

Assessment Quality at two Ethiopian Universities: Requirements versus Practices

Firdissa Jebessa Aga^{1*}

¹ Institute of Educational Research, Addis Ababa University

* Corresponding Author: firdissa.jebessa@aau.edu.et

(Received: 07/12/2022; Accepted: 08/12/2022; Published: 09/06/2022)

DOI: <https://doi.org/10.37906/real.2022.2>

Abstract: The purpose of this study was to examine the extent to which learning assessment practices fulfill assessment quality requirements. Two sorts of questionnaires were used to collect data from teachers/researchers, and MA/MSc and PhD students of Education and English Language at two Ethiopian universities. The results have shown that assessment tasks, contents, tools and procedures at the universities hardly fulfilled some desirable assessment requirements including making assessment tasks and contents free from all kinds of biases, mirroring the required skills/knowledge in real life, empowering teachers to assess what they value, encouraging deep approaches to learning, and guaranteeing learners' awarded grades to represent the levels. Some assessment tools, procedures, and practices also failed to fulfill assessment quality requirements. It has, therefore, been recommended that all parties should make utmost conscious efforts to ascertain alignment of assessment practices, procedures and contents with available theories/requirements on learning assessment quality.

Keywords: Learning Assessment, Assessment Quality, Assessment Requirements, Assessment Practices

1. Introduction

Learning assessment is a driver for deep learning. It determines what, how and why to learn. As, Ibarra-Sáiz, Rodríguez-Gómez, and Boud (2021) indicate, “[l]earning is not just determined by the curriculum, but by how it is assessed”. Consequently, assessment is considered as a catalyst, an inevitable, integral and critical component/part of any education system and/or the entire teaching and learning process (James, Mcinnis, & Devlin, 2010; Paniagua, & Swygert, 2016). It acts as a central catalyst to good teaching and is inevitably a key component in facilitating students’ learning with understanding (Donovan & Bransford, 2005 & Edwards, 2013).

Consequently, the quality of learning assessment has fundamentally attracted the interest of stakeholders who have different purposes. “Quality assessment practices” are therefore, needed “to include a consideration of the ‘fitness for purpose’ of an assessment task/activity as well as consideration of the characteristics of the learners themselves, so that best choices are made regarding the nature and timing of assessment” (Edwards, 2013; Gardner, 2006; Harlen & James, 1996). As to this author, doing this well allows for the confident use of assessment data by teachers in their decision making, and in turn leading to the improvement of current and future teaching and learning.

Learning assessment quality deals with fitting the purposes of “different audiences” (Brown & Knight, 1994) and/ or “several actor groups” (Luoma, 2001). Assessment, therefore, should reflect the simultaneous demands of multiple audiences and/or actor groups for multiple purposes, among others: test takers, students, score users, teachers, the governments, university management, employers, financing bodies, funding stakeholders, and the society at large (Brown & Knight, 1994; Luoma, 2001; OET, 2017).

As Brown and Knight (1994) further indicate, “[a] student may well want feedback to enable him or her to work on points in need of attention rather than to keep practicing points of strength”. For Bryan and Clegg (2006) “... assessment which supports learning is flexible and takes into account the need for individuals to make sense of feedback in the context of their own experiences”.

As Brown and Knight (1994) went on indicating that:

Employers may want accounts of what the student can do and the student him or herself, towards the end of his or her studies, will also want such a summative verdict. University management will want information about how much students have gained from their undergraduate years, but mentors, to take another audience, will initially need to know what the student may be trusted to do and where the main points for care and development lie.

OET (2017) also has the view that:

Teachers need to check for student understanding, and parents, students, and leaders need to know how students are doing overall in order to help them successfully prepare for college and work. In addition to supporting learning across content areas, technology-enabled assessments can help reduce the time, resources, and disruption to learning required for the administration of paper assessments. Assessments delivered using technology also can provide a more complete and nuanced picture of student needs, interests, and abilities than can traditional assessments, allowing educators to personalize learning.

Through embedded assessments, educators can see evidence of students’ thinking during the learning process and provide near real-time feedback through learning dashboards so they can act in the moment. Families can be more informed about what and how their children learned during the school day. In the long term, educators, schools, districts, states, and the nation can use the information to support continuous improvement and innovations in learning.

This shows that there are different purposes for assessment. The purposes, nonetheless, are “neither separate nor entirely compatible” (Brown & Knight, 1994). The different purposes for assessment lead to searching answer to a question: what is a quality assessment? Quality assessment for Ainslee (2018) “basically focuses on the targeted areas with complete precision”. He went on describing that, assessment in the education industry should have content validity, reliability, generating interest by the student, and consequential relevance. Redecker and Johannessen (2013) have the view that “the quality of teacher-made tests would improve greatly if they were not administered immediately but given to a few colleagues for review first”. This, then increases validity, reliability, interest, and relevance aspect of the test.

The validity of the test content for Ainslee (2018) deals with the test content to “be highly organized and should come across as clear and simple to the candidates attempting the test. It should not comprise of faulty language or spelling defaults. The content should be in accordance with the subject that is being assessed and should not be out of the syllabus or topic”.

Reliability with reference to assessment signifies that each and every aspect of the assessment has a measurable outcome, and the quality of being accurately measured without the buildup of any flaw.

Ainslee (2018) further explains that generating interest by the student deals with “the reason why tests should be objective in nature. Subjective tests are lengthy in nature not even generating interest of the

teachers, leave alone the students. So, assessments should be explicit and creative which does not give a sense of boredom to the candidates”.

Finally, consequential relevance deals with the reason for conducting an assessment, which requires a lot of time, dedication, and resources. This is because, “nobody would want so much of hard work to go in vain”. By implication, assessment result should be so exact so that it can be used as a tool to compare and analyze the data for future reference of the candidate’s performance (Ainslee, 2018).

The interests in the quality of learning assessment in higher education by stakeholders have come with due recognition of: 1) the fact that the quality of higher education graduates depends on what they have effectively learnt and authentically assessed; and 2) the need to account for the politics of accountability. The politics of accountability can be achieved by assessing quality outcomes of higher education, guaranteeing fair assessment practices responsive to human diversity, assuring success in higher education, and readiness to facing the technological future of higher education (Messick, 1999). The concerns as well as the recognitions in turn have caused moves, for instance in UK, the need for “new kinds of assessment designed to assess ‘key skills’, ‘transferable skills’, ‘generic skills’ or ‘graduate attributes’ rather than assessing solely the acquisition of knowledge” (Bryan & Clegg, 2006). This then has required: 1) revitalizing taxonomies of educational objectives in a way high-level learning are assured (Bryan & Clegg, 2006); 2) re-specifying curricula in terms of learning outcomes; and 3) looking for diverse knowledge, skills, and abilities to “conducting quality student learning outcomes” (Horst & Prendergast). Alongside, the assessment change drivers, assessment regulations and external quality assurance demands are constraining assessment options, driven by worries about requirements, reliability, and plagiarism (Bryan & Clegg, 2006).

Cognizant of the fast-changing assessment practices and contexts, Ethiopia has put in place curriculum requirements, and assessment modalities in which expected learning outcomes (Özturgut, 2011) (LO) are pre-defined and stated in national and institutional curricula.

National and institutional policies and regulations also plea for employing appropriate learning assessment methods to ensure effective implementation of its programs.

Higher education Institutions (HEIs) Ethiopia have, therefore, given a serious consideration to competency-based learning and assessment related to LOs (AAU, 2014a; 2014b). Instructors in Ethiopian HEIs are, therefore, acquainted with competency-based learning and assessment, and modularization through Higher Diploma Program (HDP). HDP is a practice-based training program for teacher educators at higher education institutions in Ethiopia. Basically, HDP has a one-year duration whereby teachers attend two hours discussion classes for two days per week, supplemented with additional classroom observations and secondary school visits for a week or two. Recently, nonetheless, Addis Ababa University has customized the Program to its context by reducing the duration to a maximum of intensive four months by integrating different competencies, truncating redundant topics, and arranging intensive schedules (Firdissa, 2021).

Assessment is an integral part in both cases (the National framework and that of Addis Ababa University). Whereas the National framework on HDP has four modules dealing with ‘Reflective Teacher Educator,’ ‘Developing Active Learning,’ ‘Improving Assessment,’ and ‘Action Research, Making a Difference’ (MoE 2006); that of AAU has five modules dealing with Understanding Higher Education, Modularization and Modular Curriculum, Managing Learning and Assessment, Subject Area Teaching, and Action Research and Field-based Learning (AAU, 2014a).

Particularly, Module three on Managing Learning and Assessment presents basic elements of assessment with assessment rationales, principles, methods, importance and grading procedures. Instructors are, therefore, aware of assessment practices and tenets.

AAU (2014) suggests that a variety of assessment methods should be designed to satisfy all LOs. In designing or redesigning modules, it is, therefore, vital to identify and reach at a consensus by instructors and academic leadership on appropriate parameters of assessment; and to decide which can be left to individual teachers or subject coordinators. Concerning the general provisions on examinations, AAU (2019, Article 82, No. 821) indicates that:

Student learning shall be assessed on a variety of ways/continuous assessment in the form of tests, assignments, presentations, etc. to determine the final grade earned. This shall account for 50% of the total module/course grade. The remaining 50% shall be allotted for a final exam conducted at the end of module/course delivery. Instructors shall monitor the student's academic performance by keeping track of records.

In the same vein, AAU (2014) presents the following points concerning the assessment of the modular curriculum:

- 1) performance of learners in a module should be evaluated in relation to the achievement of the modular-objectives (criterion-referenced) rather than on competitive basis (norm-referenced) and normal distributions;
- 2) the old system of using the normal curve for determining grades should be replaced by initial planning of correspondence between number-grades and letter- grades while determining the latter;
- 3) failing grades for a module can be determined by learner performance below 60 percent of the total. it is suggested, however, that each instructor with the consultation of his/her department can modify the suggested grading scale;
- 4) assessment of student work should be continuous, valid, and reliable; and
- 5) there should be a meaningful and effective system of evaluating, revising, up- grading or phasing out academic programs.

Overall, many of the available national and institutional curricula documents and/or guidelines advocate that students achieve the requirements set out in criterion-referenced assessment modalities, master the LOs which are inherent within the modalities, and achieve learning with understanding. These in turn call for the alignment of competency-based outcomes and learning activities with assessment; and intentionally designing curricula around competencies with explicit, measurable, transferable LOs and integrate with instruction, and assessment. The needs for the said alignment, design and integration have come with the growing body of research into higher education assessment on which academics, leaders and policy makers can begin to build robust policy and practice decisions (Bloxham & Boyd, 2007). Equally, we are witnessing that universities are becoming accountable for the quality of their assessed graduates. Universities are no longer remote, ivory towers, and can no longer be regarded as diarchies whereby institutional autonomy and academic freedom seem to obscure accountability for inefficiencies including poor assessment practices.

Inherent within the heightened interests in assessment matters (globally, nationally as well as institutional) is a quest for assuring quality outcomes of higher education one of which can be achieved through learning assessment quality.

2. Problem Statement

Though the catalytic role of learning assessment as a driver for real learning; and as an inevitable, integral and a critical component part of any education system and/or the entire teaching and learning process have got due recognition than ever before; learning assessment is still challenged by a number of factors. Among others, increased class size, changing curricula, the need to support students better, declining resources, assignments and study time, 'modularization' and assessment, plagiarism, computer-aided assessment, declining student retention, specifications and assessment of new kinds of learning outcomes, paradigm shift, technological influences, increasing cohort size and the shrinking unit of resource, the changing of student body, and policy climate have pressurize HEIs to make changes in their assessment practice (James, Mcinnis, & Devlin, 2010; Paniagua, & Swygert, 2016; Donovan & Bransford, 2005; Brown & Knight, 1994; Luoma, 2001).

Equally, learning assessment practice lags well behind its equivalent in the school sector, relying largely on a limited range of tried (but not always tested) methods. It is dealt with in an ad hoc way and the situation is not mitigated by the 'amateur' status of many academics regarding assessment (Bloxham, & Boyd, 2007; Murphy 2006; Swann & Ecclestone, 1999; Ramsden 2003). Bloxham and Boyd's (Price, 2005) have indicated that teachers "learn the craft of assessment informally through being assessed [themselves] and through being part of a community of practice, but lack scholarship regarding assessment". As the same authors indicate, most teachers "have survived this approach to professional learning reasonably unscathed but it is not a recipe for enhancement; it provides no reliable route for ensuring that research on assessment reaches those doing the assessing."

As to my knowledge, quality learning assessment requirements vis-à-vis practice has never been a topic of research in Ethiopian HEIs. Equally, "...there has been little investigation into the effect of classroom-based assessment on instructional and learning practices" (Muñoz & Álvarez, 2009).

Despite the recent growth in interest towards enhancing the quality of graduates and being sensitive to answerability (accountability), "assessment in higher education remains under-conceptualized" (Bryan, & Clegg, 2006; Gerritsen-van Leeuwenkamp, Brinke, & Kester, 2017). Moreover, there is also lack of established practices to regularly review newly prepared teacher made-test as assessment tools. The case has been exacerbated with the complexity of quality conception in general and learning assessment quality in particular. The word quality is rarely defined, and is a more complex concept than traditional assessment requirements suggest – quality cannot be reduced to a set of easily quantified learning outcomes (UNESCO, 2001; Bryan & Clegg, 2006). This, therefore, calls for examining the extent to which learning assessment practices and procedures at the selected universities fulfill desirable quality requirements.

3. Objective of the study

The study had the purpose of gauging requirements versus practices on student learning assessment quality taking two Ethiopian universities as a case study. More specifically, the study intended to assess the extent to which learning assessment practices, and procedures at the selected universities fulfill desirable assessment quality requirements. The objective has the assumption that teachers and students are expected to have proper understanding of the link between student learning assessment requirements/theories and practices so as to enable them appreciate the what, how, and why of learning.

4. The Research Methodology

Quantitative data were collected using two sorts of questionnaires from 161 subjects (72 teachers and 94 students) at two purposively selected Ethiopian universities. For the sake of anonymity, the universities have been labeled as U1, and U2. Whereas U1 was selected based on its age and productivity in offering

postgraduate programs, U2 was selected for convenience purpose. The data were generated both from teachers and students using close-ended questions of the questionnaires. All the returned copies of the questionnaires were numbered as: TR1-72, and SR1-94, representing respectively teacher respondents and student respondents. They were entered into the Statistical Package for the Social Sciences (SPSS) software version 23, and analyzed, tabulated, interpreted, and discussed.

5. Results

166 copies of the questionnaires (from 72 teachers and 94 students) were filled and returned. The return rate was 87% from the dispatched 190 copies. 49 and 23 of the teachers were respectively from U1 and U2. Similarly, 53 and 41 of the student respondents were respectively from U1 and U2. Of the 53 student respondents from U1, 5 were PhD students from other universities studying at U1; and two of them indicated that they were not employees of any university.

6. Analysis of the Respondents

Student- and teacher- respondents were requested respectively to indicate their study programs and qualifications. Whereas 85 of the students and 71 of the teachers properly filled and returned, respectively 9 and 1 were missing systems as can be seen from Table 1.

Table 1: The students' study programs or levels and Teachers' Qualification

		Students' study programs		Teachers' Qualification	
		Frequency	%	Frequency	%
Valid	MSc/MA	52	61	19	27
	PhD	33	39	52	73
	Total	85	100	71	100
Missing	System	9		1	
Total		94	100	72	100

Table 1 shows that 52(61%) and 33(39%) of the students were respectively attending MSC/MA and PhD programs. It can further be seen from the Table that the majority (73%) of the teacher respondents had a PhD qualification followed by 27% master's holders.

5 copies of the student questionnaires, nevertheless, were jettisoned as they were not properly filled. In analyzing the data, therefore, 161 (72 from teachers and 89 from students) of the properly filled copies of the questionnaires have been used.

Requested to indicate their teaching/research experiences in years, all the teachers and 85 of the students responded properly whereas 4 was a missing system from that of the students.

Table 2: Respondents' teaching/research experiences in years

Experiences		Frequency	Valid Percent
Valid	Under 3	14	9
	3-6 Years	21	13
	7-10 Years	26	17
	Above 10 Years	96	61
	Total	157	100.0
Missing	System	4	
Total		161	

Table 2 shows that the majority of the respondents had teaching/research experiences of above 10 years. As can be seen from the Table, 96 (61%) of them had teaching/research experiences of above 10 years. A further separate frequency analysis for the same has shown that all those who had under 3 years of teaching/research experiences were student respondents. Of those who had above 10 years of teaching/research experiences, 85% and 43% respectively were teachers and students.

Teacher respondents were also requested to indicate their respective ranks and the results have been shown in Table 3.

Table 3: Teacher respondents' Ranks

	Rank	Frequency	Valid Percent
Valid	Associate Professor	13	18
	Assistant Professor	41	57
	Senior Lecturer	4	6
	Lecturer	13	18
	Assistant Lecturer	1	1
	Total	72	100

Table 3 shows that the majority (57%) of the respondents had the rank of assistant professorship, followed by 18% associate professorship and equally lecturer.

7. Requirements versus practices in learning assessment quality

The issue of requirements versus practices in learning assessment quality was explored using 21 closed items of the questionnaires. The reliability of the 21 items is .95. This is very high from statistical point of view. When seen for each of the items, all are in between .94 and above. All the items are, therefore, reliable for generating dependable evidences.

14 of the closed questions requested the extent to which assessment tasks and contents, and assessment tools and procedures used at the universities fulfill some theoretically desirable assessment requirements. The respondents were directed to circle "1" for Very little, "2" for Little, "3" for Medium, "4" for Greatly, "5" for Very greatly for each of the statements. The results have been presented in Table 4 and Table 5 (7 of them in Table 4, and 7 in Table 5).

Table 4. Practiced assessment tasks and contents vis-à-vis requirements

The extent to which Assessment tasks & contents at the university:	1		2		3		4		5	
	Cnt	%	Cnt	%	Cnt	%	Cnt	%	Cnt	%
1. Are made "fit for purpose" (i.e. learning-oriented)?	22	14.1	Cnt	%	59	37.8	28	17.9	8	5.1
2. Encourage 'deep' approaches to learning in the learners?	21	13.6	39	25.0	57	37.0	14	9.1	2	1.3
3. Mirror the required skills/knowledge in a real life/in the workplace?	19	12.1	60	39.0	46	29.3	20	12.7	5	3.2
4. Are made free from all kinds of biases that may disadvantage particular learner groups?	24	16.0	67	42.7	34	22.7	15	10.0	3	2.0
5. Are in line with the syllabi?	22	14.1	74	49.3	59	37.8	28	17.9	8	5.1
6. Guarantee learners' awarded grades to meaningfully represent the levels?	21	13.6	39	25.0	57	37.0	28	17.9	2	1.3
7. Empower teachers assess what they teach and what they value?	19	12.1	60	39.0	46	29.3	14	9.1	5	3.2

Table 4 shows that 74(49.3%) of the respondents rated the extent to which assessment tasks and contents at the universities were made free from all kinds of biases that may disadvantage particular learner groups as little. In the same vein, 67 (42.7%) rated the extent to which assessment tasks and contents at the universities mirror the required skills/knowledge in a real life/in the workplace, and empower teachers assess what they teach and what they value as little. Moreover, 60 (39%) rated as little the extent to which assessment tasks and contents at the universities encourage deep approaches to learning in the learners, and guarantee learners' awarded grades to meaningfully represent the levels.

Similarly, 59 (37%) of the respondents rated the extent to which assessment tasks and contents at the universities were made "fit for purpose" (i.e. learning-oriented), and were in line with the syllabi as medium. On the other hand, whereas all the counts for very greatly fall under 10, those for greatly fall between 14 and 28.

A descriptive analysis of the same data has shown that the average mean is 3.44. This indicates that the extent to which assessment tasks and contents at the universities fulfill some theoretically required assessment tasks and contents fall close to medium.

The seven questions presented to the respondents to rate on the extent to which assessment tools and procedures used at the universities fulfill some assessment requirements have also been presented in Table 5.

Table 5: Assessment tools and procedures used

The extent to which Assessment tools and procedures used	1		2		3		4		5	
	Cnt	%	Cnt	%	Cnt	%	Cnt	%	Cnt	%
1. Appropriately measure what they are supposed to measure?	16	10.3	67	42.9	51	32.7	18	11.5	4	2.6
2. Are made free from errors of measurement?	15	9.7	45	29.0	54	34.8	37	23.9	4	2.6
3. Lead to dependable scores?	14	9.0	47	30.3	64	41.3	23	14.8	7	4.5
4. Yield consistent results upon testing and retesting?	14	9.1	41	26.6	57	37.0	38	24.7	4	2.6
5. Lead to measurable outcomes?	25	16.6	45	29.8	52	34.4	25	16.6	4	2.6
6. Protect academic requirements?	28	18.1	44	28.4	48	31.0	27	17.4	8	5.2
7. Lead to mastery of learning outcomes?	21	13.4	67	42.7	38	24.2	24	15.3	7	4.5

Table 5 shows that 67 (42.9%) of the respondents rated the extent to which assessment tools and procedures used at the universities appropriately measure what they are supposed to measure, and lead to mastery of learning outcomes as little. On the other hand, 64(41.3%), 57 (37%), 54(34.8%), 52(34.4%), and 51(32.7%) rated as medium respectively the extent to which assessment tools and procedures used lead to dependable scores, yield consistent results upon testing and retesting, are made free from errors of measurement, lead to measurable outcomes, and appropriately measure what they are supposed to measure.

A descriptive analysis of the same data has shown that the average mean for the extent to which assessment tools and procedures used fulfill the listed requirements is 3.32. This indicates that the extent to which assessment tools and procedures at the universities fulfill some theoretically required tenets fall close to medium.

Five questions were also presented to the subjects to rate on how often teachers undertook some desirable theoretical tasks by indicating their choices by circling “1” for Never, “2” for Rarely, “3” for Sometimes, “4” for Often, “5” for Always for each of the statements. A descriptive result of the same have been presented in Table 6.

Table 6: Descriptive Statistics on the frequency that teachers perform assessment tasks

How often teachers at your universities:	N	Min	Max	\bar{X}	SD
1. Design assessment tasks that foster valued study habits?	160	1	5	3.01	.942
2. Use objective assessment tools?	152	1	5	3.53	1.042
3. Use assessment procedures as a means of improving teaching and learning?	155	1	5	3.45	1.088
4. Write Test specifications for test tasks?	158	1	5	2.44	1.202
5. Provide meaningful feedbacks that guide learning?	161	1	5	2.96	.935
Average	157	1	5	3.07	1.041

\bar{X} - Mean, SD – Standard Deviation

As Table 6 shows, whereas the means for using objective assessment tools, and using assessment procedures as a means of improving teaching and learning stand respectively 3.53 and 3.45 both of which fall between sometimes and often; the means for writing test specifications for test tasks, and providing meaningful feedbacks that guide learning are respectively 2.44, and 2.96 both of which are in between rarely and sometimes. Designing assessment tasks that foster valued study habits also has a mean of 3.01, which means sometimes. On average, the mean for the five questions was 3.07 with standard deviation of 1.041 indicating the frequency of practicing some theoretically required learning assessment quality tasks fall close to sometimes.

Furthermore, two questions were also presented to the subjects to rate the extent to which learning assessment was treated by all staff at the universities as an integral component of the entire teaching and learning process, and assessment practices at the universities matched with available theories/requirements on learning assessment quality. In both cases, alternatives were given as: 1. to a very little extent, 2. to a little extent, 3. to some extent, 4. to a great extent, 5. to a very great extent, and the results have been presented in Table 7.

Table 7: Learning assessment within Teaching learning process & its match with requirements

The extent to which:	1		2		3		4		5	
	Count	%	Count	%	Count	%	Count	%	Count	%
1. learning assessment is treated by all staff as an integral component of the entire teaching and learning process	14	8.8	71	44.7	14	8.8	52	32.7	8	5.0
2. assessment practices match with available theories/requirements on learning assessment quality	18	11.5	75	47.8	10	6.4	51	32.5	3	1.9

Table 7 shows that 75 (47.8%) and 71 (44.7%) of the respondents respectively rated the extent to which assessment practices match with available theories/requirements on learning assessment quality, and learning assessment is treated by all staff as an integral component of the entire teaching and learning process as to a little extent. On the other hand, 51 and 52 of them rated the same as to a large extent. Whereas insignificant number of the respondents rated the case as to a very large extent, 18 and 14 rated the same as to a very little extent.

A descriptive analysis of the same data has shown that the mean for the extent to which learning assessment is treated by all staff as an integral component of the entire teaching and learning process, and assessment practices match with available theories/requirements on learning assessment quality were respectively 2.81 and 2.66; and the average mean is 2.73. This indicates that the extent to which learning assessment is treated treating learning assessment by all staff as an integral component of the entire teaching and learning process, and matching assessment practices with available theories/requirements on learning assessment quality were just to some extent.

7. Discussions

The majority (61%) of the respondents who filled and returned the questionnaire on the issue of requirements versus practices in learning assessment quality had teaching/research experiences of above 10 years. Noticeably, no one from the teacher respondents had under 3 years teaching/research experiences. Also, of those who had above 10 years of teaching/research experiences, 85% and 43% respectively were teachers and students. By all means, the respondents had reasonable teaching/research experiences that would enable them to judge requirements versus practices related to quality of student learning assessment.

Similarly, the majority of the teacher respondents (73%) had a PhD qualification followed by 27% master's holders. When it comes to their rank, the majority (57%) of the teacher respondents were assistant professors, followed by 18% associate professorship and equally a lecturer-ship. This, therefore, shows that the teacher respondents had reasonable qualifications and ranks that would enable them to evaluate the practices of student learning assessment vis-à-vis available theories and requirements.

The fact that the reliability analysis of the data on requirements versus practices in learning assessment quality has given .95 signifies a very high result from statistical point of view. Analysis of frequency and descriptive statistics have also shown that the majority of the respondents rated as little the extent to which assessment tasks and contents at the universities were made free from all kinds of biases that may disadvantage particular learner groups (74 of them), mirror the required skills/knowledge in a real life/in the workplace (67 of them), empower instructors assess what they teach and what they value (67 of them), encourage 'deep' approaches to learning in the learners (60 of them), and guarantee learners' awarded grades to meaningfully represent the levels (60 of them).

Similarly, substantive number of the respondents rated the extent to which assessment tasks and contents at the university were made "fit for purpose" (i.e. learning-oriented), and were in line with the syllabi as medium. The result of the descriptive analysis of the same data has given 3.44 indicating the extent to which assessment tasks and contents at the universities fulfill some desirable requirements fall close to medium.

Moreover, whereas 67(42.7%) of the respondents rated the extent to which assessment tools and procedures used at the universities appropriately measure what they are supposed to measure, and lead to mastery of learning outcomes as little; 64(41.3%), 57 (37%), 54(34.8%), 52(34.4%), and 51(32.7%) rated as medium respectively the extent to which assessment tools and procedures used lead to dependable scores, yield consistent results upon testing and retesting, are made free from errors of measurement, lead to measurable outcomes, and appropriately measure what they are supposed to measure. A descriptive analysis result has given 3.32 indicating that the extent to which assessment tools and procedures at the universities fulfill some desirable quality requirements fall close to medium.

Furthermore, whereas the mean frequency for using objective assessment tools, and using assessment procedures as a means of improving teaching and learning stand respectively 3.53 and 3.45 both of which fall between sometimes and often; the means for writing test specifications for test tasks, and providing meaningful feedbacks that guide learning are respectively 2.44, and 2.96 both of which are in between rarely and sometimes. Designing assessment tasks that foster valued study habits also has a mean of 3.01, which means sometimes.

On average, the mean for the five questions was 3.07 with standard deviation of 1.041 indicating the frequency of practicing some desirable requirements for learning assessment quality fall close to sometimes.

Finally, the extent to which learning assessment is treated by all staff as an integral component of the entire teaching and learning process, and assessment practices match with available theories/requirements on learning assessment quality were rated as to a little extent by the majority of the respondents.

An average descriptive analysis result on these two issues was 2.73 with standard deviation of 1.123 indicating that treating learning assessment by all staff as an integral component of the entire teaching and learning process, and matching assessment practices with available theories/requirements on learning assessment quality were just to some extent.

8. Conclusions and Implications

The fact that the majority of the respondents had above 10 years of teaching/research experiences, reasonable qualification, and rank shows that they are appropriate to sensibly judge the level of student learning assessment vis-à-vis available theories and requirements. In addition, the very high reliability of the results (.95) has resulted in identifying limitations of assessment tasks, contents, assessment tools, and procedures at the universities in terms of: 1) standing free from all kinds of biases that may disadvantage particular learner groups, 2) mirroring the required skills/knowledge in a real life/in the workplace, 3) empowering the instructors to assess what they teach and what they value; 4) encouraging deep approaches to learning, 5) guaranteeing learners' awarded grades to meaningfully represent the levels, 6) appropriately measuring what they are supposed to measure, and 7) leading to mastery of learning outcomes.

Moreover, it can also be concluded from a synthesis of the results that there were shortfalls in matching assessment practices with available theories/requirements on learning assessment quality in terms of making the learners' awarded grades to meaningfully represent the levels; aligning with the syllabi; 'fitting for purpose' (i.e. learning-oriented); measuring what they are supposed to measure; writing test specifications for test tasks; providing meaningful feedbacks that guide learning; designing assessment tasks that foster valued study habits; and treating learning assessment as an integral component of the entire teaching and learning process.

By all means, the prevailing practices regarding student learning assessment quality at the universities fail to achieve available theories and/or requirements. The case implies that very little conscious efforts have been made to enhance the practices of learning assessment quality commensurate with available requirements/theories and requirements. The results have a far-reaching implication and have been born out of multiple factors that have resulted in misunderstanding and confusion about assessment requirements, and suitable practices (Horst, & Prendergast; Hutchings, 2010; Nicholas & Slotnick, 2018; James, Mcinnis, and Devlin, 2010:6).

The findings go with Sanga's (2016) view that practices dichotomize assessment and teaching-learning processes instead of viewing assessment as an integral part of the teaching-learning process. Consequently, teachers perceive assessment as a one-shot activity in terms of tests and examinations resulting in disjoining the inherent interdependence of teaching, learning and assessment practices. In such a culture, teachers are "caught in a dilemma of either facilitating students' meaningful learning or preparing students who can earn high grades" without achieving the required competencies (Sanga, 2016:1). Students may, therefore, be tempted "to earn good grades, rather than to acquire required knowledge" (Firdissa, 2022:10).

In principle, possession of thorough understanding of the linkage between requirements/theories and practices of assessment quality by teachers and students could serve as a surface symptom for proper implementation of the requirements/theories. Inasmuch as knowledge is linked to the context, and abstractions to concrete experiences; requirements/theories ought to be linked to practices (Chung, 1997). Though the finding appears general, it in one way or another has contributed to failures of fulfilling the required assessment tasks, practices and procedures towards quality learning assessment at the universities.

9. Recommendations

The results have shown that learning assessment practices and procedures at universities failed to fulfill desirable assessment quality requirements. It has, therefore, been recommended that:

- 1) the universities should make utmost efforts to fitting learning assessment practices, procedures, and contents for the purpose wherein learners achieve the required learning outcomes rather than just laboring for grades;
- 2) instructors should take the drivers' seat to ascertain mastery of learning outcomes, dependable scores, measurable outcomes, and direct learners to higher-order objectives (beyond shallow learning) by way of connecting testing with learning, and providing meaningful feedbacks that guide learning; and
- 3) all parties should make conscious efforts to ascertain alignment of assessment practices, procedures and contents with available theories/requirements on learning assessment quality.

References:

- Ainslee, J. (2018). The Significance of Quality Assessment in the Education System (available at <https://elearningindustry.com/quality-assessment-in-the-education-system-significance>)
- AAU [Addis Ababa University]. (2019). Addis Ababa University Senate Legislation
- AAU. (2014) a. Higher Diploma Program. Module Two: Modularization and Modular Curriculum
- AAU. (2014) b. Higher Diploma Program: Module Three: Managing Learning and Assessment
- Bloxham, S.; and Boyd, P. (2007). *Developing Effective Assessment in Higher Education: a practical guide*. England: Open University Press.
- Brown, S.; and Knight, P. (1994). *Assessing Learners in Higher Education*. London & Newyork: RoutledgeFalmer.
- Bryan, K. and Clegg, K. (Eds.), (2006). *Assessment in Higher education*. New York: Rutledge
- Chung, M. A. (1997). Individual and small group interactions in learning to teach with a hypermedia case. Unpublished doctoral thesis, Purdue University.
- Cook, and Campbell (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally
- Edwards, F. (2013). Quality assessment by science teachers: Five focus areas. In *Science Education International*, 24(2), 212-226
- Firdissa J. (2021). Learning Assessment Practices in Ethiopian Universities: Quests versus Upheavals. In *Research in Education Assessment and Learning*, 6(2), 1-14, ISSN: 1947-5497, doi.org/10.37906/real.2021.6.1
- Gerritsen-van Leeuwenkamp, K.J.; Brinke, D.J.; and Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. In *Studies in Educational Evaluation*, Vol. 55, 94-116, <http://dx.doi.org/10.1016/j.stueduc.2017.08.001>
- Horst, S. J. and Prendergast, C.O. (nd.) The Assessment Skills Framework: A Taxonomy of Assessment Knowledge, Skills and Attitudes. In *research and practice in assessment*, 15(1), 1-25
- Ibarra-Sáiz, M.S.; Rodríguez-Gómez, G. & Boud, D. (2021). The quality of assessment tasks as a determinant of learning. In *Assessment & Evaluation in Higher Education*, 46(6), 943-955, DOI: 10.1080/02602938.2020.1828268
- James, R.; Mcinnis, C.; and Devlin, M. (2002). *Assessing Learning in Australian Universities: Ideas, strategies and resources for quality in student assessment* (available at

<https://phongdbcl.ntu.edu.vn/uploads/47/files/old/Tu%20lieu%20tham%20khao/Phuong%20pha%20danh%20gia/assessing%20learning.pdf>

- Luoma, S. (2001). *What does your test measure? Construct definition in language test development and validation*. Jyväskylä, Finland: Centre for Applied Language, ISBN 951-39-0897-6
- Messick, S.J. (Ed.), (1999). *Assessment in higher education: Issues of access, quality, student development, and public policy*. Mahwah, NJ: Lawrence Erlbaum
- MoE [Ministry of Education]. (2006). *Higher Diploma for Teacher Educators. A Training Portfolio* (up).
- Muñoz, A.P. and Álvarez, M.E. (2009). Wash back of an oral assessment system in the EFL classroom. *Language Testing*, 27(1) 1–17, DOI: 10.1177/0265532209347148
- OET [Office of Educational Technology] (2017). *Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update* (available at <http://tech.ed.gov>)
- Özturgut, O. (2011). Standardized Testing In the case of China and the Lessons to be learned for the U.S. *Journal of International Education Research*, 7(2), 1-6
- Paniagua, M. A.; Swygert, K. A. (, Eds.). (2016). *Constructing Written Test Questions for the Basic and Clinical Sciences*. USA: National Board of Medical Examiners
- Redecker, Ch. & Johannessen, Ø. (2013). Changing Assessment – Towards a New Assessment Paradigm Using ICT. *European Journal of Education*, 48(1), 77-96
- Sanga, Ph. L. (2016). Implications of Teacher Educators’ Practices in Assessment for Student Learning in Tanzania. In *Makerere Journal of Higher Education*, 8 (1), 1–24; DOI: <http://dx.doi.org/10.4314/majohe.v8i1.1>
- UNESCO [United Nations Educational, Scientific and Cultural Organization], eds. Kellaghan, T. and Greaney, V. (2001). *Using assessment to improve the quality of education*. Paris: Imprimerie alençonnaise