Educational Dimension, 2027, Vol. **16**, pp. 245–270 https://doi.org/10.31812/ed.973

Self-perceived assessment competencies and practices of instructors at Tanzanian universities: an evaluation of university lecturers' compliance with assessment standards

Asia Mbwebwe Rubeba

Department of Psychology and Curriculum Studies, The University of Dodoma, 1 Benjamin Mkapa Rd., 41218 Iyumbu, Dodoma, Tanzania

Abstract. This study examined self-perceived assessment competencies and practices among university instructors in Tanzania. The study was conducted across three public universities, using ex post facto and transcendental phenomenological designs. Stratified proportional sampling, simple random sampling, and purposive sampling were used to select and collect data from a sample of 205 instructors. Questionnaires, interviews, observations and documentary analysis were used concurrently to gather data. The qualitative data were analysed thematically, whereas the quantitative data were analysed using logistic regression. The analysis revealed significant differences among instructors on indicators for self-perceived assessment practices and demographic parameters. This study, therefore, proposes that universities make deliberate efforts to hold frequent workshops and seminars for instructors to improve their assessment practices.

Keywords: assessment competencies, assessment literacy, test construction practices, university instructors

1. Introduction

The assessment of students' tasks has been done since time immemorial. It plays a central role in the teaching and learning process and serves as a driving force for school attainment, student performance, and educational reforms worldwide. By its definition, assessment is a systematic, continuous process of monitoring the various pieces of learning to evaluate student achievement and instructional effectiveness [39]. It is based on a collection of information about what learners know and what they can do [2]. The teaching and learning process requires continuous monitoring, while students' progress requires assessment [62]. Literatures support the importance of assessment in improving teaching and learning processes [28, 54].

Assessment helps lecturers to determine the effectiveness of their teaching techniques and learning materials. The lecturer may use assessment feedback to refine and redefine the learning outcomes. Also, it helps to determine the general trend in the development of the teaching/learning process. Besides, lecturers conduct assessments to determine students' understanding and whether they can correct misconceptions. Furthermore, assessment results inform policymakers and other educational stakeholders in determining whether the knowledge students possess meets current job-market demand.

Assessment in universities is done in various ways. It is done as a formative procedure, called continuous assessment, and as a summative procedure, called the university examination – the former aims to promote students' learning, and the latter

0000-0001-7136-5486 (A. M. Rubeba)
rubeba2016@gmail.com, asia.rubeba@udom.ac.tz (A. M. Rubeba)

for grading purposes and curriculum improvement. Formative assessment means frequent, interactive assessments of students' progress and understanding to identify learning needs and adjust teaching appropriately [46]. Other scholars define formative assessment as a range of formal assessments conducted during the learning process to modify teaching and learning activities and improve students' attainment [44, 48]. The purpose of this is to determine whether students can do what they were previously unable to do after the learning experience. The tools used to measure formative assessment include individual/group assignments, tutorial questions, tests, seminar presentations, and quizzes.

However, other authentic and comprehensive assessment tools, such as portfolios, practical work, artefacts, and peer and self-assessment, are used intermittently [28, 54]. Formative assessment practices differ from one instructor to another because of the modes of assessment outlined in the course outlines and the lack of moderation [52]. Thus, each instructor has autonomy over what to include in the assessment tool and how to conduct it [35].

Summative assessment refers to formal assessment conducted at the end of an educational activity to provide feedback that summarises the teaching and learning process at a particular level [63]. Summative assessment at the university level is conducted at the end of each semester, during which university-wide examinations are held. In this category, strict procedures are followed from planning and construction through moderation and administration to the release of results. The aim is to ensure fairness and justice are maintained. However, traditional assessments (paper-and-pencil) in the form of essay, multiple-choice, matching, and short-answer items are frequently practised in universities in Tanzania [52].

1.1. Assessment literacy framework

Assessment literacy is increasingly recognised as a multidimensional construct that extends beyond basic knowledge of testing procedures [55]. It encompasses a coherent integration of testing literacy, measurement literacy, and data literacy [16], each contributing essential knowledge and skills required for effective assessment practice. Assessment literacy can be defined as the ability of instructors to understand the meaning, forms, aims, tactics, and techniques of assessment and use them appropriately [20]. Assessment literacy, according to Michigan Assessment Consortium [37], encompasses a range of attitudes, practices, and knowledge that help instructors and other stakeholders use assessment to improve student learning and performance.

1.1.1. Components of assessment literacy framework

Assessment literacy is composed of three key interrelated competencies:

1. *Testing literacy* refers to instructors' understanding of the principles of test design, development, and use [16]. This includes the ability to construct, select, and apply both formal and informal assessment tools to gather evidence of student learning [29, 51]. For instructors, testing literacy is critical in ensuring alignment between learning outcomes, instructional activities, and assessment tasks [9]. It also supports the appropriate use of diverse assessment methods, such as classroom questioning, observations, quizzes, and performance-based assessments, rather than relying solely on standardised tests [2]. Strong testing literacy enables instructors to design assessments that are instructionally meaningful and supportive of learning.
2. *Measurement literacy* forms a second essential pillar of instructors' assessment literacy. It involves a practical understanding of core measurement principles, particularly validity, reliability, and measurement error [9, 36]. Instructors with measurement literacy are better positioned to interpret assessment results

appropriately and to recognise the limitations and uncertainty inherent in all measurement processes. Rather than requiring advanced statistical expertise, measurement literacy emphasises applied judgment, enabling instructors to evaluate the quality of assessment evidence and avoid misinterpretation or misuse of scores [10, 29, 49, 51]. This competence is essential for making fair, defensible instructional and evaluative decisions.

3. *Data literacy* refers to the capability to manage, understand, and assess extensive data in today's technologically advanced world of interconnected information [4]. Data literacy enhances instructors' assessment literacy by equipping them with the skills to organise, manipulate, analyse, and interpret assessment data to enhance instruction. As the use of assessment data becomes more prevalent in educational settings, instructors must be able to manage data through tasks such as sorting, filtering, combining data sources, and generating reports [31, 32]. Within the context of assessment literacy, data literacy focuses specifically on transforming assessment data into actionable information that informs teaching and learning [15, 32]. These skills enable instructors to identify patterns in student performance, monitor progress, and adjust instructional strategies accordingly [4].

Although testing, measurement, and data literacies are conceptually distinct, they are interdependent in practice. Effective assessment literacy among instructors emerges from integrating these three domains rather than their isolated application [31, 37, 58]. Together, they enable instructors to move beyond administering assessments toward using assessment as a tool for learning, reflection, and continuous improvement, thereby enhancing the quality and equity of educational practice [6, 7, 30].

1.1.2. Conceptual framework of instructors' assessment literacy

The framework illustrates instructors' assessment literacy as a central construct supported by three interrelated competencies: testing literacy, measurement literacy, and data literacy (figure 1). The interaction of these competencies enables effective assessment design, interpretation, and data use to support instructional decision-making and student learning.

1.2. Assessment standards

Instructors are required to adhere to the following assessment standards:

1. *Competence in choosing assessment methods* amongst university instructors is essential for the standardisation of items. Omo-Egbekuse, Afemikhe and Imobekhai [45] assert that instructors need to be fully competent in assessment construction and selection of items. Any chosen assessment methods need to be valid in whatever way they are used. Rezvani Kalajahi and Abdullah [51] assert that lecturers should be able to understand the strengths and weaknesses of each assessment method used to measure students' understanding. In the same regard, Kitta [26] emphasises that understanding the pitfalls of the method will enable lecturers to set test/examination items within their limits. They can also provide feedback using familiar methods to assist their students. Rezvani Kalajahi and Abdullah [51] observed that by choosing an appropriate assessment method, a lecturer can diagnose students' learning difficulties, identify misconceptions, and correct them effectively.
2. *Competence in developing assessment methods* is another important standard for assessment practices. According to Rezvani Kalajahi and Abdullah [51], lecturers should have assessment methods skills appropriate for instructional decisions.

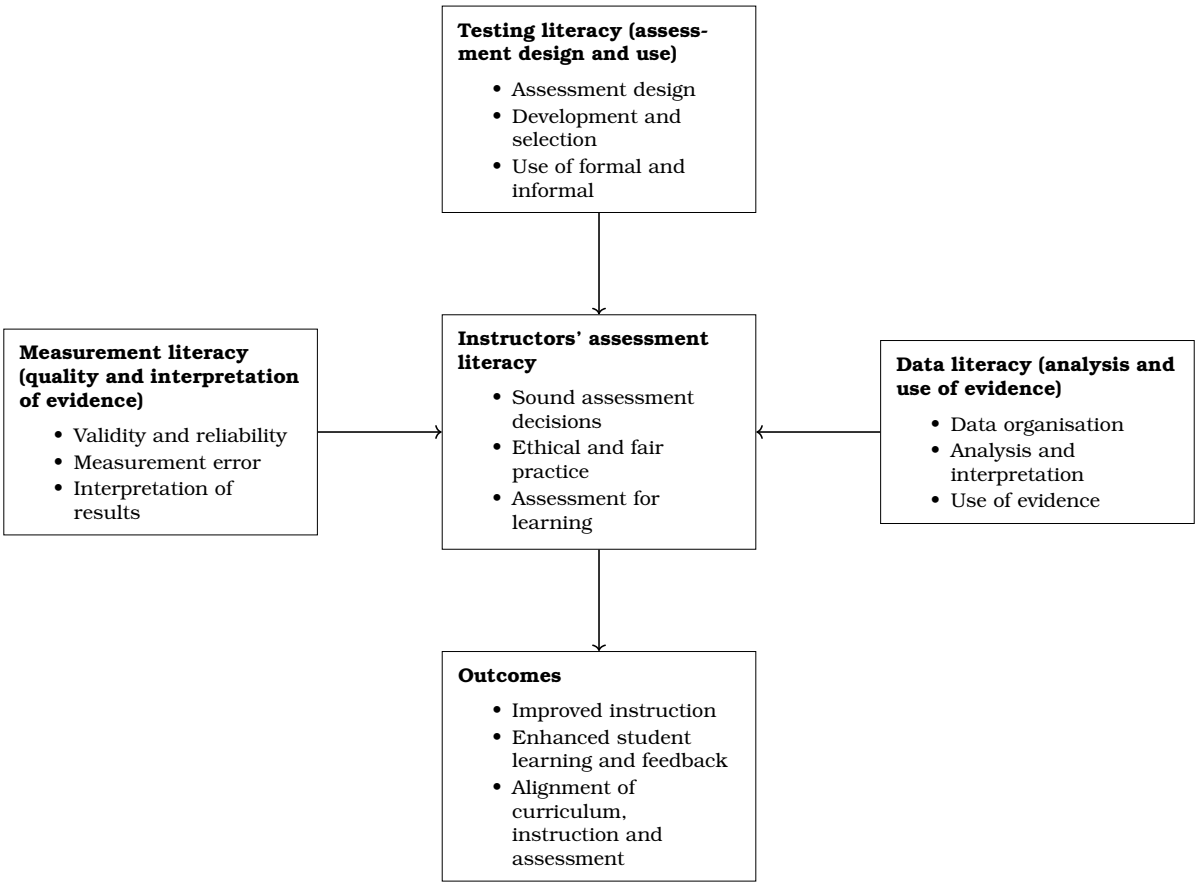


Figure 1: Conceptual framework of instructors' assessment literacy.

In the same regard, Kitula, Kireti and Wambiya [27] argue that developing an effective assessment method will allow the lecturer to plan teaching effectively and use a variety of teaching strategies in relation to the assessment method. These scholars add that decisions lecturers make should be guided by the approaches they develop and, therefore, implement [27]. For example, Waugh and Gronlund [63] states that, in developing assessment methods, the lecturer must consider the course's learning outcomes, the level of difficulty, and the discrimination that can be used to determine each student's ability.

3. *Competence in administering, scoring and interpreting the results.* The administration, scoring, and interpretation of results are of paramount importance to ensure valid and reliable results. Waugh and Gronlund [63] suggest that a sound educational assessment in administering, scoring and interpreting the results requires a clear conception of all intended learning outcomes of the instruction and a variety of assessment procedures that are relevant to the instruction, adequate to sample student performance, and fair to everyone. Miller, Linn and Gronlund [38] emphasise that during the assessment process, the instructor must ensure that the instructional plans align with the assessment. Therefore, the use of assessment should not be taken for granted, as assessment results directly influence students.
4. *Competence in using assessment results for decision making.* The instructor should be able to use the assessment results to make decisions about individual students, planning and teaching, developing curriculum and school improvement. According to Waugh and Gronlund [63], assessment results help lecturers to obtain feedback on the general quality of instruction and identify areas of

strength and weakness. When using assessment results to make decisions about individual students, the teacher/lecturer may use formative, diagnostic, and summative assessment tools. Using the assessment results, the instructor will be better positioned to determine students' needs and adjust the teaching accordingly. To avoid misuse of the assessment, instructors are also required to skillfully interpret test scores and apply them to the assessment's intended purpose, ensuring the results are meaningful. Otherwise, the instructor may use assessment results to harm students rather than help them.

5. *Competence in developing valid grading procedures*. Lecturers need to be skilled in developing valid grading procedures that use students' assessments. Generally, grades are determined by competence and competition: individuals compete on an equal basis to demonstrate their claim to competence [48]. Sales [53] argues that grading is of major importance to educational practice and society as it plays a gate-keeping role, opening or closing doors for individuals in the labour market. This is the meritocratic basis of our modern society that allows free competition based on academic ability. Furthermore, *Alternative Assessment in Higher Education: A Practical Guide to Assessing Learning* [2], Kimaro and Kapinga [25] explain that lecturers are required to be skilled in developing valid grading procedures that use students' assessments. Lecturers should be aware of the assessment principles that underpin valid grading. Therefore, in grading students' performance, the lecturer will not only evaluate his/her teaching but also their attainment. Given such a situation, the grading procedures should not be used for fault-finding but rather to promote the lecturer-student relationship, which, in turn, may lead to effective learning.
6. *Competence in recognising unethical, illegal and inappropriate assessment methods*. Lecturers need to be skilled at recognising unethical, illegal, and otherwise inappropriate assessment methods and the use of assessment information. Zembazemba [66] states that a lecturer should be able to follow assessment procedures and principles. He/she should also be fair in communicating students' results and protecting their rights. The skill in maintaining moral values not only promotes lecturers' integrity but also encourages students to be more confident during the teaching and learning process and to enjoy learning. Students are human beings indeed; therefore, lecturers need to have a sense of humour when dealing with their academic matters. Therefore, in reporting students' results, lecturers should maintain confidentiality. The literature shows that instructors need not be experts in educational measurement and evaluation to construct valid and reliable tests, but there are basic test-construction skills that every instructor ought to possess to construct high-quality tests [27, 66]. These skills help instructors draw inferences from students' responses, and, in judging students' results, fairness and objectivity are to be observed.

Despite these standards, several systemic factors limit instructors' ability to engage in effective assessment and improve their competencies, as follows.

One of the key limitations arises from inadequate policy frameworks and weak institutional guidelines, which often fail to provide coherent standards for effective assessment practice [59]. In many developing contexts, national and institutional assessment policies are fragmented or poorly implemented, leaving instructors without consistent direction or support.

Another constraint is limited leadership commitment and governance. Strong institutional leadership is essential for creating a culture that values assessment for learning, supports professional development, and provides strategic direction for teaching and assessment improvement [19]. When institutional leaders neglect

assessment capacity in strategic planning and budgeting, instructors receive little encouragement or accountability to improve their competencies. This governance gap reduces institutional ownership of assessment quality, resulting in ad hoc or compliance-based practices rather than meaningful learning assessment [17].

Professional development opportunities are often inadequate. Effective assessment relies on understanding feedback, aligning learning outcomes with assessment tasks, and applying formative assessment strategies. However, many institutions fail to provide structured, ongoing training in these critical areas [10]. Furthermore, the absence of mentorship or peer-support systems limits opportunities for instructors to reflect on and refine their assessment practices [21, 40, 42].

Curriculum misalignment presents another systemic challenge. Biggs and Tang [5] emphasise the importance of constructive alignment, where teaching, learning activities, and assessment are coherently linked. When learning outcomes, teaching strategies, and assessments are disconnected, instructors tend to rely on traditional examinations instead of more authentic or competency-based approaches.

Resource constraints, including limited funding, inadequate technological tools, and insufficient access to digital platforms, further restrict instructors' ability to implement modern assessment approaches [60]. In resource-limited settings, instructors often rely on traditional examinations rather than authentic or performance-based assessments that require more time, training, and institutional support [44].

Institutional culture strongly influences how assessment is perceived and used. If assessment is viewed primarily as a grading mechanism rather than a learning tool, instructors may not prioritise feedback or reflective assessment practices. Studies emphasise the importance of supportive feedback cultures that promote learning-oriented assessment [8, 11]. Weak monitoring and feedback mechanisms mean that instructors rarely receive constructive input on their assessment practices, reducing motivation for continuous improvement [11]. In general, these systemic factors create an environment that hinders effective assessment practices, thereby hindering the teaching and learning process.

Several studies, including Agu, Onyekuba and Anyichie [1], Chalchisa [12], Istoroyekti [23], McCallum and Milner [35], Ogula and Onsongo [43], Srivastava and Kumari [57], have demonstrated that assessment is a key determinant of effective teaching and learning. However, there are limited empirical studies that explain instructors' self-perceived assessment competencies and practices in universities in Tanzania. This situation posed questions such as, "What are the dimensions of instructors' assessment competence in universities? Do assessment practices differ from instructors' demographic characteristics?" It is evident that unless university instructors have a clear understanding of the principles guiding the construction of quality assessment items, they cannot construct valid and reliable assessments [44]. Hence, the current study aimed to examine whether instructors' self-perceived assessment competencies were reflected in their test construction practices, that is, how competent instructors are at constructing valid and reliable tests, regardless of the existing limitations, and to what extent instructors are capable of constructing tests that adhere to assessment standards. Furthermore, the study hypothesised whether there is a significant difference between demographic parameters and indicators of instructors' self-perceived assessment competencies that adhere to assessment standards in Tanzanian universities. Therefore, based on the given explanations and the problem under study, this study was conducted to inform education stakeholders, employers, policymakers, quality assurance officers, and university instructors of the importance of instructors adhering to assessment standards to improve teaching and learning.

2. Methodological issues

The data of this study were collected from three universities: Mzumbe University, Sokoine University and University of Dar es Salaam, subject to the given approval from the University of Dodoma (UDOM/GR/209/VOL.1/37), Mzumbe University (MU/R.2/1/Vol.11/174), Sokoine University of Agriculture (SUA/ADM/R.1/8/VOL.IV/2140 and University of Dar es Salaam (AB3/31). Furthermore, an informed consent form was provided to the respondents to ensure their voluntary participation. The selection of the three universities was based on the accessibility of data collection, as these universities have instructors with extensive teaching experience, the focus of this study. The study used a mixed-methods approach informed by both ex-post facto and transcendental phenomenological designs. Henceforth, the quantitative approach involved an ex post facto design, and the qualitative approach involved a transcendental-phenomenological design. The study adopted an ex post facto research design because instructors' assessment competencies and practices already exist and cannot be manipulated. This design allowed the investigation of university lecturers' self-perceived assessment competencies and practices in Tanzanian universities, focusing on how their experiences, qualifications, and institutional contexts influence compliance with assessment standards, without interfering with actual teaching processes. It was also important to employ a transcendental-phenomenological design, as it provides the researcher with a greater opportunity to understand instructors' lived experiences of their self-perceived assessment practices. The study believed that instructors play multiple roles in assessing students' activities, from designing coursework to releasing results, for example, in university examinations. Therefore, interviewing the instructors would yield sufficient data that might be useful for improving examination practices in Tanzanian Universities.

Yamane's [64] formula for a known population $n = \frac{N}{1+N(e)^2}$ was employed ($N = 1864$), followed by stratified proportional sampling and simple random sampling techniques to select and generate data from a sample of 205 instructors. These included professors, senior lecturers, lecturers, and assistant lecturers. Heads of department and quality assurance coordinators were purposively selected. These informants play a crucial role in conducting assessments effectively and assessing instructors' compliance with assessment standards. Furthermore, these participants were selected because they live the reality being studied; they observe both formative and summative assessment practices by lecturers, they see the challenges instructors face during assessment, they experience the process, and they understand what works and what does not. Their first-hand perspectives enriched, clarified, and grounded the findings in real classroom practice. The study employed questionnaires, interviews, observations and document review to collect data.

The researcher ensured that the validity and reliability of the data collected were maintained. The questionnaire on self-perceived test construction practices was validated through expert appraisal and field testing by experts in educational research and measurement, assessment, and evaluation. The assessment included relevance, content coverage, appropriateness of the items, and the language and readability of the items. These experts rated the questionnaire and the structured interview guide. Later, the Pearson product-moment correlation coefficient (r) was used to assess whether the items on the questionnaires represented the content domains, how clear they were, and the extent to which they maintained the theoretical factor structure, as assessed by factor analysis. Thus, the PPMCC value obtained for test construction practices was 0.83, indicating a high correlation.

Cronbach's α was used to assess the internal consistency of the items and to

determine whether all items in the instrument measured the same phenomenon. The Cronbach's α value for assessment practices was 0.879, indicating a high level of internal consistency for the scale, as it exceeds 0.7 [13, 14]. Table 1 shows Cronbach's α of each item and their descriptive statistics (means and standard deviations).

Table 1
Correlation of self-perceived test construction practices.

S/n	Test construction practices	Mean	Standard deviation	Cronbach's alpha if item deleted
1	I use Bloom's taxonomy to construct test items for measuring students' all-around development.	3.52	1.013	.873
2	I analyse students' performance in every test, quizzes or project works which I give them.	3.59	1.052	.873
3	I evaluate instructional procedures.	3.51	1.187	.875
4	I make an effort to motivate students who show low interest in my course.	3.48	1.235	.876
5	I determine appropriateness of assessment procedure.	3.68	1.077	.877
6	I interpret informal assessment results.	3.90	1.096	.874
7	I use the marking scheme to mark subjective question items.	3.64	1.170	.873
8	I modify grading procedures to improve confidence in interpretations.	3.64	1.083	.874
9	I avoid faulty grading procedures.	3.53	1.069	.877
10	I do construct valid test items.	3.41	1.158	.874
11	I interpret formal teacher-produced assessment results.	3.48	1.008	.875
12	I use standardised tests to assess students.	3.30	1.064	.877
13	I use a summary index, such as mean and standard deviation, to interpret students' performance.	3.59	1.089	.875
14	I can analyse test item difficulty and discrimination.	3.47	1.109	.872
15	I do construct the table of specification.	3.62	1.080	.871
16	I am fair and flexible in judging students' works according to their nature.	3.41	1.179	.874
17	I use assessment results for improving teaching.	3.63	1.057	.873
18	I consider students' strengths and weaknesses when evaluating their overall performance.	3.55	1.173	.875
19	I make efforts to adjust assessment strategies to the proper level for individual students.	3.43	1.193	.873
20	I do interpret assessment results correctly to plan instruction and curriculum.	3.66	1.102	.875
21	I use accumulated assessment information to organise an instructional plan.	3.47	1.105	.873
22	I do explain results to students.	3.43	1.160	.872
23	I communicate with students and other educational stakeholders about students' progress.	3.53	1.064	.877
24	I do inform students about the objectives and procedures of evaluation.	3.60	1.023	.877
25	I make efforts to explain why the grades assigned are rational and justified.	3.60	1.087	.877

Furthermore, to ensure validity and reliability, the current study underwent pilot testing at one of the universities in Tanzania, and the data from this testing were not used in the final analysis. The university where the pilot study was conducted had

characteristics similar to those of other universities in Tanzania – the pilot study aimed to assess the relevance, readability, suitability, and applicability of the approach. Thus, the pilot study helped confirm the research design, sample size, and the instrument used for data collection in relation to the study's problem and research objectives. Modifications were made based on suggestions from the pilot study.

Additionally, prior to data analysis, data cleaning was performed to identify and remove errors and extreme outliers that could mislead the study's findings and reduce the study's credibility and validity. Various data-cleaning methods were used. These included human judgement, member checking (informant feedback), and possible code cleaning, and examining the distributions of responses to each data set. Human judgment involved removing data that were not required for the study and retaining the most focused responses. For example, some items in the questionnaires were partly skipped by the respondent. In that case, these questionnaires were not included during data entry and analysis.

Lastly, to ensure the data's validity and reliability, member checking was conducted during each interview. During member checking, the researcher summarised the data and gave the informants the chance to correct or improve whatever they had explained. What appeared to be a common agreement was then taken for data analysis. Possible code cleaning was performed to remove codes that were not assigned during data coding to avoid errors and different meanings during data analysis and interpretation. This allowed the data file to remain with the assigned codes for the choices. In examining the data set, frequency distribution tables for demographic parameters were generated and examined to assess the distribution of the data set. The frequency tables, which indicated wrong entered values, were corrected.

After data cleaning, the qualitative data were subjected to interpretation analysis. Creswell [14] claims that the interpretive analysis involves a systematic set of procedures for coding and classifying qualitative data to ensure that important constructs, such as themes and sub-themes, emerge. Data from observations, document analysis, and interviews were coded to identify themes and sub-themes. It was through repeated ideas that themes and sub-themes emerged. Themes were derived from the study's main objectives, and sub-themes were developed inductively from a critical analysis of the gathered data.

Besides, quantitative data from questionnaires were categorised, coded, and entered into the computer for the computation of descriptive and inferential statistics. The study employed exploratory factor analysis (EFA) to determine the factor loadings. Data compliance for factor analysis was assessed using the Kaiser-Meyer-Olkin measure of sampling adequacy, Bartlett's test, and inter-variable correlations. Table 2 presents the results of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity.

Table 2
Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test.

Test	**Value**
Kaiser-Meyer-Olkin measure of sampling adequacy	0.842
Bartlett's test of sphericity	
χ^2	336.020
df	28
p	<0.001

The obtained KMO value was 0.842, which exceeds the cut-off of 0.5. This means that the sample is adequate for factor analysis. Similarly, the recorded significant

p-value for Bartlett's test (< 0.001) indicates that the original R-matrix is significantly different from an identity matrix. These findings suggest that there are correlations between test variables measuring instructors' perceived assessment competencies, that the data are suitable for factor analysis, and that the instrument could be used to determine the content and construct validity.

2.1. Communalities after extraction

25 items were used to measure test construction practice. However, during the extraction of the factor analysis, 12 items were removed. Among the 12 items, seven (7) items were removed as they belonged to more than one component (complex structure). This included: the extent to which "I can construct free-error test items"; the degree to which "I can use scales to rate students' performance"; and the extent to which "I can be fair and flexible in judging students' work according to its nature". The amount of effort "I make to motivate students who show a low interest in my course", the level at which 'I can determine the appropriateness of assessment procedure", and the amount "I can use assessment results for improving teaching".

The other five (5) items were removed from the analysis as they had communality of less than 0.5 as follows; the extent to which I can give oral feedback to students, the extent to which I can interpret formal teacher-produced assessment results, the extent to which I can use a summary index such as means standard deviation to interpret students' performance, the extent to which I consider the strengths and weaknesses of the students while evaluating their overall performance, and how much effort I can give written feedback to students.

Table 3 provides the communalities of the remaining items after extraction. The communalities ranged from 0.561 to 0.730, all above the cut-off point of 0.5.

Table 3
Communalities after extraction.

S/n	Item	Initial	Extraction
1	I avoid faulty grading procedures.	1	0.654
2	I do construct valid test items.	1	0.561
3	I do analyse students' performance in every test, quiz or project work which I give them.	1	0.585
4	I do explain to students about the results.	1	0.715
5	I use accumulated assessment information to organise the instructional plan.	1	0.575
6	I communicate with students and other educational stakeholders about students' progress.	1	0.678
7	I do construct the table of specifications.	1	0.684
8	I use Bloom's taxonomy to construct test items for measuring students' all-around development.	1	0.584
9	I do interpret assessment results correctly to plan instruction and curriculum.	1	0.730
10	I do make efforts to adjust assessment strategies to the proper level for individual students.	1	0.587
11	I do modify grading procedures to improve confidence in interpretations.	1	0.682
12	I do inform students about the objectives and procedures of evaluation.	1	0.674
13	I make efforts to explain why the grades assigned are rational and justified.	1	0.723

2.2. Factor extraction

The eigenvalues associated with each factor before extraction and after rotation are presented in table 4.

Table 4
Total variance explained by extracted factors.

Comp.	Initial eigenvalues			Extraction sums			Rotation sums		
	Total	% Var.	Cum. %	Total	% Var.	Cum. %	Total	% Var.	Cum. %
1	2.896	36.198	36.198	2.896	36.198	36.198	1.996	24.955	24.955
2	1.276	15.949	52.147	1.276	15.949	52.147	1.613	20.163	45.118
3	1.015	12.690	64.837	1.015	12.690	64.837	1.577	19.719	64.837
4	0.695	8.689	73.526	0.695	8.689	73.526	1.321	8.689	73.526
5	0.598	4.246	77.772						
6	0.495	3.689	81.461						
7	0.461	3.261	84.722						
8	0.451	3.195	87.917						
9	0.324	2.911	90.828						
10	0.292	2.724	93.552						
11	0.285	2.516	96.068						
12	0.202	2.457	98.525						
13	0.169	1.475	100.000						

Initially, the analysis identified 13 linear components in the dataset. The analysis then extracted factors with eigenvalues of 1.0 or higher based on Kaiser's recommendation ([18], [47, p. 184]). The unrotated factor solution retained four (4) factors, which explained 73.526 per cent of the total variance. The first factor accounted for about 36.198%. The second factor explained 15.949% of the total variance, the third factor accounted for 12.690%, and the fourth factor explained 8.689% of the variance in the dataset. Before rotation, factor 1 accounted for considerably more variance (36.198%) compared to the second factor (15.949%), the third factor (12.690%) and the fourth factor (8.689%). However, after rotation, factor 1 accounted for 24.955% of variance, which is not very different from that of the second factor (20.163%), and the third factor (19.719%) accounted for 8.689%. This implies that the relative importance of the three factors has been equalised after rotation.

2.3. Indicators of test construction practices

The results of the factor analysis for test construction practice revealed that four (4) factors can represent the remaining 13 variables. Table 5 presents the factor loadings of the obtained factors and their respective indicators after varimax rotation.

Table 5 indicates that factor 1 consisted of three indicators: "I do construct the table of specification", "I do construct valid test items", and "I use Bloom's taxonomy to construct test items for measuring students' all-around development". Factor 2 had two ways: "I analyse students' performance in every test, quizzes or project works which I give them", and "I modify grading procedures to improve confidence in interpretations". Factor 3 consisted of four ways: "I avoid faulty grading procedures", "I make efforts to adjust assessment strategies to the proper level for individual students", "I do interpret assessment results correctly to plan instruction and curriculum", and "I use accumulated assessment information to organise instructional plan" and the last factor consisted of four items: "I explain to students about results", "I communicate to students and other educational stakeholders about students' progress", "I do inform students about the objectives and procedures of evaluation", "I make efforts to explain why grades assigned are rational and justified".

Table 5
Pattern matrix for exploratory factor analysis after varimax rotation.

S/n	Indicators of test construction practices	Components			
		1	2	3	4
1	I do construct the table of specification.	0.723			
2	I do construct valid test items.	0.783			
3	I use Bloom's taxonomy to construct test items for measuring students' all-around development.	0.806			
4	I analyse students' performance in every test, quizzes or project works which I give them.		0.791		
5	I modify grading procedures to improve confidence in interpretations.		0.654		
6	I avoid faulty grading procedures.			0.817	
7	I make efforts to adjust assessment strategies to the proper level for individual students.			0.612	
8	I do interpret assessment results correctly to plan instruction and curriculum.			0.811	
9	I use accumulated assessment information to organise the instructional plan.			0.715	
10	I explain to students about the results.				0.723
11	I communicate with students and other educational stakeholders about students' progress.				0.685
12	I do inform students about the objectives and procedures of evaluation.				0.814
13	I make efforts to explain why the grades assigned are rational and justified.				0.752

The Statistical Package for the Social Sciences (SPSS Statistics, version 20) was used to run descriptive analyses to produce means and standard deviations for various characteristics of the respondents. Data were presented using means and standard deviations. Inferential statistics, specifically logistic regression, were used to assess the association between different levels of assessment practices and demographic parameters, as shown in the following equation.

$$P(Y=1|Z) = \frac{e^{\beta_0+\beta_1 Z_1+\beta_2 Z_2+\ldots+\beta_k Z_k}}{1+e^{\beta_0+\beta_1 Z_1+\beta_2 Z_2+\ldots+\beta_k Z_k}} \quad (1)$$

where $P(Y=1|Z)$ is the chance of having a high perceived assessment competency score for a subject with Z demographic parameters. The confidence level was set at 95% and the significance level at 0.05.

3. Findings and discussion

The study findings have been explained by focusing on the key objectives.

3.1. Indicators of instructors' self-perceived assessment competencies

In summary, the following are indicators of self-perceived assessment competencies and practices demonstrated by university instructors, as revealed in the figures 2, 3, 4, and 5. In this study, a cut-off point of 4 was adopted to interpret self-perceived assessment competence. Scores of 4 or higher indicate high competence, distinguishing lecturers who demonstrate meaningful proficiency in assessment practices from those with moderate or lower competence. This approach was intended to minimise discrepancies that may arise during interpretation. Furthermore, categorising competence in

this way highlights areas that may benefit from further development, consistent with the study's focus on evaluation.

The analysis in figure 2 presents the indicators of instructors' competence in communicating assessment procedures and results. The findings reveal that instructors demonstrated low competence in almost all indicators that measure competence in test/examination results instructors' communicating and assessment procedures to students (i.e. "I make efforts to explain why grades assigned are rational and justified" ($\bar{x}$=3.72), "I do explain to students about results" ($\bar{x}$=3.48), "I do communicate to students and other educational stakeholders about students' progress" ($\bar{x}$=3.46), and "I do inform students about the objectives and procedures of evaluation" ($\bar{x}$=3.32). Assessment results should be communicated effectively and routinely after a given task. Students were generally eager to know their performance levels, as were parents and other audiences who wanted to know the status of their children. If the assessment results are not communicated effectively, the information obtained may be distorted and misused [34]. The effective communication of students' results not only improves students' learning and performance but also promotes teaching effectiveness, as the instructor will be in a position to identify what remained unclear in their course content. Besides, student feedback is important for satisfaction and change [48].

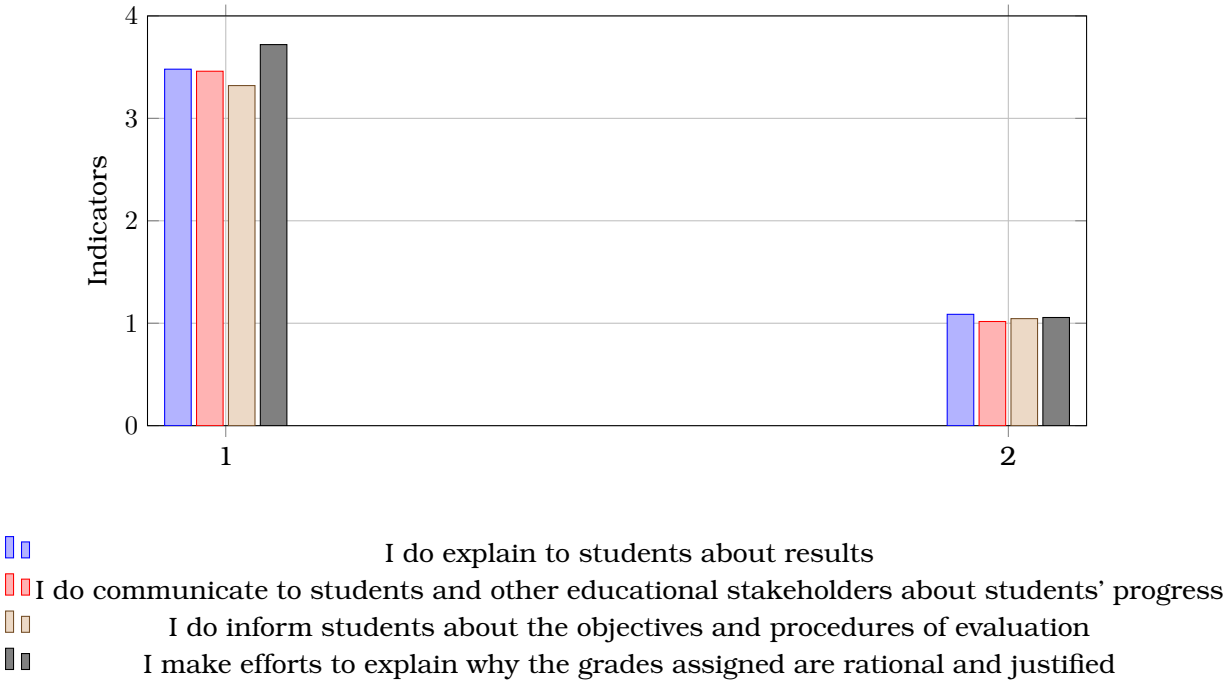
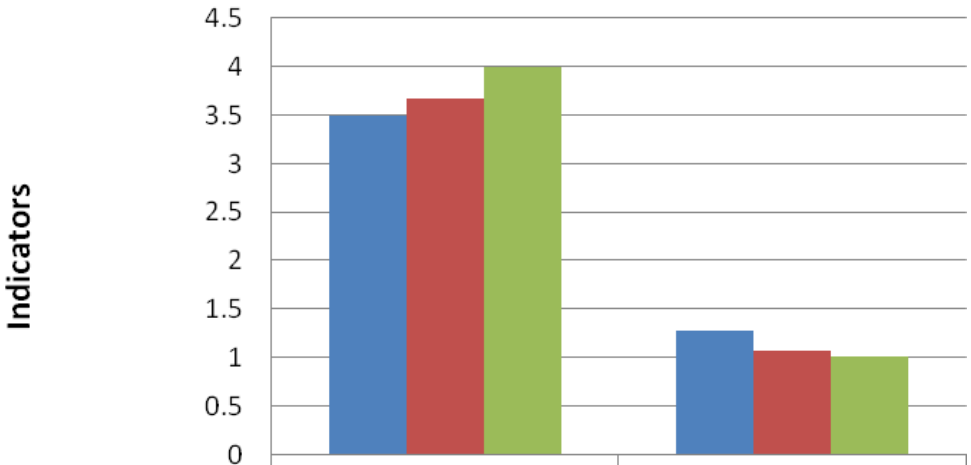


Figure 2: Indicators of instructors' competence in communicating assessment procedures and results: mean (1) and standard deviation (2).

The analysis in figure 3 presents the indicators for measuring instructors' competence in analysing and interpreting test/examination results. The findings indicate that instructors demonstrate high analysis and competence in interpreting test/examination results in one item of avoiding faulty grading procedures ($\bar{x}$=4.00). The ability to modify grading confidence procedures to improve interpretations was found to be low ($\bar{x}$=3.68), and instructors' ability in analysing students' performance in every test, quizzes or project works which they give students ($\bar{x}$=3.49). These findings reflect the institutional culture and poor feedback systems. In many universities, assessment is viewed primarily as a grading tool rather than a process to enhance learning [8]. Weak monitoring and feedback mechanisms, which are rarely provided to



	1	2
I analyses students' performance in every test, quizzes or project works which I give them	3.49	1.29
I modify grading procedures to improve confidence in interpretations	3.68	1.076
I avoid faulty grading procedures.	4	1.017

Figure 3: Indicators for measuring instructors' competency in analysing and interpreting test/examination results: mean (1) and standard deviation (2).

instructors, reduce motivation for continuous improvement [11].

The data suggest that instructors be aware of the assessment principles that support valid grading. Besides, in grading students' performance, an instructor will not only evaluate his/her teaching but also their attainment. Given such a situation, the grading procedures should not be used for fault-finding, but rather to promote teacher-student relationships that, in turn, may lead to effective learning.

The data in figure 4 present indicators of instructors' competence in using test results to plan and organise classroom instruction. The findings reveal that almost all indicators for measuring competency in using test results to plan and organise classroom instruction were less practised by instructors. Their mean scores are as follows: "I make efforts to adjust assessment strategies to the proper level for individual students" ($\bar{x}$=3.75), "I do interpret assessment results correctly to plan instruction and curriculum" ($\bar{x}$=3.56), and "I do use accumulated assessment information to organise instructional plan" ($\bar{x}$=3.49). Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching based on learning outcomes, developing curriculum, and supporting school improvement. Failure to use assessment results may lead to wrong interpretations of educational outcomes and students' learning. These findings concur with Hamafyelto, Hamman-Tukur and Hamafyelto [22] in their study on assessing teacher competence in test construction and the content validity of teacher-made examination questions in commerce in Borno State, Nigeria. The findings revealed that teachers were not competent in developing valid student grading procedures, communicating results, and recognising unethical and illegal assessment methods.

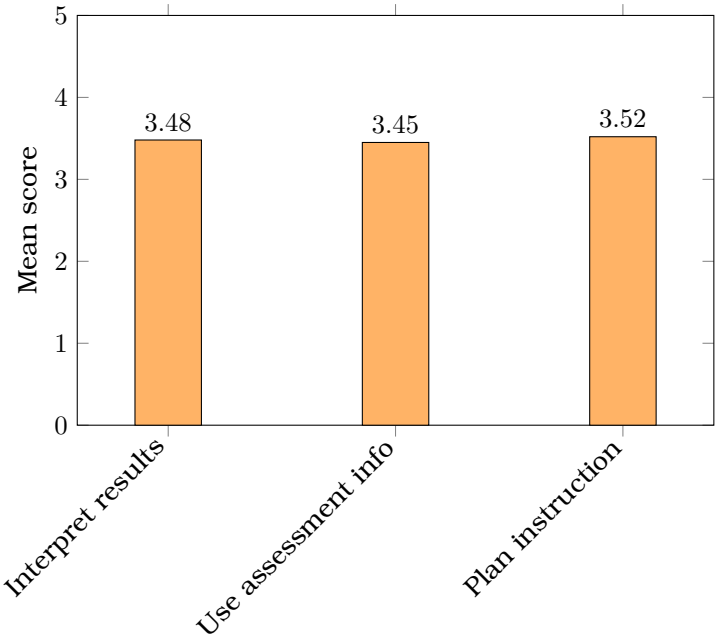


Figure 4: Indicators for measuring instructors' competency in using test results in planning and organising classroom instruction.

The analysis in figure 5 presents indicators for measuring knowledge of test/examination procedures and principles of test construction. The findings show that instructors had low competence in constructing valid test items ($\bar{x}$=3.57), constructing the table of specification ($\bar{x}$ =3.55), and using Bloom's taxonomy to construct test items for measuring students' all-around development ($\bar{x}$=3.51). These findings reflect inadequate policy frameworks and weak institutional guidelines, which often fail to provide coherent standards for effective assessment practice [59]. Anderson et al. [3] suggest that, in constructing the test, instructors need to be aware and competent in the procedures and their principles. They emphasise the use of the table of specification and Bloom's taxonomy during test construction to construct a valid test. Furthermore, any assessment should not only measure the content but also the competence level to determine students' understanding and application of the learned materials in different environments.

Quantitative data were also supported by qualitative data from interviews, document reviews, and observations, as summarised in table 6.

The analysis in table 6 presents self-perceived assessment competencies and practices as revealed by informants during interviews. The findings portray that instructors were not using the table of specifications to construct their test/examination items. Most of the interviewed instructors reported having limited time or being unaware of what is called a table of specifications. During interviews, one instructor who was an expert in the subject and had not attended any teaching training said:

> The thing you are saying is new to me. Actually, I do not know it, to say the truth. I may explain it when you give me some clues. I have never used it, madam, I do not want to tell lies to you that I am using it... (Interview with lecturer 1)

Responding to the same question, another instructor who is also the head of the department confirmed that:

> I know the table of specifications because I attended the University's wide

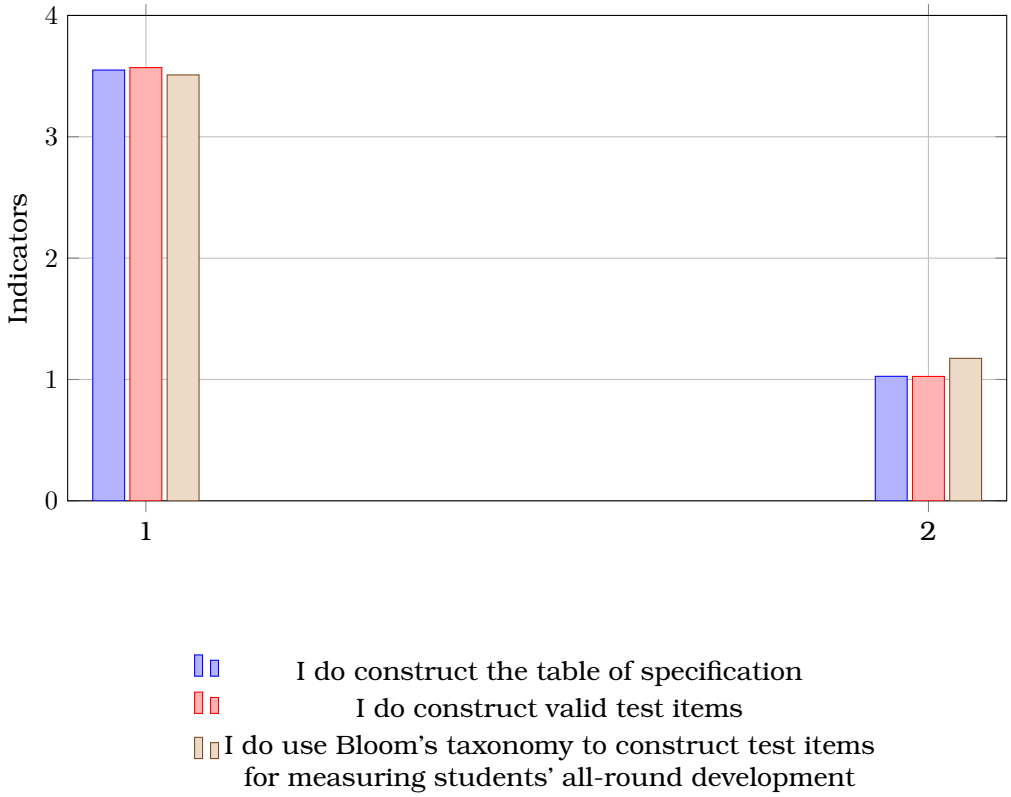


Figure 5: Indicators for measuring knowledge of test/examination procedures and principles of test construction: mean (1) and standard deviation (2).

Table 6
Self-perceived assessment competencies and practices by sampled university instructors,

Subtheme	Important quotes
Using the table of specifications	"I don't use the table of specifications because I don't know it well; it is my first time hearing about it."
Steps followed in constructing tests/exams	"First, I consider content coverage and then construct the test that measures all levels."
Considering students' characteristics	"I only take into consideration the learners with special needs, such as the blind ones."
Ways instructors ensure that their tests are valid and reliable	"I ensure that my test is reliable and valid by asking questions somehow different from what I have taught. For example, if I teach about crop pests, I will ask students how crop pests affect production."
Procedures followed in test item construction	"During the test, I just list down things I need to assess and use the list to construct questions."
Information instructors specify in the examination	"I tell my students the number of questions they should respond to and the time they should use."

programme called UTLIP (University Teaching and Learning Improvement Programme). The programme has been focusing on improving lecturers' skills in teaching, assessment, and evaluation of the teaching activities. Test and examination constructions are key emphases in that training … but I don't use them in constructing test items. This is the truth… I have my own ways of setting questions in the test. (Interview with senior lecturer 3)

From these two quotes, one can draw the following conclusions. First, some of the instructors had a limited understanding of what the table of specifications was and, hence, were not using it in test/examination construction. Second, although some of the instructors had attended training on student assessment and knew what a table of specifications was, they did not use it during test/examination item construction.

External examiners and internal quality assurance officers also considered the table of specifications as an instructor's blueprint for constructing test/examination items. The curriculum's internal quality assurance officers regarded it as the yardstick for test/exam item construction, which, therefore, could help instructors ensure that all cognitive levels were fairly and adequately assessed. Regarding this, one quality assurance coordinator who had an educational background had this to say:

> Using the table of specifications is very important in constructing a test/examination because it is a blueprint of what to assess. It helps the instructor ensure that all levels of cognitive domains are assessed. However, the challenge we have in this university is that some instructors are not conversant with it and even those who know it, particularly teachers, rarely or never use it. (Interview with the quality assurance coordinator 3)

This was also supported by pre- and post-modifiers' and examiners' reports from some of the visited universities. Most of the reports reviewed found that moderators and external examiners found that most instructors were not constructing tests in accordance with the table of specifications. They were not even assessing all the cognitive domains as emphasised in Bloom's taxonomy. Generally, the equality and equity in the assessment were overlooked. Given below are some of the comments by external moderators and external examiners in various courses in the respective universities:

Course A: The paper lacks balanced questions to examine high and low levels of knowledge according to Bloom's taxonomy. It also does not cover a broad range of the course content outlined in the course outline. The examination does not discriminate between low and high achievers, which is one of the main objectives of testing and examining. (External examiner 1)

Course B: The paper was set without taking all levels of understanding for the learners into account. Most questions tested higher levels only, but ignored the coverage of the content in the course outline. (External examiner 2)

During the interviews, further elaboration on the use of the table of specifications and Bloom's taxonomy of learning outcomes was provided by one senior lecturer who had been an external examiner for about 5 years and had also served as the head of department. The lecturer argued that some instructors were assessing only the lower cognitive domains, while ignoring the higher-order levels. Others were focusing only on higher cognitive levels, neglecting the lower levels that were equally important.

Consequently, based on the documents reviewed, the researcher came across similar cases regarding how tests and examinations were observed. For example, in one of the courses, she observed a mismatch between what was on the examination paper and what ought to have been taught in the classroom, as outlined in the course. The paper lacked a broad range of question types, from objective to essay. In view of this, the external examiner had these to comment on in one of the courses:

> …the paper uses a lower-level type of question in almost all sections. The use of lower-level test questions denies students the opportunity to practise writing skills, organising ideas, and developing strong arguments and reasoning. (External examiner 2)

This shows that no balance of items has been set for tests/examinations. Consequently, the test/examination items may fail to discriminate between higher- and lower-performing students. A good examination paper ought to consider, among other important aspects, a mixture of examination items that test lower and higher levels of understanding and coverage of the course content that students learned in the classroom. It should be noted that innovative modes of assessment require instructors to move from routine, limited-factual questions to more open-ended questions and problems that involve tasks, which evoke broader discussions and thinking in the classroom. Instructors' assessment practices articulated in this study align with those reported in other studies conducted in various parts of the world. For instance, Punongbayan and Bauyon [50], Sales [53], Thakur and Shekhawat [61] reported that university lecturers mostly assessed lower cognitive skills rather than higher ones, which require innovation and creativity.

It is evident from this study's findings that all students in classes were assessed through the same assessment activities. This means that the university instructors did not consider individual students' knowledge, talents, interests, and age to identify each student's background knowledge, talents, and interests. This is contrary to what Jones et al. [24] suggested, who argued that conducting assessments effectively would help instructors study and understand every student. This is done through a diagnostic test before teaching a new topic so that the teacher can assess each student's background knowledge and experiences with that topic, as well as their interests and talents. This may help the instructor teach and assess every student while understanding each student's capacity and attitudes. Therefore, this can even assign different activities to students depending on their capacity, interests, talents, and attitudes.

3.2. Logistic regression model for demographic parameters associated with indicators of instructors' self-perceived competencies in universities

To determine the level of significance of the comparison between self-perceived assessment competencies and demographic characteristics, a logistic regression model was used, as shown in table 7.

The results of the fitted logistic regression model presented in table 7 show that instructors' self-perceived assessment competencies and practices were significantly associated with teacher by profession ($p = 0.0058$) and teaching experience ($p = 0.0048$). Respondents who were not professional teachers ($OR = 0.45$, $p = 0.0058$) were significantly less likely to have a high level of assessment competence. Regarding teaching experience, respondents with 5–9 years of experience had greater odds of having a high level of assessment competence than those with 0–4 years of experience ($OR = 7.36$, $p = 0.0012$). Similarly, individuals with 10 or more years of experience were significantly more likely to have a high level of test assessment competence than those with 0–4 years of experience ($OR = 4.43$, $p = 0.0104$). The findings of studies conducted in Tanzania and abroad [33, 48, 50, 56] provide adequate support, as they are also consistent with the current study. These studies have shown that the level of instructors' qualifications plays a significant role in enhancing the quality of their assessment practices. Therefore, to be an effective instructor in assessment, the instructor must possess high-level qualifications. As a result, one is likely to satisfy the diverse demands of assessment practices and achieve the best outcomes.

Table 7
Results of logistic regression model for instructors' self-perceived assessment competencies and demographic parameters.

Variables		*OR* (95% CI)	*p*
Sex			0.7217
	Female	Reference	
	Male	0.89 (0.48, 1.65)	
Educational qualifications			0.8230
	Master degree	Reference	
	PhD	0.94 (0.54, 1.64)	
Professional rank			0.8368
	Assistant lecturer	Reference	
	Lecturer	1.26 (0.69, 2.31)	
	Senior lecturer	1.38 (0.53, 3.62)	
	Professor	0.97 (0.30, 3.08)	
A teacher by profession			0.0058
	Yes	Reference	
	No	0.45 (0.25, 0.79)	
Teaching experience			0.0048
	0–4 years	Reference	
	5–9 years	7.36 (2.20, 24.60)	0.0012
	10 and above years	4.42 (1.42, 13.82)	0.0104

Similarly, university lecturers with 10 or more years of experience were significantly more likely to have a high level of assessment competence and practices than those with 0–4 years of experience ($OR = 4.43$, $p = 0.0104$). The findings on this aspect diverge from those of other studies [50, 53, 61]. For example, Nguon [41] conducted a study of assessment practices at a Cambodian University. The study revealed that instructors' experience had little influence on their assessment practices. The study found that even instructors with extensive work experience had difficulty observing assessment principles, while some did not follow the recommended steps necessary for item construction at all. Studies indicate that deficiencies in professional development significantly hinder instructors' assessment competency. Many institutions lack structured, continuous professional learning programmes focused on assessment design, feedback, and data interpretation [10].

Generally, these findings concur with Rezvani Kalajahi and Abdullah [51] in their study on assessing assessment competency and practices among lecturers in Malaysia. The findings revealed that lecturers had low assessment competency across all established areas. The study suggests that, for effective teaching and learning in the classroom, determining the level of assessment literacy and practice among lecturers should always be given top priority.

Moreover, the findings show that communicating assessment procedures and results was being less practised by instructors in the sampled universities. Thus, for effective teaching and learning, instructors are required to provide feedback to students, as students are generally eager to know their performance level. Parents and other audiences would also like to know the status of their children. Effective communication of students' results not only improves their learning and performance but also promotes teaching effectiveness, as the instructor can identify what remains unclear in their course content. If the assessment results are not communicated effectively, the information obtained may be distorted and misused [65]. Besides, one underscores

the importance of feedback to students as it enhances satisfaction and change [48].

Surprisingly, instructors with 5–9 years of experience ($OR = 7.36$) had higher odds than those with 10+ years ($OR = 4.42$) in the assessment literacy. Experienced professionals indeed possess a wealth of knowledge and skills in teaching; however, having a longer tenure in teaching does not guarantee competence in conducting assessments. This may be due to less engagement in professional development activities as experienced lecturers perform other multiple tasks (such as leadership roles, supervision of both master's and PhD students' work, conducting community outreach, consultancies and projects, research works, giving mentorship to juniors in university roles, etc.) and being less motivated in adopting new changes in current assessment methods and technologies. In contrast, mid-careers are readily able to adopt changes and are motivated to learn new assessment strategies and technologies, and to engage in various training as they need to know more about modern assessment techniques and best practices.

Furthermore, the use of assessment results in planning and organising classroom instruction was being less practised by the sampled university instructors. This might be because each semester has different courses to teach, and sometimes an instructor may teach a different class than the previous one. Therefore, an instructor may find that the results obtained do not apply to planning and organising classroom instruction because of these changes. Literatures suggest the inclusion of assessment for teaching and learning tasks for effective implementation of instructional plans and decision-making in universities [48, 50]. Remarkably, instructors are required to be skilled in using assessment results when making decisions about individual students, planning teaching based on learning outcomes, developing curriculum, and advancing university improvement. Failure to use assessment results may lead to wrong interpretations of educational outcomes and students' learning.

4. Conclusions

The study established that university instructors demonstrated varying levels of self-perceived assessment competencies. Notably, competencies related to communicating assessment procedures and results, analysing and interpreting assessment outcomes, using assessment results to inform instruction and following standardised procedures and principles in test item construction, such as the use of tables of specifications and Bloom's taxonomy, were less frequently implemented. In contrast, avoiding fault grading was highly practised by instructors. Moreover, the regression analysis underscores that professional teaching background and teaching experience are significant predictors of high assessment competence. Those without formal teaching qualifications or with fewer years of experience tend to demonstrate lower assessment literacy and effectiveness.

Both quantitative and qualitative data reveal a gap between awareness and actual application of sound assessment practices. While some instructors possess theoretical knowledge acquired through training, such as UTLIP, many do not apply it effectively. The lack of use of the table of specifications and imbalanced testing practices favouring either low- or high-order cognitive skills highlight deficiencies in instructional alignment and fairness in student evaluation.

These findings point to a broader systemic issue within higher education institutions, where assessment literacy is insufficiently prioritised, thus leading to inconsistencies in assessment quality and potentially unfair outcomes for students.

5. Recommendations

To enhance university instructors' assessment competencies, institutions should prioritise ongoing professional development focused on effective assessment practices. This includes training on constructing valid and reliable assessments using tools such as Bloom's taxonomy and table of specifications. Special attention should be given to instructors without formal pedagogical training or with limited teaching experience. Institutions should also establish clear guidelines and accountability measures that standardise assessment procedures across departments. These efforts will help ensure fairness, consistency, and alignment between learning objectives and evaluation methods.

Instructors should be encouraged and supported to use assessment results to inform instructional decisions and give meaningful feedback to students. Practical steps should include developing institutional templates for analysing assessment outcomes, integrating assessment review into regular departmental meetings, and requiring evidence-based justifications for test design. By embedding these practices into teaching routines and performance evaluations, universities can foster a culture of reflective and data-driven instruction that ultimately improves student learning outcomes.

Furthermore, as this study was limited to three public universities, future research could extend to private universities to enable comparative analysis. The resulting insights could inform the development of comprehensive, long-term strategies for strengthening assessment literacy across Tanzania's higher education sector.

Funding: This study formed part of my Doctor of Philosophy (PhD) research, funded by the University of Dodoma.

Data availability statement: Data will be made available on request.

Conflicts of interest: The authors declare no conflict of interest.

Acknowledgments: I express my sincere gratitude to the universities' management for the permission to collect data in their respective universities. Special thanks to the instructors, heads of department, and quality assurance officers for their time and patience during data collection.

References

[1] Agu, N.N., Onyekuba, C. and Anyichie, A.C., 2013. Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research and Reviews*, 8(8), pp.431–439. Available from: https://www.researchgate.net/publication/316939632.

[2] *Alternative Assessment in Higher Education: A Practical Guide to Assessing Learning*, 2021. Putrajaya, Malaysia: Department of Higher Education, Ministry of Higher Education Malaysia. Available from: https://cadelead.upm.edu.my/upload/dokumen/20221223155224eBookBKA_l_Alternative_Assessment_in_HE-compressed.pdf.

[3] Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. and Wittrock, M.C., eds, 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Abridged ed. New York: Longman. Available from: https://eclass.uoa.gr/modules/document/file.php/PPP242/Lorin%20W.%20Anderson%2C%20David%20R.%20Krathwohl%20-%20A%20taxonomy%20for%20learning%20teaching%20and%20assessing_%20a%20revision%20of%20Bloom%60s%20taxonomy%20of%20educational%20objetives-Longman%20%282001%29.pdf.

[4] Athanases, S.Z., Bennett, L.H. and Wahleithner, J.M., 2013. Fostering Data Literacy Through Preservice Teacher Inquiry in English Language Arts. *The Teacher Educator*, 48(1), pp.8–28. Available from: https://doi.org/10.1080/08878730.2012.740151.

[5] Biggs, J. and Tang, C., 2020. Constructive Alignment: An Outcomes-Based Approach to Teaching Anatomy. In: L.K. Chan and W. Pawlina, eds. *Teaching Anatomy: A Practical Guide*. Cham: Springer International Publishing, pp.23–30. Available from: https://doi.org/10.1007/978-3-030-43283-6_3.

[6] Black, P. and Wiliam, D., 2009. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), pp.5–31. Available from: https://doi.org/10.1007/s11092-008-9068-5.

[7] Black, P. and Wiliam, D., 2012. Assessment for Learning in the Classroom. In: J. Gardner, ed. *Assessment and Learning*. 2nd ed. SAGE Publications Ltd, pp.11–32. Available from: https://doi.org/10.4135/9781446250808.n2.

[8] Boud, D. and Falchikov, N., 2006. Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), pp.399–413. Available from: https://doi.org/10.1080/02602930600679050.

[9] Brookhart, S.M., 2003. Developing Measurement Theory for Classroom Assessment Purposes and Uses. *Educational Measurement: Issues and Practice*, 22(4), pp.5–12. Available from: https://doi.org/10.1111/j.1745-3992.2003.tb00139.x.

[10] Brookhart, S.M., 2011. Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30(1), pp.3–12. Available from: https://doi.org/10.1111/j.1745-3992.2010.00195.x.

[11] Carless, D., 2015. *Excellence in University Assessment: Learning from award-winning practice*. London: Routledge. Available from: https://doi.org/10.4324/9781315740621.

[12] Chalchisa, D., 2014. Practices of Assessing Graduate Students' Learning Outcomes in Selected Ethiopian Higher Education Institutions. *Journal of International Cooperation in Education*, 16(2), pp.157–180. Available from: https://cice.hiroshima-u.ac.jp/wp-content/uploads/2014/10/16-2-10.pdf.

[13] Cohen, L., Manion, L. and Morrison, K., 2017. *Research Methods in Education*. 8th ed. London: Routledge. Available from: https://tinyurl.com/2hv99fza.

[14] Creswell, J.W., 2012. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. 4th ed. Boston, MA: Pearson Education. Available from: https://eclass.uoa.gr/modules/document/file.php/ECD433/%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7%20%CE%B4%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CF%89%CE%BD%20%CE%BC%CE%B5%20%CF%84%CE%BF%20SPSS/%CE%92%CE%B9%CE%B2%CE%BB%CE%B9%CE%BF%CE%B3%CF%81%CE%B1%CF%86%CE%AF%CE%B1/John%20W.%20Creswell%20-%20Educational%20Research_%20Planning%2C%20Conducting%2C%20and%20Evaluating%20Quantitative%20and%20Qualitative%20Research%2C%204th%20Edition%20%282011%2C%20Addison%20Wesley%29.pdf.

[15] Datnow, A. and Hubbard, L., 2015. Teachers' Use of Assessment Data to Inform Instruction: Lessons from the past and Prospects for the Future. *Teachers College Record*, 117(4), pp.1–26. Available from: https://doi.org/10.1177/016146811511700408.

[16] DePascale, C., Sharp, A., Ryan, K. and Betebenner, D., 2018. *Building a Conceptual Framework for Assessment Literacy*. National Center for the Improvement of Educational Assessment. Available from: https://www.nciea.org/wp-content/uploads/2021/11/Assessment-Literacy-Framework-5-18.pdf.

[17] Earl, L. and Timperley, H., 2015. *Evaluative thinking for successful educational*

innovation, OECD Education Working Papers 122. Paris: OECD Publishing. Available from: https://doi.org/10.1787/5jrxtk1jtdwf-en.

[18] Field, A., 2009. Logistic regression. *Discovering Statistics Using SPSS (and sex and drugs and rock 'n' roll)*. 3rd ed. London: SAGE Publications, chap. 8, pp.264–316. Available from: https://www.researchgate.net/profile/Abdelrahman-Zueter/post/What_are_the_conditions_for_using_Ordinal_Logistic_regression_Can_anyone_share_the_various_regression_methods_and_their_application/attachment/59d637d8c49f478072ea5080/AS%3A273691429015552%401442264529487/download/DISCOVERING+STATISTICS.pdf.

[19] Fullan, M., 2025. *The New Meaning of Educational Change*. 6th ed. Teachers College Press.

[20] Gaikwad, S., Wadegaonkar, A., Mitra, G. and Chakravarty, D., 2023. Assessment Literacy, Current Assessment Practices and Future Training: Reflections of Teachers in Higher Education. *International Journal of Learning, Teaching and Educational Research*, 22(7), pp.1–29. Available from: https://doi.org/10.26803/ijlter.22.7.1.

[21] Gordon College of Education and UNICEF Montenegro, 2025. Assessment of the Teachers' Professional Development System in Montenegro. Available from: https://www.unicef.org/montenegro/en/media/25686/file/Assessment%20of%20the%20Teachers%E2%80%99%20Professional%20Development%20System%20in%20Montenegro.pdf.

[22] Hamafyelto, R.S., Hamman-Tukur, A. and Hamafyelto, S.S., 2015. Assessing Teacher Competence in Test Construction and Content Validity of Teacher Made Examination Questions in Commerce in Borno State, Nigeria. *Journal of Education and Practice*, 5(5), pp.123–128. Available from: http://article.sapub.org/10.5923.j.edu.20150505.01.html.

[23] Istoroyekti, Y.M., 2016. Issues challenging universities: A case of Tanzanian higher education. *Ahmad Dahlan Journal of English Studies*, 3(1), pp.51–62. Available from: https://doi.org/10.26555/adjes.v3i1.3625.

[24] Jones, B., Chang, S., Heritage, M., Tobiason, G. and hERMAN j, 2015. *Supporting students in close reading*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Available from: https://files.eric.ed.gov/fulltext/ED565336.pdf.

[25] Kimaro, A.R. and Kapinga, B.B., 2020. An assessment of instructors' classroom assessment practice in selected higher learning institutions, Tanzania. *Tengeru Community Development Journal*, 7(1), pp.54–66. Available from: https://ticd.ac.tz/wp-content/uploads/2021/05/kimaro-4.pdf.

[26] Kitta, S., 2017. *TCU quality assurance training workshop, 10th February, 2017. Principles and procedures of test construction*. Dar es Salaam: Tanzania Institute of Education (TIE).

[27] Kitula, P.R., Kireti, P. and Wambiya, P., 2018. Perceived Efficacy of University Lecturers in Conducting Assessment among Selected Universities in Tanzania. *International Journal of Education and Research*, 6(6), pp.121–130. Available from: https://www.ijern.com/journal/2018/June-2018/09.pdf.

[28] Koloi-Keaikitse, S., 2017. Assessment of teacher perceived skill in classroom assessment practices using IRT Models. *Cogent Education*, 4(1), p.1281202. Available from: https://doi.org/10.1080/2331186X.2017.1281202.

[29] Levi, T. and Inbar-Lourie, O., 2020. Assessment Literacy or Language Assessment Literacy: Learning from the Teachers. *Language Assessment Quarterly*, 17(2), pp.168–182. Available from: https://doi.org/10.1080/15434303.2019.1692347.

[30] Levy-Feldman, I., 2025. The Role of Assessment in Improving Education and Promoting Educational Equity. *Education Sciences*, 15(2), p.224. Available from:

https://doi.org/10.3390/educsci15020224.
[31] Mandinach, E.B. and Gummer, E.S., 2012. *Navigating the Landscape of Data Literacy: It IS Complex*. WestEd. Available from: https://files.eric.ed.gov/fulltext/ED582807.pdf.
[32] Mandinach, E.B. and Gummer, E.S., 2013. A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, 42(1), pp.30–37. Available from: https://doi.org/10.3102/0013189X12459803.
[33] Matovu, M. and Zubairi, A.M., 2015. Assessment Practices in the Developing World: Predictors of Assessment Practices in Ugandan Institutions of Higher Learning. *IIUM Journal of Educational Studies*, 3(2), p.75–112. Available from: https://doi.org/10.31436/ijes.v3i2.96.
[34] Mbua, F.N., 2014. *Educational Administration: Issues and Perspectives*. Limbe, Cameroon: Pressprint.
[35] McCallum, S. and Milner, M.M., 2021. The effectiveness of formative assessment: student views and staff reflections. *Assessment & Evaluation in Higher Education*, 46(1), pp.1–16. Available from: https://doi.org/10.1080/02602938.2020.1754761.
[36] Meyer, J.P., 2010. *Reliability*. New York, NY: Oxford University Press.
[37] Michigan Assessment Consortium, 2020. Assessment Literacy Standards: A National Imperative. Available from: http://michiganassessmentconsortium.org/sites/default/files/mac_AssessLitStds_mobile.pdf.
[38] Miller, M.D., Linn, R.L. and Gronlund, N.E., 2009. *Measurement and Assessment in Teaching*. 10th ed. Upper Saddle River, New Jersey: Pearson Education, Inc. Available from: https://fliphtml5.com/dyclx/dpwx/Measurement_and_Assessment_in_Teaching_%28Tenth_Edition%29_by_M._David_Miller%2C_Robert_L._Linn%2C_Norman_E._Gronlund_%28z-lib.org%29/5/.
[39] Munene, C.K. and Ogula, P., 1999. *Handbook on educational assessment and evaluation*. Nairobi, Kenya: New Kemit Publishers.
[40] Mwesigwa, F. and Nakato, N., 2025. Mentorship in Teacher Education Programs: A Review of Practices, Outcomes, and Challenges. *Acta Pedagogia Asiana*, 5(1), p.15–29. Available from: https://doi.org/10.53623/apga.v5i1.679.
[41] Nguon, S., 2013. *Assessment practices in a Cambodian university: Through the lens of lecturers and students*. Master's thesis. Victoria University of Wellington. Available from: https://doi.org/10.26686/wgtn.17004847.
[42] Nwakanma, C.H., 2024. The Influence of Mentoring on Career Development of University Tutors: A Narrative Review. *Didaktika: Jurnal Kependidikan*, 13(1 Februari), pp.1037–1050. Available from: https://doi.org/10.58230/27454312.545.
[43] Ogula, P.A. and Onsongo, J.K., 2009. *Handbook on teaching and learning in higher education*. Nairobi: Catholic University of Eastern Africa (CUEA) Press.
[44] Omari, I.M., 2011. *Concepts and Techniques in Educational Assessments and Evaluation*, vol. 1. Dar es Salaam: Oxford University Press.
[45] Omo-Egbekuse, J., Afemikhe, O.A. and Imobekhai, S., 2010. Standards for teacher competence in educational assessment of students: Nigerian teachers' ratings of their need. *The 36th Annual Conference of the International Association for Educational Assessment (IAEA 2010), Bangkok, Thailand, 22-27 August 2010*. Available from: https://tinyurl.com/44fbzz6z.
[46] Organization for Economic Co-operation and Development, 2008. Assessment for Learning: Formative Assessment. *OECD/CERI International Conference "Learning in the 21st Century: Research, Innovation and Policy"*. Paris: OECD/CERI. Available from: https://www.utoledo.edu/aapr/assessment/pdfs/15_April-1_Formative%20Assessment.pdf.

[47] Pallant, J., 2011. *SPSS Survival Manual: A step by step guide to data analysis using SPSS*. 4th ed. Crows Nest, Australia: Allen & Unwin. Available from: https://www.researchgate.net/profile/Mohd-Aliff-Abdul-Majid/post/Can_i_run_correlation_between_categorical_and_continuous_variableeg_age_gender_salary_level_of_education_and_job_satisfaction/attachment/5e6166033843b0499fed181d/AS%3A865830183120896%401583441410004/download/Pallant+%282010%29+SPSS+Survival+Manual+%286th+ed%29.pdf.

[48] Popham, W.J., 2008. *Transformative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

[49] Popham, W.J., 2009. Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48(1), pp.4–11. Available from: https://doi.org/10.1080/00405840802577536.

[50] Punongbayan, E.J. and Bauyon, S.M., 2015. Instructional Performance of Teacher Education Faculty Members in One State University in the Philippines. *Asia Pacific Journal of Multidisciplinary Research*, 2(5), pp.135–143. Available from: https://www.academia.edu/123987390/.

[51] Rezvani Kalajahi, S.A. and Abdullah, A.N., 2016. Assessing Assessment Literacy and Practices among Lecturers. *Pedagogika*, 124(4), p.232–248. Available from: https://doi.org/10.15823/p.2016.65.

[52] Rubeba, A.M. and William, F., 2019. nstructors' Test Construction Competences and their Implications on Teaching and Learning in Tanzanian Universities. *Journal of Adult Education in Tanzania*, 22(1), pp.111–130. Available from: https://jaet.iae.ac.tz/index.php/adulteducation/article/view/43.

[53] Sales, R., 2013. *Assessment practices in higher education: the experiences of newly appointed academics in professional fields from a phenomenological perspective*. Ph.D. thesis. University of the West of England. Available from: https://uwe-repository.worktribe.com/output/928705.

[54] Shumba, A. and Naong, M., 2012. Factors Influencing Students' Career Choice and Aspirations in South Africa. *Journal of social sciences*, 33(2), pp.169–178. Available from: http://krepublishers.com/02-Journals/JSS/JSS-33-0-000-12-Web/JSS-33-2-000-12-Abst-PDF/JSS-33-2-169-12-1397-Shumba-A/JSS-33-2-169-178-12-1397-Shumba-A-Tx%5B4%5D.pdf.

[55] Smith, C.D., Worsfold, K., Davies, L., Fisher, R. and McPhail, R., 2013. Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education*, 38(1), pp.44–60. Available from: https://doi.org/10.1080/02602938.2011.598636.

[56] Sokoli, D. and Koren, A., 2017. Qualification of lecturers as indicator of quality teaching in higher education. *Management Challenges in a Network Economy. 17–19 May 2017. Lublin, Poland*. pp.445–451. Available from: https://tinyurl.com/3w6jdtu6.

[57] Srivastava, D.S. and Kumari, S., 2005. *Education: Assessment, evaluation and remedial*. Delhi, India: Isha Books.

[58] Stiggins, R., 2014. Improve assessment literacy outside of schools too. *Phi Delta Kappan*, 96(2), pp.67–72. Available from: https://doi.org/10.1177/0031721714553413.

[59] Stiggins, R.J., 2002. Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan*, 83(10), pp.758–765. Available from: https://doi.org/10.1177/003172170208301010.

[60] Tadesse, T. and Gillies, R.M., 2015. Nurturing Cooperative Learning Pedagogies in Higher Education Classrooms: Evidence of Instructional Reform and Potential Challenges. *Current Issues in Education*, 18(2). Available from: https://cie.asu.

edu/ojs/index.php/cieatasu/article/view/1374.

[61] Thakur, A. and Shekhawat, M., 2014. The Study of Different Components of Teacher Competencies and their E ffectiveness on Student Performance.(According to Students). *International Journal of Engineering Research & Technology (IJERT)*, 3(7), pp.1426–1428. Available from: https://tinyurl.com/46usf87c.

[62] Walde, G.S., 2016. Assessment of the implementation of continuous assessment: the case of METTU university. *European Journal of Science and Mathematics Education*, 4(4), pp.534–544. Available from: https://eric.ed.gov/?id=EJ1118167.

[63] Waugh, C.K. and Gronlund, N.E., 2012. *Assessment of Student Achievement*. Tenth ed. Pearson Education. Available from: https://newteacherlibraryandtools.square.site/uploads/b/e4443390-f724-11ec-9291-b5b724deb1dc/Assessment%20of%20Student%20Achievement%20(C.%20Keith%20Waugh%20-%20Norman%20E.%20Gronlund)%20(Z-Library).pdf.

[64] Yamane, T., 1967. *Statistics: An Introductory Analysis*. 2nd ed. New York: Harper and Row.

[65] Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L. and Lao, H., 2021. A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), pp.228–260. Available from: https://doi.org/10.1080/0969594X.2021.1884042.

[66] Zembazemba, R.D., 2017. Educational Assessment Practices in Tanzania: A Critical Reflection. *Business Education Journal*, 1(3). Available from: https://doi.org/10.26555/adjes.v3i1.3625.