2	33.3%	9	45.0%	2	40.0%	6.808	0.038
	Often	0	0.0%	1	5.0%	1	20.0%		
	Always	0	0.0%	4	20.0%	1	20.0%		
Institutional support in providing resources or training on the use of Bloom's Taxonomy for test construction	Not supportive at all	0	0.0%	2	10.0%	2	40.0%		
	Slightly supportive	3	50.0%	8	40.0%	1	20.0%		
	Moderately supportive	3	50.0%	7	35.0%	2	40.0%	6.329	0.558
	Very supportive	0	0.0%	3	15.0%	0	0.0%		
	Extremely supportive	0	0.0%	0	0.0%	0	0.0%		

Additionally, how frequently educators utilise Bloom's Taxonomy as a framework for developing high-quality order test questions correlates significantly with their teaching experience. Educators with 10 years or more teaching experience frequently use this framework for designing their tests of this framework, compared to 50% of educators with 1 to 5 years of experience, who indicated that they rarely employ it. Notably, other factors explored in the study, such as the availability of institutional support and resources for implementing Bloom's Taxonomy, did not show significant associations. Besides, the study performed a Logistic model fit Test of factors influencing the implementation of Bloom's Taxonomy as detailed in the Hosmer and Lemeshow Test in Table 9.

Table 9: Hosmer and Lemeshow Test

Model	Chi-Squire	df	Sign.
Factors influencing the implementation of Bloom's Taxonomy	**15.0713**	**5**	**0.15**

At the 5% significance level, no significant evidence suggests that the logistic regression model does not fit the data well. In other words, the Hosmer and Lemeshow Test indicates that the model fits the data adequately, meaning that the predicted values are reasonably close to the observed values. Therefore, based on this test, the model can be considered a good fit for the data.

Table 8: Logistic regression equation showing the factors influencing the implementation of Bloom's Taxonomy

Use Boom Taxonomy	**Odds ratio**	**St. Err.**	**t-value**	**p-value**	**[95% Conf**	**Interval**
Familiar with Boom Taxonomy	6.525	6.494	1.88	.042	1.928	45.887
Institution Support	2.416	2.453	0.87	.03	1.33	17.677
Lack of formal Training	.319	.281	1.29	.01	1 .056	1.8
Constant	1.585	6.26	7.12	.007	.001	3655.123

Mean dependent var	0.677	SD dependent var	0.475
Pseudo r-squared	0.260	Number of obs	31
Chi-square	10.134	Prob > chi2	0.017
Akaike crit. (AIC)	36.852	Bayesian crit. (BIC)	42.588

In this model, the variable familiarity with Bloom's Taxonomy shows a strong positive relationship with the likelihood of using Bloom's Taxonomy in test construction. The odds ratio for familiarity is 6.525, meaning lecturers familiar with Bloom's Taxonomy are approximately 6.5 times more likely to use it in their test creation than those unfamiliar with the framework. The p-value of 0.042 is below the 0.05 significance threshold, indicating that this relationship is statistically significant. This suggests that familiarity with Bloom's Taxonomy is a crucial factor in its implementation, and those who clearly understand the framework are much more likely to incorporate it into their assessments. The 95% confidence interval [1.928, 45.887] does not include 1, further supporting the conclusion that familiarity significantly influences the likelihood of using Bloom's Taxonomy.

On the contrary, Institutional support also plays a significant role in the adoption of Bloom's Taxonomy. The odds ratio of institutional support is 2.416, indicating that lecturers with support from their institutions are about 2.4 times more likely to adopt Bloom's Taxonomy than those without any such support. The p-value of 0.03 is less than 0.05, and therefore, this is statistically significant at the 5% level. This finding suggests that institutional support, whether in training, resources, or motivation, can play an enormous positive role towards lecturers' use of Bloom's Taxonomy. The 95% confidence interval [1.33, 17.677] corroborates the significance of this relationship, as it does not include 1.

Moreover, the variable lack of formal training in educational assessment negatively affects the possibility of using Bloom's Taxonomy. The odds ratio of lack of formal training is 0.319, which means that lecturers who lack formal training in educational assessment are approximately 68% less likely to use Bloom's Taxonomy than their counterparts who have received formal training. The p-value 0.01 is below 0.05, indicating that this factor is statistically significant at the 5% significance level. This shows that formal training in educational measurement is an important factor in enabling lecturers to apply Bloom's Taxonomy in test construction effectively. The 95% confidence interval [1.056, 1.8] suggests that this factor has a significant impact, as the confidence interval does not include 1.

The performed model indicates that the constant term in the logistic regression equation represents the baseline odds of using Bloom's Taxonomy when all the predictors (familiarity, institutional support, and formal training) are at their reference values (i.e., no familiarity, no institutional support, and lack of formal training). The odds ratio for the constant is 1.585, suggesting a baseline likelihood of using Bloom's Taxonomy even when the predictors are not in favour. The p-value of 0.007 is below 0.05, indicating that the constant term is statistically significant. However, the constant's practical interpretation is limited since it represents the baseline level of use when other factors are at reference levels, and the other predictors influence its value in the model.

In addition, the model Fit Statistics on the Pseudo R-squared value of 0.660 indicates that the model explains 66% of the variance in the dependent variable, i.e., the likelihood of using Bloom's Taxonomy. While this is a high level of explanatory power, it does suggest that the predictors included in the model (familiarity, institutional support, and lack of formal training) account for a moderate portion of the variance. The Chi-square statistic of 10.134 with a p-value of 0.017 indicates that the model as a whole is statistically significant. This means that the factors included in the model have a meaningful impact on

the likelihood of implementing Bloom's Taxonomy, and the model fits the data well.

5. Conclusion and Recommendations

5.1 Conclusion

Testing students' learning achievements cannot be under-emphasised because teaching and learning can never be complete without it. Although some educators have undergone training, many struggle to formulate clear, relevant, and aligned test questions for accurately assessing student learning outcomes. Difficulties in creating unambiguous questions, formulating plausible distractors, and aligning questions with course content and learning objectives were reported to be major obstacles. Although educators seem to struggle to construct high-quality objectives and questions, the lack of formal training in educational assessment has emerged as a critical factor contributing to these challenges. However, some educators do not know much about applying Bloom's Taxonomy, with many educators not fully utilising this framework to guide their assessment practices.

5.2 Recommendations

Test construction still seems challenging for many educators in higher learning institutions. Therefore, the study recommends that educational institutions organise regular professional development workshops on assessment design and item construction to enhance educators' skills; educators should be reminded to follow stipulated test construction procedures set by their institutions and use the skills attained from their various training to set their tests; new staff should be oriented on the stipulated test construction procedures the institution had the necessities of adhering to them; and institutions should strengthen feedback systems that allow lecturers to receive constructive critiques on their test items from peers or external examiners, which will help improve clarity and relevance.

References

Abubakar, M., Husain, N., & Mardiana, M. (2022). Developing Higher Order Thinking Skills Test Instruments for English Speaking Classes: Sing a Song Material. Eduvelop: *Journal of English Education and Development, 6*, 40–54. https://doi.org/10.31605/eduvelop.v6i1.1870

Adebunmi, M. (2020). Lecturers' Perceptions of Their Knowledge and Skills in The Construction of Suitable Test Items. 7(4).

Alessi, E. (2010). Conducting an Internet-based Survey: Benefits, Pitfalls, and Lessons Learned. *Social Work Research, 34*, 122–128. https://doi.org/10.1093/swr/34.2.122

Amani, J., Kitta, S., Kapinga, O., & Mbilinyi, C. (2021). Secondary School Teachers' Knowledge of Procedures for Constructing Quality Classroom Tests in Tanzania. ÜNİVERSİTEPARK Bülten, 10. https://doi.org/10.22521/unibulletin.2021.101.3

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition. Addison Wesley Longman, Inc.

Apostol, E., Sereño, R. R., Cuevas, G., Sumicad, R., & Pinatil, M. (2023). Evaluating Educators: A Comprehensive Study of Teachers' Assessment Skills and Practices. *Journal of World Englishes and Educational Practices, 5*, 56–69. https://doi.org/10.32996/jweep.2023.5.3.5

Asamoah-Gyimah, K. (2022). Influence of Knowledge of Assessment on Test Construction Skills among Tutors of the College of Education in Ghana. *American Journal of Education and Practice, 6*(3), Article 3. https://doi.org/10.47672/ajep.1177

Bharti, S. (2024). Assessment of students by using tools to assess the holistic performance of students at the primary level of students. *International Journal of Multidisciplinary Research in Arts, Science and Technology, 2*, 10–18. https://doi.org/10.61778/ijmrast.v2i5.57

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education, 32*(3), 347-364.

Biku, T., Demas, T., Woldehawariat, N., Getahun, M., & Mekonnen, A. (2018). The effect of teaching without pedagogical training in St. Paul's hospital millennium medical college, Addis Ababa, Ethiopia. Advances in Medical Education and Practice, 893-904.

Bloom, B. J., David, K. K., & Betram, M. B. (1984). Taxonomy of educational objectives: The classification of educational goals. Longman.

Capan Melser, M., Steiner-Hofbauer, V., Lilaj, B., Agis, H., Knaus, A., & Holzinger, A. (2020). Knowledge, application and how about competence? Qualitative assessment of multiple-choice questions for dental

students. *Medical Education Online, 25*(1), 1714199. https://doi.org/10.1080/10872981.2020.1714199

Chali, M., Eshete, S., & Debela, K. (2022). Learning How Research Design Methods Work: A Review of Creswell's Research Design: Qualitative, Quantitative and Mixed Methods Approaches. *The Qualitative Report, 27*(12), 2956–2960. https://doi.org/10.46743/2160-3715/2022.5901

Chandio, M., Zafar, N., & Muhiyuddin, G. (2021). Bloom's Taxonomy: Reforming Pedagogy Through Assessment. *Journal of Education and Educational Development, 8*, 110. https://doi.org/10.22555/joeed.v8i1.308

Child, S. F. J., & Vitello, S. (n.d.). Faster, higher, stronger? A comparison of content coverage and teacher pedagogy in Technical Awards and GCSEs in sport-related subjects.

Coleman, V. (2017). On the reliability of applying educational taxonomies. https://doi.org/10.17863/CAM.100360

Creswell, J. W. (2013). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed). SAGE Publications.

DiBattista, D., & Kurzawa, L. (2011). Examination of the Quality of Multiple-choice Items on Classroom Tests. Canadian Journal for the Scholarship of Teaching and Learning, 2. https://doi.org/10.5206/cjsotl-rcacea.2011.2.4

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research, 38*(1), 1–18. https://doi.org/10.1016/j.ssresearch.2008.03.007

Ehigbor, B. O., & Obaze Agbonluae, P. D. (2023). Assessment of Multiple-Choice Construction Competence Among Public Junior Secondary School Teachers in Edo Central Senatorial District, Nigeria. *International Journal of Research and Innovation in Social Science, 7*(4), 516–525.

Fialho, I., Cid, M., & Coppi, M. (2023, May). Grounding and improving assessment in higher education: a way of promoting quality education. In Frontiers in Education (Vol. 8, p. 1143356). Frontiers Media SA.

Fields, M. (2019). Common Errors in Teacher-Made Test Design. In Handbook of Research on Assessment Literacy and Teacher-Made Testing in the Language Classroom (pp. 328-346). IGI Global.

Fuentealba, C. (2011). The Role of Assessment in the Student Learning Process. *Journal of Veterinary Medical Education, 38*(2), 157–162. https://doi.org/10.3138/jvme.38.2.157

Gicharu, C. W. (2015). Factors influencing compliance with disaster risk reduction guidelines in public primary schools in Kiambaa Division, Kiambu County, Kenya. https://www.semanticscholar.org/paper/Factors-influencing-compliance-with-disaster-risk-Gicharu/3ecf02788e5769a56ed55e53862842807acbe849

Gichuhi, C. (2014). Teachers' competence in tests construction within Bloom's taxonomy for effective learning assessment: A case study of Kikuyu district, Kiambu county. [Thesis, University of Nairobi]. http://erepository.uonbi.ac.ke/handle/11295/76048

Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. Routledge.

Hamafyelto, R. S., Hamman-Tukur, A., & Hamafyelto, S. S. (2015). Assessing Teacher Competence in Test Construction and Content Validity of Teacher-Made Examination Questions in Commerce in Borno State, Nigeria. Education 3-13. https://www.semanticscholar.org/paper/Assessing-Teacher-Competence-in-Test-Construction-Hamafyelto-Hamman-Tukur/2481287f7e383ade3ef25c0ea633acea3c0a285c

Hoque, M. (2017). Three Domains of Learning: Cognitive, Affective and Psychomotor. 2, 45–51.

Iqbal, Z., Haider, M., & Qamar, A. (2022). Observing Teachers Skills about the Assessment of Students: A Standard-based Teacher Evaluation. Global Educational Studies Review, VII, 59–70. https://doi.org/10.31703/gesr.2022(VII-IV).06

Kapinga, B. B., & Kimaro, A. R. (2019). Test Construction Competence among Primary School Teachers in Dar es Salaam Region, Tanzania. *Journal of Adult Education in Tanzania, 22*(1), Article 1. https://jaet.iae.ac.tz/index.php/adulteducation/article

/view/42

Kellenberg, D., & Mobarak, A. M. (2011). The Economics of Natural Disasters (SSRN Scholarly Paper 1928371). https://doi.org/10.1146/annurev-resource-073009-104211

Kembo, J. (2020). How effectively are university students tested? A case study. *East African Journal of Education Studies, 2*(1), 59-66.

Kissi, P. (2020). Multiple-choice construction competencies and items‟ quality: Evidence from selected senior high school subject teachers in Kwahu-South District [Thesis, University of Cape Coast]. http://ir.ucc.edu.gh/jspui/handle/123456789/4641.

Kissi, P., Baidoo-Anu, D., & Anane, E. (2023). Teachers' test construction competencies in examination-oriented educational system: Exploring teachers' multiple-choice test construction competence. Frontiers in Education, 8. https://doi.org/10.3389/feduc.2023.1154592.

Mamoon-Al-Bashir, M., Kabir, M. R., & Rahman, I. (2016). The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education. Journal of Education and Practice. https://www.semanticscholar.org/paper/The-Value-and-Effectiveness-of-Feedback-in-Learning-Mamoon-Al-Bashir-Kabir/532e30684334206e2fd8e120f038f1fe399bea91.

Monica, S. (2021). The Challenges in Constructing and Administering an Objective Test for Beginner Students During the COVID-19 Pandemic. https://doi.org/10.2991/assehr.k.210227.075.

Mpuangnan, K. (2024). Determining skills of basic schoolteachers in test item construction. Review of Education, 12. https://doi.org/10.1002/rev3.3478

Munna, A., & Kalam, M. A. (2021). Teaching and learning process to enhance teaching effectiveness: Literature review. *International Journal of Humanities and Innovation (IJHI), 4*, 1–4. https://doi.org/10.33750/ijhi.v4i1.102.

Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The Value of Using Tests in Education as Tools for Learning—Not Just for Assessment. *Educational Psychology Review, 35*(3), 89. https://doi.org/10.1007/s10648-023-09808-3

Mutakin, T., & Rahman Hakim, A. (2020). Teachers' Ability in Designing Test Assessments. https://doi.org/10.2991/assehr.k.201230.070

Mwakamele, M. I. (2018). Using teachers' knowledge of Bloom's Taxonomy in determining the quality of *classroom assessment in secondary schools. Journal of Issues and Practice in Education, 10*(1), 44-66.

Nwani, S. U., Ezugwu, I., & Nnenanya, G. (2022). Performance Evaluation of teachers as a correlate of instructional effectiveness and students' academic achievement in Basic science. 9, 256–265.

Nyinge, B., Matete, R., & William, F. K. (2024). Exploring the Relationship between Authentic Assessment and Teaching Professional Competence Acquisition among Undergraduate Science Student-Teachers in Higher Education Institutions in Tanzania. International Journal of Current Educational Studies, 3(1). (mbadala wa nduguru

O'Neill, G., & Murphy, F. (2010). Guide to taxonomies of learning. https://www.semanticscholar.org/paper/Guide-to-taxonomies-of-learning-O%E2%80%99Neill-Murphy/f630b6ef1a76f7b0566627bd0a359881308be296

Obilor, E. I. (2019). Feedback and Students' Learning. *International Journal of Innovative Research in Education, 7*, 40–47.

Opie, O., Oko-Ngaji, V., & Nsor, J. (2021). An Assessment of Science Teachers' Utilization of the Knowledge of Test Construction Procedure in Multiple Choice Objective Tests in Secondary Schools in Yala LGA, Cross River State. 9, 54–62.

Quansah, F., & Amoako, I. (2018). Attitude of Senior High School Teachers Toward Test Construction: Developing and Validating a Standardised Instrument. Research on Humanities and Social Sciences. https://www.semanticscholar.org/paper/Attitude-of-Senior-High-School-Teachers-Toward-Test-Quansah-Amoako/ac4b28a168041b79faf36989e75e3c18352da390.

Quansah, F., Amoako, İ., & Ankomah, F. (2019). Teachers' Test Construction Skills in Senior High Schools in Ghana: Document Analysis. *International Journal of Assessment Tools in Education, 6*(1), 1–8. https://doi.org/10.21449/ijate.481164

Quileste, R. M., & Moreno, L. A. S. R. (2020). The pedagogical content knowledge of the school of education pre-service teachers in test construction. *JPAIR Institutional Research Journal, 14*(1), 1-1.

Ragma, F., & Valdez, F. J. (2023). Test Construction Proficiency of Teachers.

Ritter, L., & Sue, V. (2007). Introduction to using online surveys. New Directions for Evaluation, 2007, 5–14. https://doi.org/10.1002/ev.230

Rubeba, A. (2024). Addressing the Challenges Facing University Instructors in Test Construction Practices in Tanzania. *Educational Challenges, 29*(1), 159-174.https://doi.org/10.34142/2709-7986.2024.29.1.11 A

Rubeba, A., & William, F. (2019). Instructors' test construction competences and its implications on teaching and learning in Tanzanian universities: A preliminary study in three universities in Tanzania. *Journal of Adult Education, 22*, 111–130.

Sanga, P. L. (2016). Implications of teacher educators' practices in assessment for student learning in Tanzania. *Makerere Journal of Higher Education, 8*(1), 3–24. http://dx.doi.org/10.4314/majohe.v8i1.1

Setiyana, R., & Muna, H. (2019). Evaluating Bloom's Taxonomy in Test Items. Proceedings of EEIC, 2, 79-83.

Sibo, M. F., & Obulor, O. C. (2019). Implementation of Bloom's Taxonomy in the Construction of Teacher-made Biology Test in Public Senior Secondary Schools in Rivers State, Nigeria. International Journal of Entrepreneurship & Business Innovation, 2(1), 12-20.

Tanner, K. (2018). Chapter 6—Survey designs. In K. Williamson & G. Johanson (Eds.), Research Methods (Second Edition) (pp. 159–192). Chandos Publishing. https://doi.org/10.1016/B978-0-08-102220-7.00006-6

Yambi, T. D. A. C., & Yambi, C. (2018). Assessment and evaluation in education. University Federal do Rio de Janeiro, Brazil.