Full Length

The Practices of Assessment, Test construction and Administration at Kotebe University of Education (KUE)

Bekalu Atnafu Taye[1], Dejen Chaka[2], Nebyou Gashu[3], Birhane Mulugeta[4], Walelign Admassu[5] Birhane Admasu[6], Yeshaw Tesema[7], Yoseph Shumi[8] , Kebede Terfesa[9]

[1]Corresponding author: E-mail: bekaluatnafutaye@gmail.com

Abstract

The main objective of this study is to assess perception and practice of assessment, test construction and administration at Kotebe University of Education (KUE). To conduct the study, descriptive research design with a mixed methods approach (i.e., quantitative and qualitative methods) was employed. In attempts to data collection, two hundred fifty-nine students from twenty-two departments, and one hundred thirty-two instructors were contacted through availability sampling. Questionnaire, semi-structured interview, final exam session's observation and document analysis were used to collect the data. The quantitative data were analyzed using descriptive and inferential statistics while qualitative data were analyzed thematically. The findings of the study depicted that teachers particularly those who have had the highest qualifications had positive attitudes towards assessments. Analyzing college level staff perceptional variation, teachers who are affiliated with the colleges of Educational Sciences and Language Education and Institute of Physical Education and Sports' staff were in favor of educational assessments. With respect to assessment tool construction, however, it was discovered that instructors at KUE did not carry out item analyses. In connection with the frequency of test administration, the majority of students and teachers confirmed that assessments were undertaken regularly while some respondents had reservation in this regard. Those informants who did not agree with the majority claimed that the commitment and dedication to administrating test frequently depends on individual teachers' traits and passions.

Key words: Assessment; Test Construction; Test Administration; Exams

1. Introduction

Assessment is a tool for gathering information about learners, providing feedback and informing instructional decisions. As Puhl (1997) noted, assessment plays a crucial role in improving the teaching-learning process. Ramsden (2003)further emphasized its functions as a means of supporting student learning, reporting on progress, and informing teaching practices.Assessment measures student achievement, provides feedback to both teachers and students, motivates and directs learning, evaluates teaching methods and programs and facilitates learning (Ebel, 1979; Hopkings & Stanley, 1981).

Assessment tools and procedures are crucial for gauging students' progress and achievement, and they also help evaluate the effectiveness of the curriculum, instruction and teaching methods(Shaaban, 2001).Effective assessment demands a variety of methods to address students' diverse needs, interests, and learning styles, taking both informative and summative assessments

into account. Conducting summative assessments alone may be perceived as judgmental specifically among students. When this happens, students become nervous, and they might be negatively impacted. Hence, administering both final and continuous assessment serves to gather valid and reliable information to see the ongoing process and to make effective decisions about students' learning(Nitko, 2006), and this consequently avoids students' feelings.

To get highly valued and reliable information about performance of students, assessment has to be regarded as an integral part of teaching learning process. However, studies indicated that instructors use continuous assessment results mainly for grading students (Seyum, 2006). Likewise, Muluken's (2006) findings noted that continuous assessment was viewed as simply giving series of tests for students to measure their performance. Similarly, Alsubaie(2015) identified some of the problems against the proper implementation of assessment which calls for the realignment of values, practices and outcomes. Some of the challenges that are associated with educators include a lack of skills in assessment tools' construction and administration, attitudes towards continuous assessment, and record keeping. Given the importance of assessment in fostering effective learning, it has become a significant focus in Ethiopia's higher education system, particularly in the context of active learning. As outlined in the Ministry of Education's policy document (MoE, 2003), the effective practice of assessment is a key objective of teacher education in Ethiopia.

In a study conducted at various academic institutions, Yalew (2008)found out that teacher-made tests had critical issues related to language use, adherence to item writing principles and guidelines, content validity, and other technical concerns. All these showed that the practice assessment, test construction and administration might face serious challenges in the education system. This study, therefore, addressed the following five questions which focused on the teachers' perceptions about assessment, the overall practice of test construction and administration at KUE, and the associations among these variables.

- What are the perceptions of teachers about assessment?
- What are the overall assessment practices of teachers?
- What are the practices of items construction at KUE?
- What are the test administration practices of teachers?
- Are there any associations among teachers' perception, assessment construction and administration practice?

2. Methods

This study has been designed to answer those five research questions. To address these questions, both quantitative and qualitative research methods were employed. The quantitative approach was utilized to assess teachers' perception on assessment, item construction and administration practice; whereas the qualitative approach was used to triangulate the findings derived from the quantitative data.

In this study, a descriptive research designed was employed. A concurrent combination of mixed-methods research approach was also chosen as it can help triangulate set of findings against others, thereby achieving a comprehensive understanding of the variables involved. In this approach,

quantitative and qualitative data are collected and analyzed independently, but the results are integrated during the interpretation phase to optimize the reliability of the study (Dornyei, 2007). Owing to this, both the quantitative and qualitative research instruments were administered to gain a full understanding of the practice of teacher-made items construction and administration at KUE. Accordingly, wide spectrums of data were gathered through the quantitative tools though that attempt provided limited insights of teachers' experiences. To address this limitation, qualitative research instruments were employed to gain in-depth enlightenment on the issues being investigated.

There were about four hundred teachers and eight hundred senior students at KUE by the time the data were collected. From the total population, more than thirty percent of teachers and the student population were taking part in the study. As a result, two hundred fifty-nine students, which was 32 per cent of the total population, from twenty-two departments and one hundred thirty-two teachers, 33 per cent of the target population, were participating in the study.

Likert scale questionnaires, semi-structured interviews, final exam session observations and document content analyses were deployed to collect data. Two different types of questionnaires were prepared and administered to teachers and students. The teachers' questionnaire included items related to tools to be used for constructing and administering continuous assessments, examining the practice on the ground, and teachers' perceptions. Similarly, students were requested to complete questionnaire about tools teachers employed to conduct continuous assessments. Of the four research tools employed in this study, interviews, observations, and document analysis were used for qualitative data collection. Semi-structured questions were designed to conduct interviews with fifteen teachers. To this end, participants were asked to reflect on their practice regarding assessment tools construction and administration, and these data served to triangulate the results found through other instruments. Final examination sessions were also observed using checklist, focusing on the exam administration process. This checklist included ten items that could be answered with a simple 'yes' or 'no.' Over ten exam sessions were observed to gather comprehensive data. Furthermore, a total of forty-two final exam papers were analyzed, taking the principles of test construction into account. These research instruments were reviewed by experts and fine-tuned before administering them to collect data.

Eventually, the data were tabulated and analyzed by using descriptive and inferential statistical values. The descriptive statistics were used to classify and summarize numerical data using mean scores and percentage whereas the inferential statistics were made in terms of correlational accounts. The inferential statistics were computed to establish whether or not the results obtained from the informants were powerful enough to generalize to the whole population. This was made because the descriptive statistics does not allow to draw general conclusion that would go beyond the sample. The qualitative data were transcribed verbatim and the accounts were thematically analyzed.

To assess the validity of data collected, the research instruments were reviewed by experts in the area. To determine the reliability, Cronbach's alpha coefficient was calculated, and it was found to be0.81, which indicates that it was a very good level of internal consistent (reliability) of items used to collect data in this study. For validation in qualitative data, Creswell (2007)proposed eight validation strategies and recommended that qualitative researchers utilize at least two of them in

any given study. Accordingly, the study employed three significant validation strategies such as external audit, member checking and triangulation.

3. Results and Discussions

This part of the study presents the results and discussions sections. First, it presents the demographic description of the study participants. Next, it shows the results of teachers' perception, practice of continuous assessments, test construction and administration and then the relation of perception to practice in constructing assessments items and administering them. It also deals with students' reflection on the practice.

3.1.The Demographic Descriptions of the Study Participants

The data for the research were collected from teachers and students. The demographic data of the study participants are presented in Table1, Table 2 and Figure1 below.

Table 1: Educational background of the teacher participants

Qualifications of the Teachers	*Academic Units of the University (i.e., KUE)*						
	Languages Education	*Social Sciences*	*Natural Sciences*	*Sport Institute*	*Educational Sciences*	*FBE & Urban*	*Total*
PHD	16	7	16	2	7	1	**49**
MA/MSc	6	26	21	1	12	16	**82**
BA/BSc	0	0	0	0	0	1	**1**
Total	**22**	**33**	**37**	**3**	**19**	**18**	**132**

As it can be seen from Table 1 above, the majority of the study participants were found to hold second and third degrees. Specifically, 49 participants had doctoral degree, while 82 were master's degree holder. Of the 132 teachers contacted, there was a participant with a bachelor's degree. Thus, it can also be noticed that the participant were from five colleges and an institute.

Table 1: The teaching experience of teachers

Years of teaching Experiences	*Academic Units of the University (i.e., KUE)*							
	Languages Education	*Social Sciences*	*Natural Sciences*	*Sport Institute*	*Educational Sciences*	*FBE & Urban*	*Total*	
							F	*%*
1-5	*0*	*1*	*1*	*0*	*0*	*3*	***5***	4
6-10	*0*	*7*	*9*	*0*	*3*	*5*	***24***	18
11-15	*5*	*16*	*8*	*0*	*3*	*8*	***40***	30
16-20	*6*	*7*	*10*	*3*	*4*	*1*	***31***	24
Above 20	*11*	*2*	*9*	*0*	*9*	*1*	***32***	24
Total	*22*	*33*	*37*	*3*	*19*	*18*	***132***	100

Table 2 above depicts that majority of teachers had rich experience in teaching. In other words, of the 132 teachers who had participated in the study, 103 (78%) had got more than ten years of

experiences. Therefore, the findings of the study revealed that only 29 teacher informants had ten and less years of teaching experience. This implies that three out of four teachers were thought to develop skills and knowledge that are required to construct valid and reliable items of assessments to gauge the performance of their students.

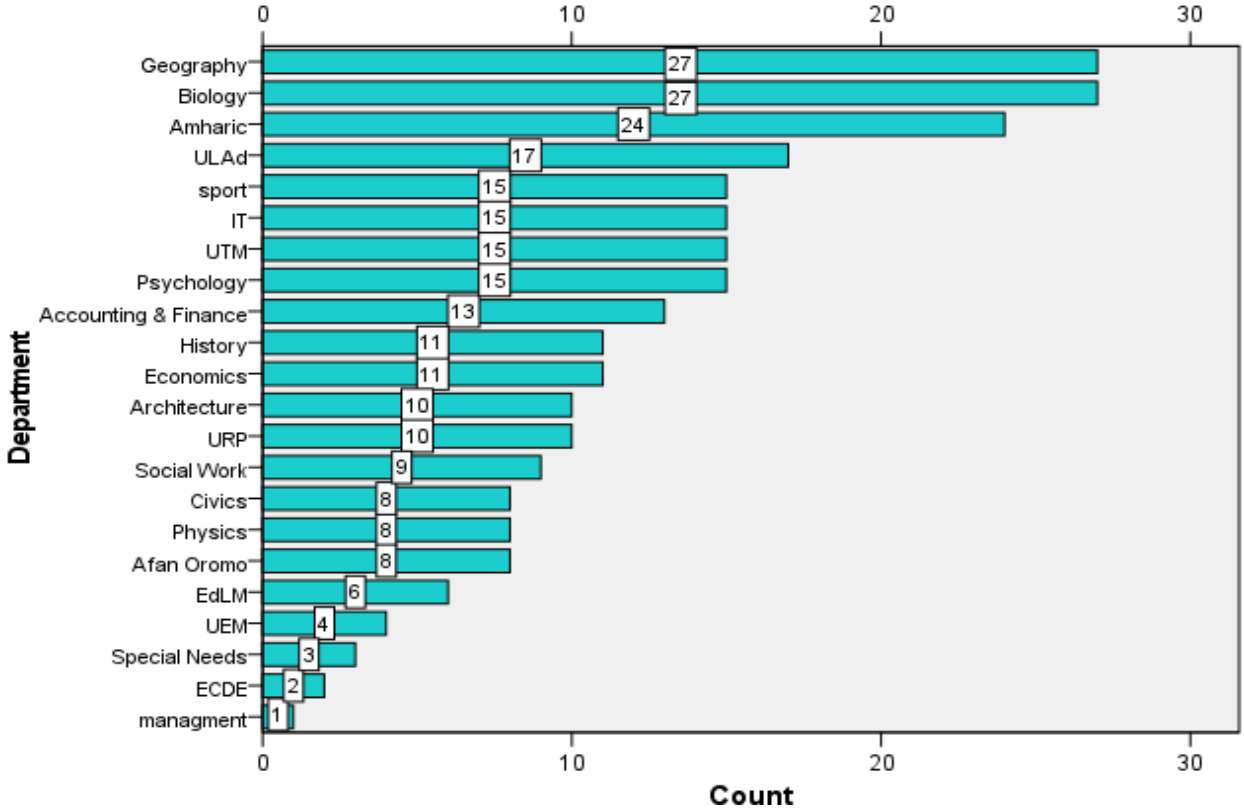


Figure1**: **Number of student participants by department

On the other hand, as shown in Figure1 above, 259 students of all batches in the undergraduate programs, 2020 to 2023, from 22 departments participated in the survey. These data were collected from different groups of students in every department and of all year levels. The histogram presented above clearly indicates that a number students ranges from an individual in Management Department to the highest number (i.e., 27) in Geography and Biology departments.

3.2.Teachers' Perception about Assessment

The data about perception of teachers on assessment in general and continuous assessment in particular were collected using questionnaires and the results are presented in Table 3-5 and Fig. 2 below.

Table 2: Instructors' perceptions about assessment

No	Statements	Skewness	M	SD
1	*Learning and assessment are inseparable*	*-1.86*	*4.32*	*1.03*
2	*Assessment improves teaching-learning process and performance*	*-2.34*	*4.45*	*0.87*
3	*Assessment encourages learners to engage in learning experience*	*-1.82*	*4.26*	*0.93*
4	*Assessment is an integral part of the teaching learning process*	*-1.65*	*4.27*	*0.90*
5	*Assessment is given only to grade students*	*1.13*	*2.00*	*1.17*

As illustrated in Table 3 above, teachers' perceptions regarding the statement that learning and assessment are inseparable in a sense that assessment improves teaching and learning and that assessment encourages learners to engage learning experience were notably high with mean scores of 4.32, 4.45, and 4.26, respectively. These scores indicate that the informant teachers generally held a positive view about assessments. The analysis of the mean scores (M) alongside the standard deviation (SD) for the provided statements can yield insights into the perceptions of the participants regarding assessment in the educational context. To this end, statements 1, 2, 3 and 4 have high mean scores (ranging from 4.26 to 4.45) with relatively low standard deviations (0.87 to 1.03). The values of SD in this context suggest that the data points tend to close to the means as the points do not go far beyond 1. In other words, the findings are good enough to show consistency. The skewness presented in Table 3 above, which ranges from -2.34 to -1.65, could also fairly confirm the consistency or replication of the results as $\pm$1 skewness value is often acceptable. On contrary, item number5 (M=2, SD=1.17) have low mean scores, indicating disagreement of participants with the statement. The lower score for statement 'assessment is only for grading' indicate that assessment serves broader educational purposes beyond just giving grades to students. In this case, the standard deviation for these statements is found to be higher which showcases the variability in the responses. Overall, the findings reveal that participants believed in the educational values of assessments.

In conclusion, the combination of high mean scores with low standard deviation for most items in this study indicates a favorable and consistent perspective on the role of educational assessments. In contrast, the lower mean scores with higher variability reflect a discrepancy in views regarding the role of assessments. This suggests that further academic dialogues need to be conducted among educational community.

Table 3: Instructors' perceptions about continuous assessment

No	***Statements***	***M***	***SD***
1	*Continuous assessment (CA) is as merely giving paper/pencil tests*	*2.37*	*1.18*
2	*CA enables learners to master the contents of the lesson*	*3.90*	*0.98*
3	*CA offers better reliable and valid information than terminal assessment*	*3.99*	*0.86*
4	*CA is able to provide on-going picture of how individual student develops*	*4.14*	*0.84*
5	*CA arouse students' desire for attention and concentration while in class*	*4.11*	*0.73*
6	*CA gives early warnings of which students are having problems with a course*	*4.01.*	*0.92*
7	*CA is beyond testing*	*4.01*	*0.94*
8	*CA helps learners to evaluate themselves.*	*4.08*	*0.86*

As shown in Table 4 above, teachers' perceptions regarding continuous assessment (CA) indicated a range of views. The first statement, "…CA is as merely giving paper/pencil tests," has a low mean score and high standard deviation (M=2.37,SD=1.18). This indicates a disagreement among participants regarding the perception of CA as merely traditional testing, suggesting that some participants may hold this view while others do not. However, statements 4 to 8 have high mean scores (ranging from 4.01 to 4.14) and relatively low standard deviations (0.73 to 0.94) which confirm that CA provides valuable insights into student development, encourages engagement, and helps learners evaluate themselves.

Statements 2 and 3 have moderate mean scores (3.90 and 3.99) with low standard deviations (0.86 and 0.98). This indicates that participants generally agree that CA helps learners master content and provides reliable information as compared to terminal assessments. Overall, the findings in this

regard reveal a clear difference in perceptions among the participants. In general, high mean scores and low standard deviations for most statements indicate a favorable view of continuous assessment as a comprehensive educational tool. However, the low mean score for the first statement highlights a misconception that needs to be addressed, suggesting that further education on the nature and purpose of continuous assessment may be beneficial for participants.

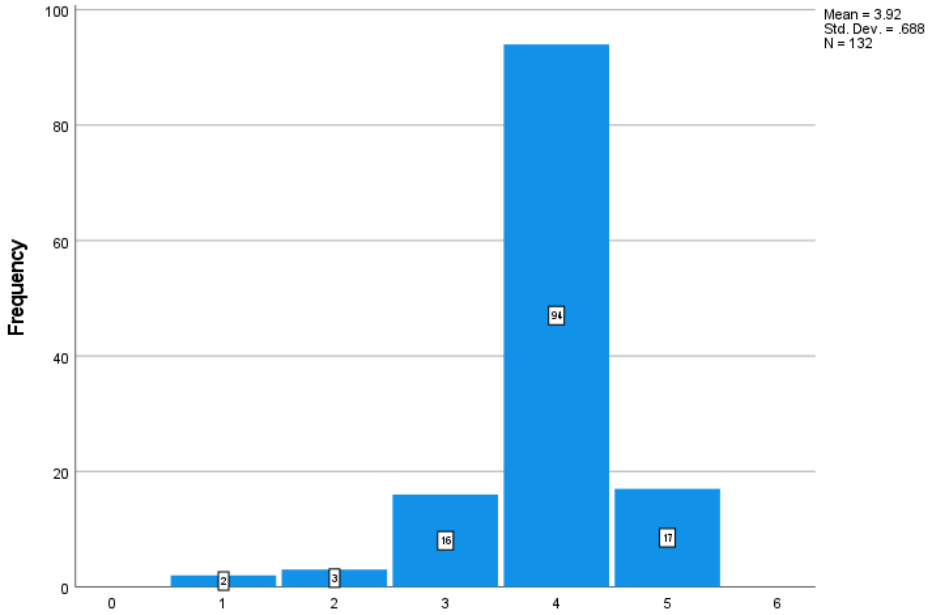


Figure 2: Teachers' level of agreement about assessment

As shown in the Figure, a significant majority of the subjects (94%) believed that assessment goes beyond mere testing. Additionally, most respondents (16% and17%) viewed assessment as being more than just administering tests. These findings indicate that the majority of the instructors possess a solid understanding of the concept of assessment in general and continuous assessment in particular. In contrast, the findings of Muluken (2006) revealed that teachers perceived continuous assessment as simply a series of tests designed to measure students' performance.

Table 4:Discrepancies of Teachers' perception about assessment by qualification and colleges

No		*Category*	*Mean*
1	Qualification	PhD	3.8
		MA/MSc	3.7
		BA/BSc	3.3
2	Colleges	Colleges of Educational Sciences	3.9
		College of Language education.	3.8
		College of Social Science education	3.7
		College of business , technology and vocational education	3.6
		College of Science and mathematics education	3.6
3	Institute	Institutes of Physical Education and Sports	3.8

Table 5 above reveals that PhD holders have the mean score of 3.8, which shows that teachers with the highest degree or qualification have developed a good perception on assessment. On the other hand, perception of teachers about assessment varied among colleges. Faculty from College of Educational Sciences has the highest mean (M=3.9) followed by those of College of Languages Education and Institute of Physical Educations and Sports with the score of 3.8. This indicated that teachers who are specialized in Educational Sciences, Languages Education and Physical Educations and Sports appeared to have good perceptions on assessment as compared to others.

Table 5: ANOVA results of mean values of teachers' experience, department and qualification

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	1.275	3	.425	1.792	.152[b]
	Residual	30.357	128	.237		
	Total	31.631	131			

a. Dependent Variable: Perception

b. Predictors: (Constant): Teachers' experience, department and qualification

The ANOVA table above provides a statistical analysis of the relationship between the dependent variable "Perception" and the independent variables (predictors) which include 'Teachers' experience, department and qualification'. This p-value in this context indicates the probability that the observed results are due to chance. Given a common threshold for significance is 0.05, the p-value (0.152) is found to be higher suggesting that the model is not statistically significant, meaning the predictors do not explain a significant amount of the variance in the dependent variable 'perception'.

3.3.Overall Assessment practices of teachers at KUE

Teachers and students were asked to indicate the practice of assessment and the types of tools used for conducting assessments. Thus, the results are presented as follows in Table 7 and Table 8.

Table 6: Teachers and students' responses regarding the application of assessment

No	Respondents	Frequency and percentage of the responses			
		Yes		No	
		F	%	F	%
1	Students	198	76.4%	61	23.6
2	Teachers	124	93.9%	8	6.1%

As it is indicated in Table 7, the majority of students (76.4%) and the great majority of teachers (93.9%) reported that continuous assessment was carried out. The qualitative findings also indicated that teachers had conducted continuous assessment regularly. In this regard, Derartu*[1] noted the following:"*I think professionals at KUE understand the concepts of ...*

All names given to participants are pseudonyms.

*continuous assessment very well. It might not be difficult for someone at this level to understand the importance of assessment and we practice it. "*Similarly, Azeb expressed her observation saying,“ *The assessment at KUE seems to be nice. I had given continuous assessments four or five times in a semester…"*As it can be seen from the data presented here above, teachers claimed to apply continuous assessment.

Table 7: Teachers and students' responses regarding the types of assessment frequently used

Types of assessment	Response of students		Response of teachers	
	F	%	F	%
Tests	60	23.2	60	45.5
Assignments	72	27.8	39	29.5
Projects	36	13.9	11	8.3
Exams	87	33.6	16	12.1
Others	4	1.5	6	4.5
	259	**100.0**	**132**	**100**

As shown in Table8, student respondents were asked to reflect on the continuous assessment (CA) tools most frequently used by their teachers. The responses indicated that exams accounted for 33.6% of the tools utilized. Assignments, tests, and project activities were reported at 27.8%, 23.2%, and 13.9%, respectively, while other CA tools comprised only 1.5% of the teacher-made assessments. According to the students' responses presented in this Table, the most frequently used assessment tools are exams (33.6%) followed by assignments (27.8%), tests (23.2%), and projects (13.9%). However, teachers self-reports indicated that the use of tests, assignments, projects, exams and other tools to conduct assessments found to be 45.5%, 29.5%, 8.3% 12.1.% and 4.5% respectively. Though there is a bit mismatch between the teachers and students' replies, findings fairly indicate that there had been a tendency of relying on few tools such as tests, assignments, projects and exams for administering assessments.

Other assessment methods such as peer assessment, self-assessment, recap exercises, observation, attendance, term papers, portfolios, assessments, discussions, group reflections, participation, presentations, projects and quizzes appear to be underutilized by teachers. These findings align with the results of other researchers, such as Bekalu (2007), who reported that teachers did not conduct continuous assessment regularly despite recognizing its importance

3.4.Test construction practice of teachers at KUE: Data from exam papers analysis

As to the test construction practice of teachers, data were collected from respondents using questionnaire and interview and the results are presented in tables 9, 10, 11 and 12 below.

Table 8: The overall conditions of exam prepared and administered at different colleges

SN	common limitations observed from the exams papers analyzed	Colleges					
		ES	LE	SS	NS	FBE	SA
1	Leaving exam cover pages out	√	√	√	√	√	√
2	Not specifying time for doing exam	√	√	√	√	√	√
3	Excessive negative statements in exam items	√	√	√	√	√	√

4	Problems of arranging items according to their difficulty levels	√√	√	√√	√√	√	√
5	Teachers use "None of the above/all of the above"	√√	√	√	√	√	√
6	The exams lack either subjective or objective components	√	√	√	√	√	√
5	Instructions lack clear indication points (marks) the items have	√√	√	√	√	√	√
8	Equal number of premises and responses in the matching items.	√	√	√√	√	√√	√
9	Unequal number of alternatives (four and five distractors).	√	√√	√	√	√√	√

NB. √= *the problem has just been observed in the exam*
√√= *the problem has markedly been observed in the exam*

Table 9 above particularly depicts that the overall scenarios observed in the final exams constructed and administered in the first semester of 2023 academic calendar at KUE. The table presents deficiencies observed in the final exams prepared and administered. Some of these problems were found to be mild while some others could be rated as severe. To illustrate these circumstances, a single tick mark (√) is employed to indicate mild problems whereas double tick marks (√√) showcases severe problems observed in the exam items.

Table 9A: Teachers' test construction practice

No	Items Measuring Test Construction	N	Min	Max	M	SD.
1	I have items bank	132	1.00	2.00	1.24	.43
2	I prepare test blueprints	132	1.00	2.00	1.29	.45
3	I get test items reviewed	132	1.00	3.00	1.59	.66

As shown in the table above, the practice of test construction regarding the statements "*I have items bank," "I prepare a test blueprint," and "I get test items reviewed"* was low, with mean scores of 1.24, 1.29, and 1.59 respectively. These responses indicate that the sample teachers less engaged in analyzing item, preparing test blueprints during test construction and seeking feedback from colleagues on their exams. Qualitative data further corroborated these findings, revealing that teachers often do not prepare item analyses or test blueprints in the test construction process. During the interview session, Messay noted "*Test blueprints are not known in many faculties and departments…. Item analysis is not widely recognized*." Similarly, Zerihun said "*I construct items based on the objectives of the lesson, but I have never conducted item analysis*. Tola added:

> *Most of the time, teachers construct test items without using a table of specifications. They prepare items from an item bank but do not perform item analysis, nor do they have a culture of preparing items collaboratively at the departmental level.*

Genet expressed:

> *Teachers prepare test items without following guidelines for constructing different types of test items and do not consider test blueprints and item analysis. Most teachers overlook the importance of test blueprints*.

Both the quantitative and qualitative findings indicate that teachers tend to ignore utilizing test blueprints and item analysis when constructing exams. Item analysis, which includes item difficulty, discrimination index, and distractor analysis, helps test designers assess the quality of test items. Similarly, a test blueprint guides teachers in preparing tests. In a study conducted in Nigeria, Ngaji et al. (2021) highlighted a significant lack of competency in test item development among lecturers. Among reasons for teachers not preparing test blueprints, lack of time could be

worthy to mention (Yao, 2015). Another contributing factor might be teachers' lack of pedagogical skills (Peshal, 2020).

Table 100B:Teachers' test construction practice

SN	ITEMS MEASURING TEST CONSTRUCTION	N	MIN	MAX	M/	SD
1	I use equal number of alternatives throughout the exam.	132	1.00	2.00	1.44	0.49
2	I use 'all of the above' and 'none of the above frequently.	132	1.00	5.00	1.83	1.03

As it can be observed from items presented in Table 10B above, the mean score of the reply for the items "*I use equal number of alternatives throughout the exam,*" (M=1.44), which indicates that teachers did not use equal number of alternatives through of the exam. This finding was in line with the document analysis results. From the document analysis, it was learnt that some items had four alternatives and others have three or five. There are a number of irregular distributions of alternatives throughout the exam. In the following sample exam papers coded as: *Exam B, G from College of Educational Sciences, Exam D, J from College of Languages Education, Exam Z, X from College of Social Science Education, and Exam C, F, G, K, M from the College of Business, Technology and Vocational Education,* there had been ineffective distractors' constructions. As presented in Table 10A, there had also been tendencies of including less effective distractors like 'all of the above' and 'none of the above'. In the test coded as *Exam G, F,E from College of Educational Sciences, and Exam A, J, 1from the Institute of Physical Education and Sports.* Summary, as shown in the above findings, teachers seemed to be reluctant in using equal number of alternatives and they chose to use distractors such as 'all of the above' and 'none of the above' which are considered ineffective as the principles of good test construction stipulate.

Table 10C: Test Item Construction Practice

SN	*Items measuring test construction*	*N*	*Min*	*Max*	*M*	*SD*
1	I arrange items based on difficult level	132	2.00	5.00	3.86	0.889
2	I avoid using negative phrasing when necessary	132	2.00	5.00	3.70	0.880
3	I use both objective and subjective test formats during exam	132	2.00	5.00	4.11	0.879

Table 10C here above presents findings related to test construction practice among 132 respondents. As it can be seen from the table, teacher informants of the present study stated that they arranged items based on difficult level (3.86, S.D=0.88). The teachers reported that they avoided negative phrasing when necessary (M=3.70 SD=0.88) and they use both objective and subjective test formats during exam (M=4.11, S.D. =0.87). These statements which were rated the mean indicated that a considerable proportion of teachers seemed to use both objective and subjective test formats, and they avoided using negative phrasing. They claimed that they arrange items based on difficult level.

The interpretation of the results data presented in table 10C would give us good understanding of the scenarios if it anchors particularly on examining the relationship between the mean (M) scores and the standard deviation (SD). For instance, the finding for the item 'Arranging items based on difficulty level' was resulted in high mean score and moderate standard deviation (M=3.86, SD=0.889). This result indicates that respondents generally agree with the practice of arranging items based on their difficulty level. The SD suggests moderate variability in responses, indicating that while most respondents agree, there are some differing opinions. In the same token, Item 2

'Avoiding negative phrasing when necessary' with (M=3.60, SD=0.88), in the same table above, reflects a positive inclination towards avoiding negative phrasing though it is slightly lower than the result found for item 1. The SD is almost similar, indicating a consistent level of agreement among respondents, but with a bit more variability compared to item 1. Furthermore, Item 3, 'Using both objective and subjective test formats' (M=4.11, SD=0.879) has the highest mean score, suggesting strong agreement among respondents regarding the use of diverse test formats. The SD is comparable to the other items, indicating that while there is strong consensus, there is still some variability in opinions. In summary, respondents showed a favorable attitude towards the practices of arranging items by difficulty levels, avoiding negative phrasing, and using varied test formats, with a consistent level of variability in their responses.

Despite the above assertion, there are some exams in which the arrangement of item was not properly written. The test content analysis showed that there are some example types which are not arranged in the right manner. That is, the exam items are not arranged based on difficulty level (True/False followed by matching and multiple choices). The samples exams '*Exam R, T, and O from the College of Natural Science Education, Exam D, and E, G from College of Educational Sciences and Exam W, Z, Y from Social Science Education'* reveals the failure in following the principles of test construction.

At it can be seen from the sampled text analysis, there are some limitations noticed in test construction. Some teachers did not seem to bother about the order of the exam items; they did not put items from easy to complex. In addition to this, during the preparation of matching items, they make the premise and the reference items equal in number which is against psychometric standards. These might happen due to lack of awareness or negligence which really affects the performance of the students. As it can be seen from the table below, teachers stated that they set points/marks for individual parts of a test(M=2.39, SD=0.79) and they allot time for individual parts of a test/exam (M=1.99, SD=0.45).

Table 11*:* Test Weighting Practice

No	Items Measuring Test Construction	N	Min	Max	M	SD
1	I set points (marks) for individual parts of a test.	132	1.00	4.00	2.39	0.79
2	I allot time for individual parts of a test.	132	1.00	3.00	1.99	0.45

Table 14 presents data regarding test construction practice, focusing on respondents' perceptions and practice related to setting points and allotting time for parts of a test. The key metrics provided are the mean (M) and standard deviation (SD). In analysis of data obtained for item 1, 'Setting points (marks) for individual parts of a test (M=2.39, SD=0.798) reported that the respondents looked to be somewhat neutral or slightly disagreeing with the statement regarding the practice of setting marks for individual test parts. This suggests that the practice may not be commonly implemented. Standard Deviation (0.79) suggests moderate variability in the responses. A high SD value indicates that there is noticeable diversity in opinions, with some respondents agreeing more strongly with the practice of assigning points than others. This variability could suggest differences in the familiarity or valuing of this practice among different respondents or contexts.

The result of Item 2 'Allotting time for individual parts of a test' (M=1.99, SD=0.453) indicates that respondents generally tended to disagree with the practice of allotting time to individual parts

of a test, suggesting that this practice is not widely emphasized or utilized. The SD of item 2 is lower than that of item 1, indicating that responses are more clustered around the mean. This shows a stronger consensus among respondents regarding this item, with less variability in opinions. Most respondents seem to uniformly disagree on the importance of allotting time for individual parts of the test. In conclusion, the findings suggest that both practices are not highly regarded or commonly implemented in the context of test construction among the sample respondents. The differences in variability of SD indicate disparity in the levels of consensus on these practices, with greater disagreement observed on the practice of setting marks compared to a more disagreement on allotting time.

From the document analysis, it was also noted that teachers did not allot time and marks when constructing exams or tests. The following are sample representative from the document analysis:

Exam T, Q, O from College of Natural Science Education
Exam G, D, B from College of Educational Sciences
Exam Y, W, X from College Social Science Education

As the findings here above indicate, teachers looked to undervalue allotting time and marks for the exam they constructed. Even those teachers who allot time for the exam might not set appropriate time for each item. The following table clearly shows the case in point. Teachers were asked how they set time for each multiple choice item in social science courses which did not require any mathematical operation. For this item, the response was tabulated as follows:

3.5. Exam time allotment concept

Unless there is a mathematical operation, a normal multiple choice item is supposed to be answered within a minute. In an interview held with a participant, it was expressed that exam times are arbitrarily allotted. In line with this thought, Feyissa in said:

> *Teachers allocate time carelessly not scientifically; they do not keep the right order of item types from simple to complex; they do not prepare clear directions or instructions for examinees; they do not state mark/s for each item; they do not prepare tests being in group or committee but they may share portions of the exam. Teachers are weak in doing item analysis."*

In addition to this, Gezahign added what seemed to be a puzzle: "*Let alone time and marks, some exams do not have directions."* This finding shows that teachers looked to have less understanding about time allotment to assign marks or time for the exam. The result of Ngaji et al's (2021) study appeared to ascertain this point. These researchers argued that teachers often times ignored the process of construction and the psychometric properties of the test instruments. This might imply that professionals lack adequate pedagogical skills. To sum up, the teachers need give times skillfully to the items or tests they construct; perhaps, the skill gap observed in test construction in general and time allotment to test items might to be filled through training.

3.6.Test administration practice at KUE

Test administration refers to the comprehensive process of organizing and overseeing the process of assessments to ensure a fair and standardized testing environment. It encompasses various tasks, including preparing the testing location and materials, implementing necessary accommodations for individuals with disabilities, providing clear instructions to participants, and monitoring the test session to prevent cheating and address any issues that arise. Administrators are responsible for timing the test appropriately, ensuring security of the materials, and later processing and scoring the assessments accurately. Additionally, evaluating the overall administration process helps identify areas for improvement, ultimately contributing to the validity and reliability of the testing outcomes(Lane et al., 2016; Stiggins, 1999). The practice of test administration at KUE can be viewed from students' responses to the twelve items presented in Table 12 below.

Table 12: Students' reflection on test administration

No	Items	Mean	SD
1	Proctors/teachers show negligence during testing	2.66	1.39
2	Teachers engage in their personal duty while administrating tests or exams	2.86	1.30
3	Teachers overlook controlling students to get access prohibited materials or handouts	2.71	1.38
4	Exams may not be properly secured	2.39	1.40
5	Students may have unauthorized access to duplication or test rooms	2.54	1.35
6	Teachers may be absent from the testing room	2.35	1.37
7	During examination, students copy from their short notes	2.94	1.37
8	Students may possess electronic devices during examination	2.63	1.54
9	Students copy answers from others	2.78	3.57
10	Proctors execute irregularities measures during exam	2.76	1.38
11	Test coordinators remain visible in the testing environment	3.06	1.36
12	Teachers tend to ignore giving feedback to students after the exam	3.00	1.35
	Test administration (grand mean).	**2.75**	**.94**
	Valid N (listwise)	**259**	

As it can be seen from the table above, the study revealed that teachers have shown negligence during testing (Mean=2.66; S.D. =1.39); exams may not be properly secured (Mean=2.39; S.D. =1.40); teachers may be absent from the testing room (Mean=2.35; S.D. =1.37); students have been possessing electronic devices during examination (Mean=2.63; S.D. =1.54). In summary, the grand mean of exam administration was found to be 2.75 with the standard deviation of 0.94 which was below the mean. This further indicates that exam administration process might have some limitations. The majority of the students' response on the measurement scale showed that the mean score was found to be below three out of five, indicating that most items were rated below the mean. This further revealed that the administration of tests appeared to be below the expectation of the sample students.

In addition to the students' response, teachers in the interview sessions also noted that there was some poor administration of tests or exams at KUE. In this regard, Gezahign reported "*Some teachers do not control students at the exam."* Similarly, Kedir noted the points below:

> *When we come to test/exam administration, there are no such problems at KUE. Most teachers properly administer tests/examinations properly. But there are few*

teachers who do not respect the protocols that are expected to be considered during exam administrations.

Gedam also pointed out that *"The quality of exams is sometimes compromised when the large number of students assigned to in a single room while administering them."* Furthermore, Teshale forwarded that he noticed "*so many critics from students that cheating is rampant using AI. All cell phones must be prohibited in exam rooms (practice)."*The excerpts that some teachers might not play their responsibilities properly, but most of the teachers have carried out their responsibility in the right manner. As per the excerpts above, the findings have shown mixed results, students tended to exhibit a sort of malpractice during exam. It was reported that they used Artificial Intelligence for cheating and teachers were unable to control this irregular acts. The following histograms have shown the results more clearly.

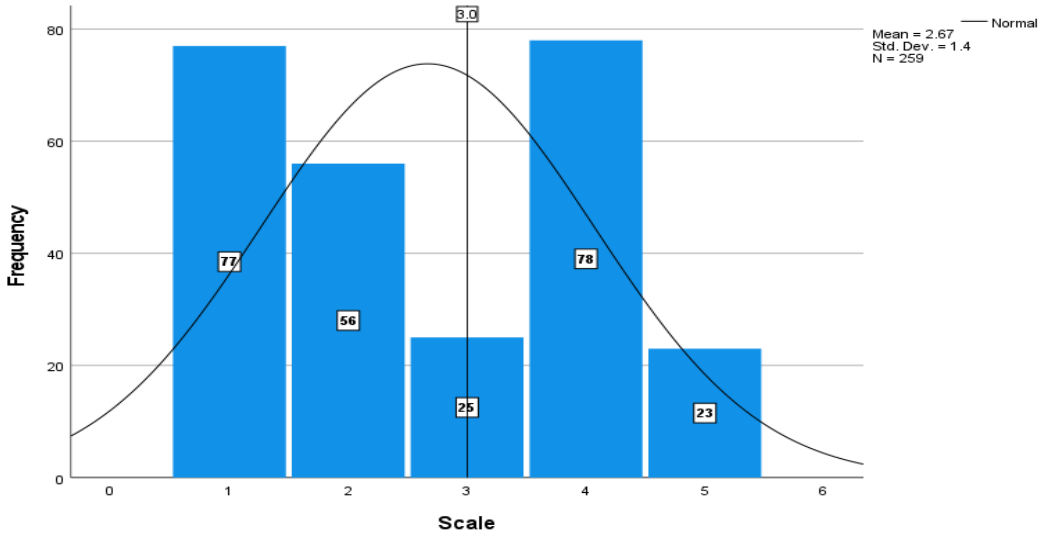


Figure 3: Invigilators' behavior while handling administration of assessment

As shown in the histogram above, the students' responses were unevenly distributed. Of the 259 students who voluntarily participated in the investigation into the practice of assessment administration at KUE, nearly equal number students reported strongly disagree (77 replies) and agree (78 replies) regarding instances of negligence of invigilators in relation to the test administration regulations established by the University. The illustration illustrates that there is a problem with following stipulations provided in the guidelines and regulations meant for administration, as reported by the students.

The majority of the students indicated that teachers showed negligence on the administration of tests which casts doubt on teachers' commitment in controlling student misbehaviors and mischief. This implies that there is a threat to academic integrity at KUE. The number of students who showed disagreement was found to be 56 (22%). This indicated that teachers might be preoccupied with their personal tasks during assessments and they may not strictly monitor students' attempts to access to the prohibited materials. The inattentiveness of invigilator teachers during assessments is believed to be linked to their engagement in personal activities while assessments are conducted in the classrooms. The students' responses appeared to confirm this presumption.

Table 13: Exams Security

Statement	Strongly disagree	Disagree	Undecided	Agree	Strongly agree	Total
The Responses of the students regarding exam security	37.5	24.7	8.5	19.3	10.0	100%

N=259

Students were asked whether exam security measures were properly observed. A significant number of respondents, 97 out of 259 (37.5%), reported that they strongly disagreed with the idea that exam security was adequately maintained. Additionally, 64 informants (24.7%) expressed disagreement regarding the perception that exams might not be properly secured. In contrast, the combined percentage of those who agreed and strongly agreed with the security measures became only 29.3%. This indicates that exam security at KUE may not be perceived as a high priority.

Table 14: The Likelihood of students' unauthorized access to test papers

Statement	Strongly disagree	Disagree	Undecided	Agree	Strongly agree	Total
The responses of the students on teachers' absenteeism during invigilation	28.2	30.1	10.8	21.2	9.7	100%

N=259

One of the responsibilities of teachers is to ensure that duplication or invigilation rooms are not accessible to unauthorized individuals. In this regard, 80 participants (30.9%) expressed suspicion that unauthorized individuals had accessed these rooms. In contrast, a majority of 58.3% (151 out of 259 informants) believed that duplication or invigilation rooms were secure and inaccessible to unauthorized personnel. As presented in the table above, there is a potential threat to exam security at KUE, although the degree of this threat may be perceived as minor.

Table 15: Punctuality of invigilators during test administration

Statement	Strongly disagree	Disagree	Undecided	Agree	Strongly agree	Total
The responses of the students on teachers' punctuality on duty of invigilation	36.3	29.0	8.1	16.6	10.00%	100%

N.=259

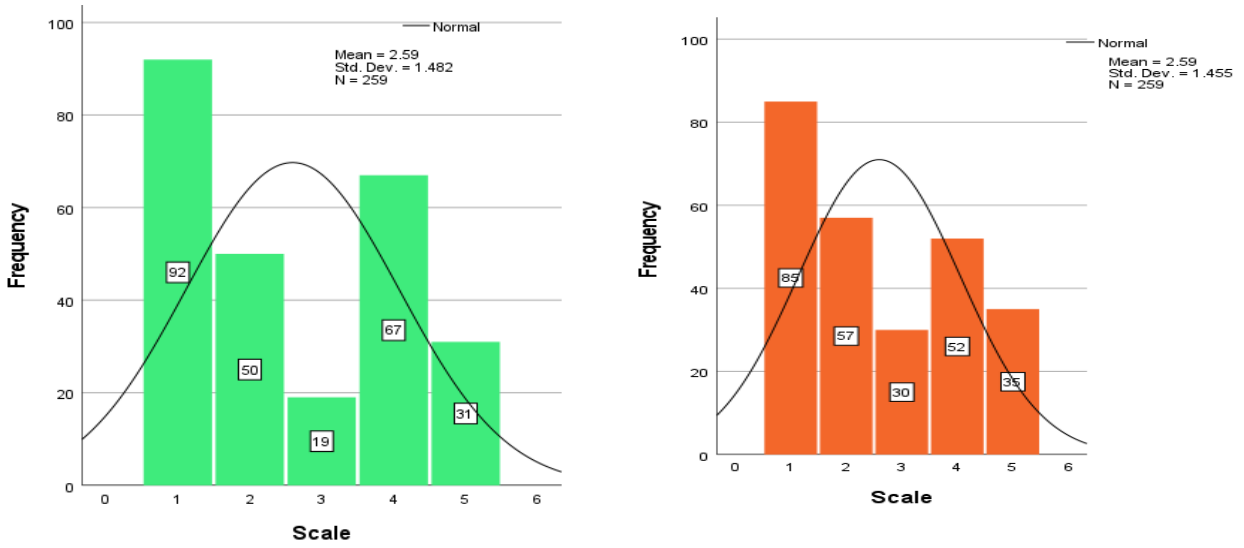


Figure 4: Academic dishonesty

Teachers' punctuality during invigilation duties is believed to significantly challenge the integrity and fairness of examinations. Unpunctuality can lead to compromised examination security, disruption of exam routines, and a negative impact on students. Among the total informants, 65.3% observed that invigilators at KUE felt responsible and did not miss their invigilation duties. In contrast, 26.65% reported that teachers failed to show up for invigilation. Students were asked to confirm their behaviors whether they engaged in the practice of carrying prohibited electronic devices during examinations and copying answers, the majority of them (71.4%) admitted to bringing forbidden devices into exam rooms, and confessed to cheating on exams, thus committing academic dishonesty.

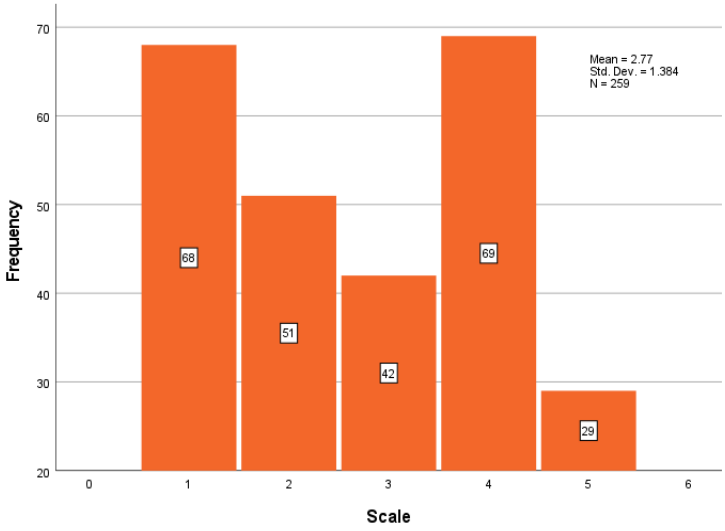


Figure 5: Irregularity measures invigilator executed during exam period

The responses of the students regarding the commitments of invigilators in enforcing irregularity measures during exams showed little variability. As illustrated in the histogram above, the

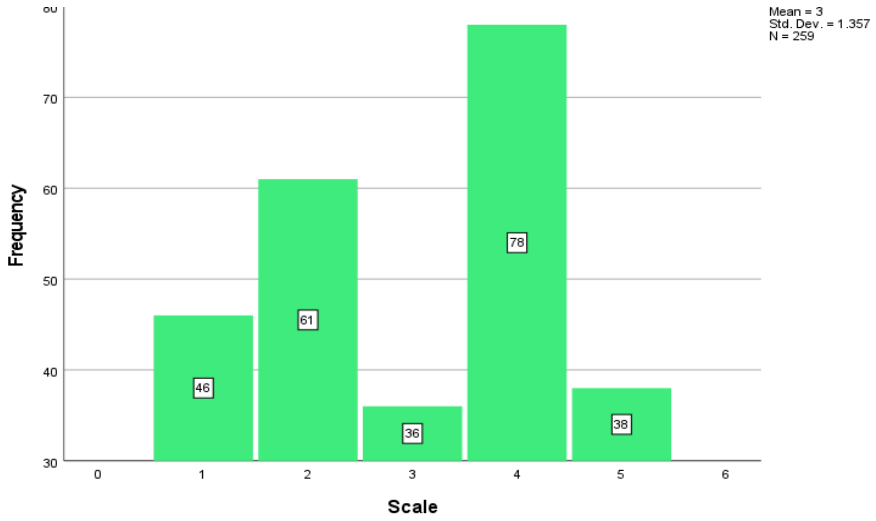


Figure 6: Teachers' practice of giving feedback

As depicted in the histogram above, more than a quarter of the students who participated in the study reported that teachers did not provide feedback after the exam.

3.7.The relationships among variables of the study

The correlation of teachers' perception to the overall test practice, test construction and administration was presented as follows:

Table 16: Teachers' overall perception of test practice, construction and administration

		Overall Test Practice	Perception	Test construction	Test Administration
Overall Test Practice	Pearson Correlation	1			
	Sig. (2-tailed)				
	N				
Perception	Pearson Correlation	-.062	1		
	Sig. (2-tailed)	.482			
	N	132			
Test construction	Pearson Correlation	.069	-.003	1	
	Sig. (2-tailed)	.429	.975		
	N	132	132		
Test Administration	Pearson Correlation	-.081	.037	-.030	1
	Sig. (2-tailed)	.356	.673	.732	
	N	132	132	132	

Table 16 presents the Pearson correlation coefficients and significance levels for the relationships between teachers' perception and overall test practice, test construction practice, and test administration practice. The table depicts that there is a very weak negative correlation of perception with overall test practice (r=-0.062), which is not statistically significant (p=0.482). This suggests that teachers' perceptions of test practice do not significantly relate to their overall

test practice. Perception-test construction correlation analysis presented in the table shows a weak positive correlation with overall test practice (r=0.069), which is also not statistically significant (p = 0.429). This indicates that there is no meaningful relationship between test construction and overall test practice.

The correlation with perception is negligible (r=-0.003) and is not significant (p = 0.975), suggesting no relationship between teachers' perceptions and test construction. The test administration correlation also exhibits a weak negative correlation with overall test practice (r=-0.081), which is not significant (p = 0.356). This implies that there is no significant relationship between test administration and overall test practice. The correlation with perception (0.037) is also weak and not significant (p = 0.673), indicating no meaningful relationship. Furthermore, the correlation with test construction (r=-0.030) is negligible and not significant (p=0.732), further suggesting no relationship. Overall, the findings indicate that there are no significant correlations among the variables of overall test practice, perception, test construction, and test administration. The weak correlations suggest that teachers' perceptions and practices in these areas may be largely independent of one another.

4. Conclusion

The findings of the study show that a considerable proportion of teachers seemed to have a reasonable perception about assessment despite the fact that there are slight differences among individuals. This further indicates that the majority of the instructors appeared to have good understanding regarding the concept of assessments. The study accentuates that teachers with the highest qualification have good perceptions on preparing and administering assessments. The study also reveals that there are differences among teachers in perceptions of assessments along with their specialization.

The majority of the students and teachers informants reported that assessments were carried out frequently at KUE. With regard to the assessment tools used, the teachers stated that they frequently employed exams, tests and assignments. Students confirmed that assignments and tests were the most frequently used tools of assessments at the University. This implies that the other assessment tools such as peer assessment, self-assessment, recap exercise, observation, attendance, term paper, portfolio assessment, discussion, group reflection, participation, presentation, project and quiz might not be duly valued by teachers. The other gap observed in the practice of test construction was that teachers did not conduct item analysis (item difficulty level, discrimination index and distractor analysis). It was also noticed that they were reluctant to get their assessment items reviewed by colleagues. Furthermore, the findings of the study revealed that teachers neither prepared test blueprints or table of specifications, nor they were consistent in constructing equal number of alternatives in preparing multiple choice items. Additionally, while constructing tests, teachers seemed to frequently employ phrases like ‘all of the above’ and ‘none of the above,’ which are considered to be poor distractors

Looking at the assessment construction practice from the perspectives of item types, a considerable proportion of teachers appeared to include both objective and subjective test formats. Nevertheless, the study depicts that teachers had gaps in properly sequencing the test items from simple to complex. Teachers also made number of premises and responses equal in the case of matching

items. The results of the study further showed that a great proportion of teachers did not seem to set time and marks for each part of test items. In many exams papers reviewed, the general directions including cover pages were missed.

Besides, the students revealed that teachers have shown negligence during testing: exams might not be properly secured, some invigilating teachers were found to be absent or late and even if they showed up, some were not vigilant to academic honesty – it was reported that students used Artificial Intelligence to do the tests. These all might be observed due to lack of awareness or negligence which really affects the performance of the students.

Generally, the results of the study revealed that most of the teachers have carried out their responsibility properly. Teachers' perception about assessment has negative but negligible relation to the overall practice of test construction and administration practice. Thus, though teachers of KUE have good perception about assessment including continuous assessment, their practices on the overall test conditions, test construction, and test administration did not seem at the expected levels. In summary, exam administration process might have limitations.

5. **Acknowledgements**

The research team is grateful for the grant support given from Kotebe University of Education. The team is indebted to the teachers and students for their being volunteered to give genuine data.

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

References

AED/BESOII. (2006). *A Concise Manual for Developing and Implementing Continuous Assessment in Teacher Education Institutions and Primary Schools of Ethiopia, Addis Ababa*.

Ajzan, I. (1991). Attitude, Personality and Behavior, Milton Keynes, Open University Press.

Alsubaie, M. A. (2015). Examples of Current Issues in the Multicultural Classroom. *Journal of Education and Practice*, *6*(10), 86–89. www.iiste.org

Asamoah-Gyimah, K. (2022).Influence of Knowledge of Assessment on Test Construction Skills among Tutors of College of Education in Ghana. *American Journal of Education and Practice,* 6(3), 60 – 71

Bekalu Atnafu (2007). An Evaluation of the Implementation of Continuous Assessment (CA): The Case of St. Mary's University College. *Proceedings of the Fifth National Conference on Private Higher Education Institutions (PHEIs) in Ethiopia.*

Birhanu Moges.(2004). *Teacher's assessment of students' Performance in selected high schools of Arsi Zone with emphasis on continuous assessment.* Unpublished MA Thesis AAU. Faculty of Education. Addis Ababa.

Brookhart S. (2002). What will Teachers Know about Assessment and how will that improve

instruction? In Lissitz. R. and Schafer, W. (eds). *Assessment in educational Reform*: Both means and ends. Allyn and Banckon.

Creswell, J. W. (2007). *Qualitative Inquiry and Research Design: Choosing among Five Approaches.* SAGE Publications.

Dornyei, Z. (2007). *Research methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies*. OUP.

Ebel, R. L., & Frisbie, D. A. (2009).*Essentials of Educational Measurement.* 5thed. New Delhi: PHI Learning Private Limited.

Ebel, R. L. (1979). *Essentials of Educational Measurement.* Prentice-Hall, Inc. Englewood Cliffs.

Flick, U. (2002). *An Introduction to Qualitative Research*. London, SAGE Publications

Gronlund, N.E. (2003).*Assessment of student's learning Achievement.* 7thed. New York:Macmillan Publishing Company

Hopkings .K.D. and Stanley, J. C. (1981). *Educational and Psychological Measurement and Evaluation.* Prentice-Hall, Inc. Englewood Cliffs.

Kennedy, C. and Kennedy, J. (1996). *Teacher attitude and Change implementation systems*, 24(3).
KMU (2018). .*Student Assessment Implementation Guideline*. Unpublished material.

Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of Test Development*, Second Edition. Retrieved from: https://doi.org/10.4324/9780203102961

Linn, R.L., andGronlund, N.E. (1996).*Measurement and assessment in teaching,*6thed. N.J.: Prentice Hall.

Mehrens, W.A. and Lehmann, I.J. (1991). *Measurement and Evaluation in Education and Psychology 4th ed.* New York: Holt, Rinehart and Winston.

MoE. (2003). *The Teacher Education System Overhaul (TESO) Programme.*

Muluken, A. (2006). *Teachers' Perceptions and Practices of Continuous Assessment in Selected Government First Cycle Primary Schools of Addis Ababa.* MA Thesis. Addis Ababa: Addis Ababa University.

Nitko, A. J. (2006). *Educational Assessment of Students.*4th ed.. Prentice Hall.

Nitko, A.J. (2004). *Educational assessment of students*4thed. Upper saddle River, N.J: Prentice Hall.

Ngaji, O. O. , Veronica , Eduwem, J. & Nsor, J. A.(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. *British Journal of Education*, 9(11), 54-62.

NOE. (2004). *Guidelines for continuous Assessment*, Addis Ababa, Ethiopia.

Onuka, A. (2006). *Continuous Assessment as an Instrument of Achieving Learning Objectives.*

Peshal, K. (2020). Key considerations in test construction, scoring and analysis: A guide to pre-service and in-service teachers. *International Journal of Research Studies in Education*, 9(5), 15-24.

Puhl, C. A. (1997). Develop, Not Judge: Continuous Assessment in the ESL classroom. *Forum Online*, *35*(2).

Ramsden, P. (2003). *Learning to Teach in Higher Education.*2nd ed.. Routledge Falmer.

Seyoum, T. (2006). *The Practice of Assessment Vis-À-Vis Mathematics Education: The Case of Dilla College of Teacher Education.* MA Thesis. Addis Ababa: Addis Ababa University.

Shaaban, K. (2001). Assessment of Young Learner. In *Ancker. W (2001) English Teaching Forum.* Washington, DC, USA.

Shepard, C. . (2006). The Role of Assessment in a Learning Culture. *Educational Research*, *20*(7), 4–14.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, *18*(1), 23–27. https://doi.org/10.1111/j.1745-3992.1999.tb00004.x

Teshome Demisse. (2001). Assessment of Pupils' Achievement in the teaching of English: Training and Practice in Focus. *The Ethiopian Journal of Education*.

Torrance H. (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.

USAID AED (2009). *Educational policy in Developing World*. Eq Review Vol. 1.

Wiersma, W. and Jurs, S.G. (1985). *Educational Measurement and Testing.* Boston: Allyn and Bacon, Inc.

Yalew Endawek. (2008). Problems of Teacher-Made Tests: A Prime Concern for Quality of Education. St. Mary's University. *Proceedings of the Sixth National Conference on Private Higher Education Institutions (PHEIs) in Ethiopia.*

Yao, Y.(2015). Teacher Perceptions of Classroom Assessment: A focus group interview. *SRATE Journal*, 24(2).