Journal of Teaching Language Skills
(JTLS)
40(2), Spring 2021, pp. 1-45
DOI: 10.22099/jtls.2021.40269.2978

Online ISSN: 2717-1604
Print ISSN: 2008-8191

Research Paper

Examining 'Assessment Literacy in Practice' in an Iranian Context: Does it Differ for Instructors and Learners?

Zahra Homayounzadeh [*] **Seyyed Ayatollah Razmjoo** [**]

Shiraz University

Abstract

The study reported here is survey research that aims to examine the assessment literacy (AL) of the instructors (N=12) as well as graduate and postgraduate students (N=46) in the Department of Foreign Languages and Linguistics at Shiraz University. To this end, interview questions were designed using Pastore and Andrade's (2019) three-dimensional AL framework. The questions, having been field-tested and their validity having been approved by three experts, were used to interview the participants. The interviews were transcribed and idea units in them were specified and coded according to a coding scheme designed based on Pastore and Andrade (2019). At the conceptual dimension, the results suggested that most of the participants in both groups believed in the learning potential of assessment, and most of the instructors used multiple tasks throughout the term to gather on the students' learning, especially at the MA and Ph.D. levels. At the praxiological dimension that concerned the integration of assessment with instruction, the students mentioned using the results of summative assessment to alter their studying approach, while only a few instructors used assessment results to modify their teaching practices. At the socio-emotional dimension, discrepancies existed among the instructors and between instructors and learners in how ethical issues should be observed in assessment. The results are discussed, and

Received: 11/04/2021 Accepted: 22/05/2021

[*] M.A, Email: yasamin.homayounzadeh@gmail.com, Corresponding Author
[**] Professor, Email: arazmjoo@rose.shirazu.ac.ir

implications are provided for designing teacher education programs and for future research.

***Keywords**:* Assessment Literacy, Conceptual Dimension, Praxiological Dimension, Socio-Emotional Dimension

The salience of assessment for student learning has set up the expectation that teachers use various assessment tasks to support students' learning and measure their progress, hence the increased emphasis placed on teacher assessment literacy (DeLuca & Klinger, 2010; Lan & Fan, 2019). Assessment literacy (AL) is commonly defined as the stakeholders' understanding and appropriate use of assessment practices (Popham, 2004), along with knowledge of the theoretical and philosophical underpinnings in measuring students' learning (Volent & Fazio, 2007). Highlighting the significance of assessment literacy, Popham (2009) argued that lacking such an understanding can "cripple the quality of education" (p. 4), hence the addition of assessment literacy to the list of literacies that are necessary to be acquired in contemporary life (Taylor, 2013). Since assessment literacy is a relatively new field, researchers have highlighted the need for research in this area in response to the critical role of assessment in education and the growth of testing and assessment worldwide (Taylor, 2013).

Several studies, both theoretical and empirical, have been conducted so far to address the various assessment literacy themes, including the specification of the assessment literacy construct (e.g., Davis, 2008; Fulcher, 2012), models for AL (e.g., Kremmel& Harding, 2020), evaluation of the assessment literacy of various stakeholders (e.g., Baker, 2016; Daygers& Malone, 2019), and training programs to promote AL (e.g., Lam, 2015; Lan & Fan, 2019). However, in the majority of these studies, the focus was on

teachers and assessment professionals, giving little heed to learners as a group of stakeholders (Lee & Butler, 2020). Such absence of learners' perspectives in studies on AL is surprising given the emphasis on learner-centered approaches in education, which encourage the incorporation of learners' voices in the curriculum, instruction, and assessment to help them become autonomous (Benson, 2006; Butler et al., 2021). The current investigation, juxtaposing the AL of university students with their instructors, is an attempt to incorporate learners' perspectives in the discussion of AL.

Literature Review

Assessment literacy and language assessment literacy as its "potentially subordinate or overlapping category" (Tylor, 2013, p. 405) have been the focus of various researchers over the past few decades. Studies have explored different issues, from the specification of the components of AL (e.g., Popham, 2009), to the development of AL measures (e.g., Brookhart, 2011; Deluca et al., 2016), evaluation of the status quo of teachers' AL (e.g., Vogt & Tsagari, 2014) to frame materials development and teaching (e.g., Tylor, 2013), to name just a few. This section provides a review of the studies that are the most relevant to the current investigation.

AL Models

Due to the importance of assessment literacy in education, scholars (e.g., Brown & Bailey, 2008; Malone, 2013; Popham, 2009; Stiggins, 1999;) have tried to specify what it is that teachers should know about assessment. The American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association (as cited in Brown & Bailey, 2008, p. 350) provided seven standards for teacher development in

assessment, namely a) selecting and b) developing appropriate assessment methods for instructional decisions, c) administering, scoring, and interpreting the results of assessments, whether externally produced or teacher produced, d) using assessment results to make decisions about individual students and plan teaching, e) developing valid assessment procedures, f) communicating assessment results, and g) recognizing unethical and inappropriate assessment methods.

Nevertheless, these seven components per se are not enough to define teachers' assessment literacy in practice. As Xu and Brown (2016) indicate, there exists a host of mediating factors that affect teachers' assessment practices and should therefore be considered aspects of teacher assessment literacy. In this regard, Xu and Brown (2016) developed a hierarchical framework for teachers' assessment literacy in practice. The model has a knowledge base at its most basic level; however, besides the stated components, or what is commonly called 'pedagogical content knowledge' (Magnusson et al., 1999), it encompasses 'disciplinary knowledge', or "teachers' knowledge of aspects of student learning that are important to assess within a particular unit of study" (Magnusson et al., 1999, p. 108). In other words, knowledge as an aspect of assessment literacy includes both what to assess and how to assess.

Teachers' practice of assessment, as Xu and Brown's (2016) model suggests, is further affected by their conception of assessment and by the sociocultural context in which they work. With respect to the former, the researchers contended that teachers adopt knowledge which is congruent with the conceptions they already hold about assessment and provided that as these conceptions are deeply held, they are resistant to change. Concerning context, Xu and Brown (2016), in agreement with several other researchers (e.g., Gu,

2014; Scarino, 2013), argue that teachers' practice of assessment might also be constrained by such factors as prespecified local/national boundaries. This entails that teacher assessment literacy in practice could be best conceptualized as the compromise teachers make between external constraints and their own values/beliefs.

This conceptualization of AL, as it can be noted, is focused on teachers and professionals as the main stakeholders (Lee & Butler, 2020; Malone, 2017); therefore, bearing in mind that in learner-centered approaches, learners are the main stakeholders, this AL model can be argued not to adequately and explicitly address them and their needs (Butler et al., 2021). Accordingly, Pastore and Andrade (2019) proposed an updated and expanded model for teacher assessment literacy. Assessment literacy, in this respect, is defined as the "knowledge and dispositions that a teacher can use to design and implement a coherent and appropriate approach to assessment within the classroom context and the school system" (Pastore & Andrade, 2019, p. 135). Accordingly, it is believed to have three main dimensions: a) a conceptual dimension that regards the conceptions a teacher has of assessment, teaching, and learning; b) praxiological dimension that encompasses the actions a teacher might carry out when navigating various assessment demands; and c) a socio-emotional dimension that concerns teachers' ability to manage the social aspects of assessment (e.g., test fairness and equity), and attend to the emotional dimensions, especially from the students' perspective (e.g., test anxiety).

The added socio-emotional dimension was intended to directly address learners' needs within the theoretical definition of AL, although, as Butler et al. (2021) indicate, it is further significant to incorporate learners' perception in the discussion of all the other dimensions of AL. This is in part owing to

the fact that assessments are an indispensable aspect of both instruction and learning (Black and William, 1998); therefore, it is significant to give learners equal power and influence in assessment processes as other stakeholders (Butler et al., 2021; Malone, 2017).

Applied Research on AL

Using the specified components for AL, numerous studies have examined teachers' (language) AL in various educational contexts, most specifically to design appropriate training programs. Volent and Fazio (2007), for instance, conducted a survey study in which a group of North American junior teacher candidates self-described their level of AL, the various assessment methods they used, as well as their further training needs. Similarly, Fulcher (2012) conducted a survey research to investigate the needs of a group of language teachers in language testing and assessment. More recently, Lan and Fan (2019), focusing on classroom-based assessment, investigated the gap between teachers' perceived level of AL and the level they expected to reach after further professional training to meet the practical challenges they face in their teaching.

Several studies have similarly been conducted to evaluate teachers' AL in the Iranian context. Ahmadi and Mirshojaee (2016), for instance, examining the AL of language teachers in high schools and private institutes in Iran, found that the participants were not satisfactorily literate in issues related to assessment reliability, bias, validity evidence, and formative assessment. Zolfaghari and Ahmadi (2016), exploring how Iranian high-school teachers in various fields of study, including humanities, English, and science, viewed AL, highlighted the significance of disciplinary knowledge in AL and suggested that this factor be considered in teacher training programs. Afsahi

and Heidari (2017) further examined the AL of EFL teachers and found positive correlations between teachers' years of experience, educational level, and their AL.

As one can notice, in the majority of these studies, teachers' (language) AL was evaluated based on the teachers' own perspectives, giving little heed to the learners. However, in the study by Vlanti (2012), discrepancies were found between teachers and middle school students in their perceptions of assessment practices; for instance, whereas teachers assumed they provided clear information about assessments, the students unanimously disagreed. Similarly, Sato and Ikeda (2015), in their investigation of the Japanese and Korean students' conception of testing constructs required for their university entrance exams, found that the participants did not have a proper understanding of the testing constructs irrespective of the high-stakes nature of the test. Such discrepancies between what the students and teachers or test developers understood about assessments highlight the significance of taking learners' ideas into account in all discussions of AL.

Accordingly, Butler et al. (2021) conducted a study on fourth and sixth-grade students to investigate their conceptions of language AL at the conceptual, praxiological, and socioemotional dimensions. The researchers found that the students already had a great deal of language assessment literacy in all three dimensions; they believed the purpose of assessment should be diagnostic and wanted more cognitively challenging and enjoyable assessment tasks. Accordingly, the researchers recommended that further research be conducted with learners of various age groups and educational backgrounds to expand the scope of learner-centered AL.

To contribute to this line of research, the current investigation aimed to examine the AL of graduate and postgraduate students in the Department of

Foreign Languages and Linguistics at Shiraz University, Iran, and compare their conceptions of AL with those of their course instructors to locate the potential discrepancies. The aim is to verify the extent to which their conceptions of AL are in line with the learner-centered educational approaches that are being advocated worldwide. The findings will be significant not only in enhancing the role of learners in the assessment process but also in providing insights for teacher education programs. The questions which the study sought to answer were as follows:

1. What are the instructors'/learners' conceptions of assessment?
2. How do instructors integrate assessment with teaching and learning?
3. How do students integrate assessment with learning?
4. What socio-emotional dimensions do the instructors consider in their assessment?
5. What socio-emotional dimensions should the instructors consider from learners' point of view?

Method

Participants

The study involved the instructors and students in the Department of Foreign Languages and Linguistics at Shiraz University, Iran, who were randomly selected from the three English majors offered in the department, namely Teaching English as a Foreign Language (TEFL), English Literature, and Linguistics. Of the twelve instructors participating in the study, seven were from TEFL, three from English Literature, and two from linguistics. They all had more than five years of university teaching experience. The pseudonyms PTEFL 1-7, PLit1-3, and PLing1-2 are used to correspondingly

refer to the instructors of the three majors. The information about the level of experience of each, as well as the courses they teach, appears in Appendix A.

Given the purpose of the study to consider the assessment literacy of both teachers and students, 46 students from the three English majors at the department also participated in the study voluntarily. Of the 22 TEFL students participating in the study,14 were Ph.D., and 8 were MA students, who had correspondingly passed three and two two-credit language testing courses at the BA, MA, and Ph.D. levels. The other 24 students participating in the study were from MA Linguistics students (N = 12) and MA English Literature students (N = 12); they were all second-year students and had passed most of their major courses. Appendix B provided a summary of learners who participated in this study.

Theoretical Framework

The study was theoretically based on Pastore and Andrade's (2019) three-dimensional model for teacher assessment literacy, which, as discussed previously, considers a conceptual, praxiological, and socio-emotional dimension for teacher assessment literacy (See Figure 3.1 below). The model was used as a guide both for framing interview questions and for coding the data.

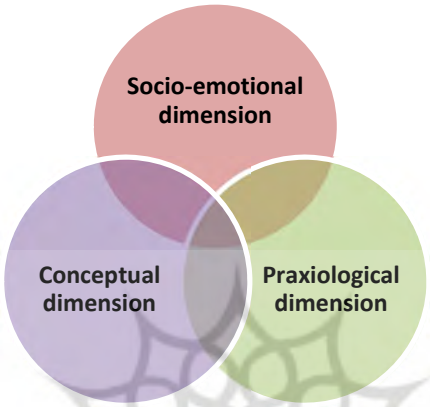

Figure 1: Three-dimensional model for teacher assessment literacy (Pastore & Andrade, 2019)

Instrument

The study used semi-structured interviews as the main instrument to collect data. The first researcher framed the questions in line with the research questions and the guidelines provided by Ary et al. (2010). In designing the questions, attempts were made to avoid any technical terms because, as stated previously, some of the participants did not have a background in language testing and it was assumed that although they might not be familiar with the technical terms, they might know the concepts and apply them in practice. In other words, the purpose of the study was to examine the participants' awareness and application, rather than merely their knowledge, of assessment principles; accordingly, to avoid any bias in favor of those with a background in language testing, no jargons (e.g., reliability, formative assessment, consequential validity) were used in the questions. Having been framed, the questions were then field-tested to ascertain their validity. To this end, three experts in language testing who were familiar with the study were asked to

review the questions and give their opinions about the appropriateness of the questions for obtaining the desired data and about whether they found any problems in them or assumed there was something missing in them; the questions were then revised accordingly. The three experts having reviewed the questions again and approved of their validity, were asked to interview a few participants to identify any probable problems in the questions that might have been overlooked. The necessary modifications were made, and the questions were thereby finalized.

Table 1 presents the interview questions for the instructors and the students. Of the twelve questions posed, the first four concerned the conceptual dimension of assessment literacy, addressing teachers' purpose in assessing the students (formative/summative), what they assess (i.e., content knowledge or student learning over time), the particular assessment instruments or strategies they employ (e.g., observation, performance tasks, tests), and how they ascertain the quality of assessment (e.g., reliability, validity). Questions 5-11, on the other hand, concerned the praxiological aspect of assessment literacy or the practice of assessment. The questions in this regard verified such issues as the participants' awareness of the importance of aligning assessment tasks with learning goals, the different assessment strategies and tools they use or prefer to be used (in case of students) to elicit information about the students' learning, how they interpret assessment results or prefer the results to be interpreted, the type of feedback they provide, and whether or not they engage the learners in the assessment process. Finally, the last two questions pertained to the socio-emotional dimension and inquired about the participants' concern about the ethical aspects of assessment such as test fairness and how they protected the students'

privacy, or in the case of the students, how they preferred their privacy to be preserved.

Table 1.

Interview Questions

Assessment Literacy Dimensions	**Questions of the interview**
Conceptual dimension	1. In general, what is your purpose in assessing the students? 2. What instruments do you use to assess the students (e.g., formal tests, observations, and presentations)? 3. What type(s) of assessment (e.g., end of the term, continuous) is/are preferred in your field of study? 4. What different methods do you use to create rubrics for grading tests? Are you familiar with any methods which can be used? 5. How do you assess students' performance (e.g., papers, tests, presentations)? 6. In your opinion, is it important for teachers to pass teacher training courses? What are other ways through which teachers can gain knowledge about assessment?
Praxiological dimension	7. How do you find the important elements to consider in your assessments? 8. Are you familiar with different types of feedback? (If yes, which one?) What type of feedback do you find the most useful in facilitating students' learning? 9. In your opinion, does it have any benefit to involve students in the process of assessment? What different strategies can you use for this aim? 10. How has your course changed or developed over the years? What changes have occurred? 11. What difficulties do you have for assessing students and the curriculum design of the course? How do you overcome these difficulties?
Socio-emotional dimension	12. How do you interpret evidence generated from the assessment? (Do you consider factors such as the students' performance during the term, cheating during the exam, etc.?) 13. How do you report the results to other stakeholders such as students?

Data Collection

Data was collected for the study through individual interviews with the instructors and group interviews with the students. The participants were reassured that their anonymity would be safeguarded and that their responses would be kept confidential and would be used only for the purpose of this research study. Throughout the interview sessions, the researcher (interviewer) tried to remain neutral, not guiding the respondents toward any responses so as to avoid interviewer bias.

Data Analysis

To analyze the data, the first researcher transcribed all the recorded interviews and identified the idea units in them, with idea unit being defined as "a clause with its pre-and post-V clause elements. Also counted as idea units are non-finite subordinate clauses and finite relative clauses where the relative pronoun is present" (Foster, Tonkyn, & Wigglesworth, 2000, p. 360). Open coding was performed to identify the relevant themes pertaining to each question; subsequently, a coding scheme was designed based on Pastore and Andrade's (2019) three-dimensional AL model to code the participants' responses to the interview questions. Table 2 below presents the scheme.

Table 2.
Coding Scheme Based on Pastore and Andrade's three-dimensional AL Model

Conceptual	Code	Praxiological	Code	Socio-emotional	Code
Why: • Formative • Summative	CP 1 2	Defines learning targets • Yes • Somehow • No	PDLT 1 2 3	Considers test ethics (e.g., cheating • Yes • Sometimes • Never	STE 1 2 3

Conceptual	**Code**	**Praxiological**	**Code**	**Socio-emotional**	**Code**
What: • knowledge/skills/dispositions • development over time • Both	CC 1 2 3	Alignsassessment with learning targets • Yes • Somehow • No	PAL 1 2 3	Protects the privacy of students' assessment data • Yes • Somehow • No	SPA 1 2 3
How: Instrument • formal tests • presentations, • classroom discussions • tasks Rigor • rubric • personal judgment (experience)	CMI 1 2 3 4 CMR 1 2	Selects and differentiates strategies to gather on student learning • Yes • Somehow • No	PSDS 1 2 3		
		Adjusts instruction with data on learning • Yes • Somehow • No	PAIDL 1 2 3		
		Communicates feedback to students • Yes • Sometimes Never	PCFS 1 2 3		
		Supports students in using assessment information to regulate their learning	PSSAL		

Conceptual	Code	Praxiological	Code	Socio-emotional	Code
		• Yes	1		
		• Sometimes	2		
		Never	3		

The coding was done by the first researcher and a Ph.D. holder in TEFL, an expert in the field of language testing. Primarily, four interviews were randomly selected and coded by the two coders; having considered the reliability and discussed the discrepancies, they continued coding the rest of the data, divided into three sets of 10 transcripts. The percentage of agreement between the two codings for each set, as presented in Table 3, suggested high intercoder reliability.

Table 3.
Percentage of Agreement Between Sets of Codings

Set	Number of transcripts	Percentage of agreement
1	4	92%
2	10	94%
3	10	95%

Results

Given the purpose of the study to juxtapose the instructors' and learners' AL For each group of the participants, the quantitative and qualitative results are presented respecting each dimension of AL.

Instructors

Conceptual dimension

The participants' conception of assessment was explored in terms of their assessment goals, the assessment instruments/strategies they used, their

approach in selecting content for assessment, and their strategies to add rigor to their assessments. Throughout the following, the participants' perspectives on each element are presented.

Purpose of Assessment. The majority of the participants in the three majors conceded that their purpose in assessment was both to evaluate the students' learning and to engage them in the learning process. PTEFL 2, for instance, posited that "exams are an extrinsic source of motivation for the students to study, so we can kill several birds with one stone; we can get the students to study the course materials, they will know their strengths and weaknesses, and we can evaluate their academic achievement".

Assessment Strategies/Instruments. The participants were further asked about the different instruments they use to assess the students. At the BA level, the majority used summative assessments (e.g., midterm and final exams) to measure the students' learning and to score them; at MA and Ph.D. levels, however, the majority claimed to use both summative and formative assessment tasks both to evaluate their learning and engage them in the learning process. For summative assessment purposes, the participants were homogeneous in using a midterm and a final exam as a means to assess the students' learning. On the other hand, the majority claimed to use classroom discussions and presentations to assess the students formatively, although a few also referred to other tasks, such as conceptual maps (PTEFL3) or quizzes at the end of any instructional unit (PTEFL4) to engage the students in the learning process. The research paper was another formative assessment task which most of the instructors, especially in the field of TEFL, suggested to use.

Notwithstanding, there were also a few instructors, especially in the fields of Linguistics and English Literature, who preferred summative assessment

tasks over formative ones, irrespective of the students' academic level. In this regard, PLit 2 stated that "at BA level, there are so many students accepted, and it is really difficult to give feedback to all the students". Moreover, she continued, "at the MA level, the students are not knowledgeable enough to be encouraged to get involved in classroom discussions" since the entrance exam which the students take to get admission to the MA level does not adequately test their knowledge in English Literature. Similarly, PLing 1 stated that he uses a summative type of assessment in his courses at MA and Ph.D. levels. His main reasons for this were time limitation and the heavy load of the materials to be covered both at MA and Ph.D. levels, as well as the students' inadequate knowledge to be involved in the assessment process.

Content of Assessment. When asked about what is important to them in assessing the students, the majority of the participants replied that what matters to them is primarily the students' learning of the materials taught in the given course and that they also include class activity to consider the students' learning over time as well. Similarly, in designing exam questions, the majority replied that they develop the questions according to the course objectives and contents taught in the course. PLit 1, for instance, stated that "the questions come basically from the materials included in the syllabus and the students' free discussions in the classroom." PTEFL 5 further added that he tries to include more general issues that can trigger the students' critical thinking and avoids minor details.

Likewise, regarding points they consider in scoring the students' performance on the tests, the participants in both groups provided rather similar responses. The majority replied that they have certain points in mind, and it is important for them that the students include them in their responses. PLing 2, for example, stated that there are certain key words he has in mind

and if students include them in their responses, they will definitely get a better score. PTEFL2 further indicated that for each question, she compares the students' responses with each other and scores them relatively according to how well they fulfill her expectations of the correct answer. PLit 2, however, stated that: "I try to keep my mind open. The answers come from the students' creative minds; if they were successful in incorporating the materials taught in a creative way, that is enough for me".

Assessment Rigor. The participants' responses to the question of how they guarantee the rigor of their assessments were rather identical in nature. The majority pointed out that they evaluate the students' performances, whether on formative or summative assessment tasks, mostly intuitively based on the criteria, which based on their experience they find significant for the students to know. A few (PTEFL 2 PTEFL 3, PTEFL 4) also declared that they use rubrics to score the students' responses, and PTEFL 3 indicated that he sometimes shares the rubric with the students.

Praxiological Dimension

The praxiological dimension of AL concerns the way teachers integrate assessment processes with their teaching practices to monitor and manage the teaching-learning process (Pastore & Andrade, 2019).

Aligning Assessment Criteria with Learning Objectives. For teachers to be able to integrate assessment with their teaching, it is significant first that they define certain learning objectives for their courses and align their assessment practices with those goals. In this regard, all the participants posited that they do define a learning objective for their courses and design their syllabi accordingly. All the 12 instructors interviewed maintained that they design their assessment tasks on the same basis. For instance, "if the

purpose of the course is for the students to learn how to write and publish research articles", PTEFL 1 stated, "I assign them a relevant task and allocate more scores to the quality of the articles they write than to the scores they get on the final exam".

Selecting/Differentiating Strategies to Gather on Students' Learning. As it was stated previously, except for a few of the participants who relied exclusively on a midterm and a final score to assess the learners, the majority used different assessment strategies, such as classroom presentations and discussions, both to gather information on the students' learning and to score them. Depending on the nature and objective of the course, a few also asked the students to write research papers.

Adjusting Instruction with Data on Learning. According to Pastore and Andrade (2019), for assessment to contribute to learning, it is significant for teachers to adjust their instruction with the data they gather on the students' learning. Nevertheless, only a few of the participants stated that they adjust their instruction with the assessment information, among whom one could refer to PTEFEL 3, 4, 5, and 7. PTEFEL 4, for instance, indicated that after relying on midterm and final exams at the MA level for one of his courses, he decided to use several quizzes after teaching each section for the students to be encouraged to study more and better learn the material.

However, the other participants indicated that they do update their teaching materials and make some modifications in their questions as well over time, but the changes they make in the assessment tasks are not necessarily based on the information they obtain from the students' learning.

Communicating Feedback to Students. Another factor that is quite consequential for assessment to contribute to learning is feedback (e.g., Bennet, 2010; Carless et al., 2006). Accordingly, the participants were asked

about whether they give feedback to students and if yes, what type of feedback they use. The majority of the participants indicated that they use direct feedback in assessing the students' exam papers or when commenting on their research papers or class presentations. PTEFL 1, for example, posited that "when students are writing proposals, they do not have enough time to figure out what the indirect feedback is; as a result, for giving written feedback, mainly a direct one is used".

Likewise, PLit 1, 2, and 3 stated that in their classes, they mainly use an indirect feedback because they believe that interrupting students for correcting them demotivates students and may make them reluctant to speak in the class. Moreover, they argued that by using indirect feedback and discussing different answers, students can learn the materials more deeply in comparison with the time when they use direct feedback. However, when it comes to giving feedback in the exams or on papers, all the English Literature participants mentioned that they use a direct type of feedback as it is very time-consuming for the students to figure out what this specific indirect feedback means.

Supporting Students in Using Assessment Information to Regulate Their Learning. According to Careless and Boud (2018), for feedback to contribute to learning, it is significant that the students be assessment literate so that they can act upon the feedback. Therefore, for a teacher to be assessment literate, it is not enough for them just to provide feedback but is further crucial that they support the students to understand the feedback and act upon it. In this respect, the participants were asked about whether and, if so, how they support learners to use the feedback. The approach taken by all participants, as they indicated, is to send a message to their students through the university website or social media, set a date and time and invite students to come to their offices to see their papers and learn from their mistakes.

However, as PTEFL 2 indicated, most of the students generally come to negotiate about their scores rather than learn from their mistakes.

A useful approach TEFL 5 has adopted is that in some courses at MA and Ph.D. levels, after taking a midterm and a final exam, he devotes one session of the class to analyzing the papers. Thus, he brings the papers to the class, gives them to students, and analyzes each question one by one. He found this approach very effective as students can learn a lot in these sessions.

Socio-Emotional Dimension

Regarding the socio-emotional dimension of AL, the present study explored how the instructors observed test fairness and protected the students' privacy (Pastore & Andrade, 2019).

Ensuring test fairness. Although all the participants homogeneously claimed to ensure test fairness, their approaches were different in this regard as they held different views about test fairness. For some (e.g., PTEFL 2, PTEFL 4, and PLit 1), test fairness involves scoring the examinees not merely based on their performance on the midterm and final scores because there might be some extraneous factors affecting their test performance, either positively (such as cheating) or negatively (such as stress, sickness, etc.). Hence, they believed when scoring the examinees' performance, these criteria must also be taken into account and they must be scored based not only on their test performance but also on their performance throughout the term. Others, however, regarded fairness in treating all learners objectively and scoring them based on the scores they get on the midterm and final exams.

Protecting the Privacy of Students' Assessment Data. To verify how the instructors protected the privacy of the students' assessment data, as an indication of their AL, they were asked about the way they communicate

assessment results; the aim was to verify whether they are sensitive to the emotional aspects of assessment. Based on the findings, the majority of the instructors interviewed (i.e., PTEFL 7, PTEFL 6, PTEFL 5, PTEFL 4, PTEFL 3, PTEFL 2, PLit 2, PLing 1 and PLing 2) suggested sensitivity to the importance of protecting the students' privacy, indicating that because some students might be sensitive about their scores, they communicate the test results individually through the university automation system. A few of the participants stated that they do not completely protect the participants' privacy as they report the exam results of all the students in one file sent to all the students, although they mention the student numbers in the file instead of the students' names. Their reason for not completely protecting the participants' privacy was that by doing so, they intend to encourage the students to study harder.

Learners

As the second group of participants, the learners were also interviewed to explore their conception of assessment literacy and verify how they preferred assessment to be practiced in that context. The aim was to explore the homogeneity of the two groups of stakeholders in terms of their assessment literacy since as Jeong (2013) states, it is significant that all stakeholders, including the students, share a common view of assessment so that the learners' needs are better fulfilled. The results are presented below.

Conceptual Dimension

Assessment purpose. The first question of the interview focused on the participants' views about the purpose of the assessment. For the majority, the assessment was a vital means to get them to carefully study the course

materials since, without exams, they admitted studying the materials haphazardly. Some of the learners also regarded assessment as a means to evaluate their learning and discriminate them according to how well they have achieved the course objective(s). Of course, a few also opposed exams altogether, finding them "stressful; what is important is the students' learning, not scores" (SLit 8). As a result, they preferred having a formative type of assessment and receiving feedback in the classroom so they "will not forget the correct answer and it will go into their long-term memory" (SLing 5).

Assessment instrument. Formal tests, most specifically multiple choice and essay type tests, classroom discussion and presentations, were the learners' most preferred assessment techniques. Nevertheless, the majority disagreed with the use of tasks in the class indicating, as STEFEL 5 stated, that they are "very time consuming and are not effective in enhancing learning". To use STEFEL 10's words, "the learners already have enough materials to study and do not have time to accomplish the tasks assigned to them". In fact, like what Yorke (2003) describes, the learner participants in the current study regarded tasks as non-essential and as unduly increasing their workload.

Content of Assessment. With respect to assessment content, differences were observed in what the participants in the three majors found significant for studying. The participants in the Linguistics and English Literature mostly indicated that they mainly focus on and study the issues which were taught in the classroom as most of the exam questions are based on the same materials. On the other hand, TEFL learners posited that they study the whole materials as they cannot distinguish which parts are more important. The reason, as STEFL 4 stated, was "the unpredictability of instructors; that is, some questions of the exams come from materials which were not discussed in the

classroom and also sometimes some instructors want the exact words mentioned in the book and if those words are not written in their responses, they will lose points; this compelled them to study the whole materials.

On the other hand, regarding the type of assessment they preferred, the majority chose continuous assessment throughout the term over a midterm or final exam. STEFEL 5, for instance, believed "a single test is never indicative of your ability". Moreover, SLit 3 stated, "it has happened several times for me that I did not have a good performance on the midterm/final exam, even though I knew the answers, so I don't think it's fair to be scored only based on a midterm and/or final exam". Of course, there were also a few students (e.g., Slit 6, and Slit 10) who disagreed with this position, positing that studying before the exam is the best way for them to learn. They also found this approach less stressful than being continuously evaluated during the term; STEFL 6 indicated "exams/evaluations are always stressful, so continuous evaluation throughout the term means being under stress for 3-4 months, which is really unbearable for me".

Assessment rigor. In response to the question about what they thought their instructors based their evaluations on, the participants provided different responses. In fact, they believed that different instructors adopt different approaches. "For some", STEFEL 2 indicated, "it is telepathic scoring; that is, they have some key words in mind, and you will get a good score only if you include those key words in your answer; otherwise, even if your answer is conceptually quite correct, you won't get a good score". On the other hand, SLit 9 thought "they [the instructors] score them mostly impressionistically or based on how complete they find their responses; after so many years of experience, they know what the complete answer is, but at times, it's just a

matter of quantity". STEFEL 5 further added that "a few also have rubrics in which they have specified criteria for good performance".

However, when asked about how they preferred to be scored, the majority replied that using a rubric would be their ideal approach. "This way", SLing 7 stated, "it is easier to negotiate about our scores and we can better learn from our mistakes". Similarly, STEFEL 2 commented, "with a rubric, I think scoring will be more objective, but I don't think it is practical for all exams and all courses".

Praxiological Dimension

The praxiological dimension, as stated previously, pertained to the integration of assessment with learning and teaching practices.

Defining Learning Objectives. One of the interview questions was designed to illustrate learners' understanding of what course/learning objectives are. For the majority, their objective for every course was simply to learn the materials taught throughout the course; after all, LLing 3 indicated, our instructors define the course materials based on some prespecified objectives, so if we learn the materials well, we can reach those objectives". "Of course", STEFL 4 admitted, "there are certainly more for us to learn than what the professors assign us, but we usually have so much to study that we can't find time to go beyond the course materials". A few MA students also replied that they identify the learning objectives by asking the top senior students who have already passed the given course to know what is important to learn, especially for the exams.

Aligning Assessment with Learning Targets. Concerning the second aspect of the praxiological dimension, the majority of the students in the three majors believed the alignment of assessments with learning objectives, which

they found the same as the course materials, happened in most of their courses, but not all. For example, LTEFL 6 stated, "I don't guess multiple-choice and fill-in-the-blank questions are appropriate types of questions for a course whose objective is for the students to learn how to conduct academic research". LTEFL 3, on the other hand, found the questions in some tests not just in line with the materials covered in the class and believed the class activities during the term do not adequately prepare them for the final exam.

Selecting/Differentiating Strategies to Gather on Students' Learning. As a Ph.D./MA student, do you think it is necessary for your instructors to constantly gather information on your learning?, the students were asked. Whereas some believed masters and doctorate students are to be autonomous and gathering on students' learning is mostly pertinent to undergraduate students, the majority disagreed. They argued that although the nature and purpose of masters and doctorate education are different from those at BA level, the instructors can and should constantly consider the students' learning, so they can provide them with support. STEFL 2, for instance, indicated

> "Although learner autonomy is quite significant, especially at the MA and Ph.D. levels, I believe instructors should still care about the students' learning; after all, MA and Ph.D. students are at the beginning of developing their research skills, for instance, and need the support of their supervisors".

What strategies do/can your instructors use to get information on your learning?, the students were asked. The response mostly was that "the instructors should give us tasks relevant to the skills they expect us to develop and should provide us with feedback on our performance, not just to evaluate and score us, but to support us and help us learn", STEFL 12 suggested. Others

proposed that the students be involved in the process of evaluation, for instance, by getting them to write questions for their own evaluation; the quality of the questions they write shows how well they have studied and learned the materials, STEFL 10 advocated. Such strategies, STEFL 8 argued, are effective in giving "the students a voice so they feel valued, which is a positive point for learning the course better".

Adjusting Learning with Assessment Data. The next question about the praxiological dimension of AL the students were asked was whether they adjust their learning approaches based on the information they obtain from the assessment tasks the instructors provide. For some, the response was negative, and as STEFL 5 explained, the reason was that "the comments we receive for the tasks we accomplish, such as the book/journal reviews or articles we write, are general comments about the overall quality of our performance, so we can't know how to write a better book review, for instance". Of course, STEFL 3 added, "some of our instructors do provide us with feedback on each individual part of our papers, but you can't still know how to improve your work. For instance, you know that the gap in the literature you are pointing to isn't significant, but you can't know how to find a significant gap in the literature".

For some of the MA students, however, the response was different. STEFL 12, for example, stated that she changed her studying approach for one of her courses after she realized from the midterm exam what points the instructor focuses on in tests. The same was true of SLing 8, who tended to listen to the recordings of her instructor's teaching after realizing that the key points that are of concern to the instructor in scoring are from his own teaching in the class.

Communicating Feedback to Students. Given the salience of feedback in improving the students' learning (e.g., Carless et al., 2006), the learners were asked about the type of feedback they find most appropriate and conducive to learning. In this regard, most of the participants conceded that they preferred direct feedback, especially on their written performances, so they would not be misguided. Some further emphasized that the feedbacks not always serve to determine their scores, but they are given a chance to improve their performance and their scores afterward.

Socio-Emotional Dimension

Considering Test Ethics. How do you define test ethics? Do you think factors such as cheating, illness, etc., that can positively/negatively affect your performance should be considered in scoring? For the majority, test ethics was that their scores on a given test be representative of their knowledge/ability level on what the test is measuring. Some believed more important than cheating, etc. is that the test itself, as well as the criteria considered in scoring, should provide a valid representation of the construct the test is assessing. "[…] For instance, I don't think remembering exact words can show your ability to conduct research", STEFL 14 claimed. Others, however, insisted that contaminating factors be considered in scoring, "if a test is to discriminate learners according to their mastery of a topic", SLing 4 argued. "It's not fair when you study hard and answer only based on your knowledge, but your classmate gets a higher score than you merely through cheating" (SLit4), or "when you just don't get enough time to write your answers completely", STEFL 12 added.

Protecting the Privacy of Students' Assessment Data. Finally, regarding the privacy of their assessment data, most of the PhD students

indicated that they do not mind their classmates knowing their scores as they indicated that scores were not important to them at this level of education. However, not all MA students agreed with this position, arguing that it is discouraging to them if they have a poor performance on a test and their classmates recognize it.

Discussion

Conceptual Dimension

The first research question concerned the conceptual dimension of the participants' AL. In this regard, the majority of the learners participating in the study agreed with their instructors regarding assessment as a means both to evaluate their learning and engage them in the learning process. Of course, there were also a number of learners and instructors who dismissed the learning potential of assessment and of alternative assessment tasks, regarding assessment as basically, a means to evaluate the students' learning. However, although the other instructors who did believe in the learning potential of assessment used the term 'formative assessment' in reference to the tasks they used throughout the term, some caution might be needed in applying the term 'formative' here. As Bennett (2011) argues, not any progressive assessment that supports learning is necessarily formative. A key element of formative assessment, which in the context of the present study appeared to be absent, is 'formative feedback', or feedback that uses evidence from the students' learning to adapt teaching to meet student needs (Black & William, 1998; Clark, 2012).

Of course, this is not in contradiction with the participants' belief in the learning potential of assessment; as Bennett (2011) indicates, a well-designed summative assessment can also contribute to learning by encouraging learners

to prepare, organize their knowledge, and rehearse the domain-relevant information to strengthen the representation of the information retrieved and develop expertise. What seems to be absent is the formative use of assessment to form both instruction and learning, which is further evident in the praxiological dimension of the participants' AL. This might, on the one hand, be attributable to the constraints which the instructors referred to, that is, the unfamiliarity of the learners with such formative assessment techniques as self-and/or peer assessment (e.g., PTEFL 2, PLing 2), the heavy load of the materials to be covered in the class (PLit 2, PLing 1), the large number of students in the classes, particularly at the BA level, as well as the instructors' priority being the publication of numerous research articles rather than the quality of instruction.

Nevertheless, with the purpose of education being to enhance the students' learning and develop their self-regulation abilities to be lifelong learners (Clark, 2012; Shepard, 2000), it is significant to engage them in the process of assessment in that:

> Assessment plays a key role in both fostering learning and the certification of students. However, unless it satisfies the *educational* purpose of ensuring students can identify high-quality work and can relate this knowledge to their own work, the likelihood that they will reach high quality themselves is much reduced. (Boud et al., 2010, p. 1)

Praxiological Dimension

The next two research questions regarded the praxiological aspect of the participants' AL, which concerns the integration of assessment with instruction and learning. In this regard, the majority provided learners with evaluative and/or descriptive feedback, used mainly to determine the learners'

summative scores rather than supporting their learning. Only a few of the instructors referred to the tendency to adapt their teaching practices according to the feedback they receive from their assessment of the learners or to engage the learners in the process of assessment. However, the learners did seem to use the information they obtained from the tests to form their studying approaches, for instance, by focusing more on the points that were significant in assessment and instruction.

This approach could, on the one hand, enhance the students' learning of the course materials, and provided that those materials are well defined, it can enhance their competence in the given field and thereby have a positive washback effect (Cumming, 2013). On the other hand, despite this effectiveness, the adoption of purely summative, as opposed to formative, assessment, as some instructors appeared to do, does not seem compatible with the goal of higher education in the age of globalization to educate autonomous lifelong learners (Gardner, 2012). Moreover, as Harlen (2012) argues, summative assessment used to help learning yields little validity evidence regarding the depth and breadth of the learning that takes place, hence questions about the transferability of the skills developed.

Thus, to promote learner autonomy and ascertain the transferability of the skills they develop, it is recommended that the instructors adopt a learning-oriented assessment approach (Carless et al., 2006), not merely in terms of the tasks they assign to students but also as Pedder and James (2012) indicates, by "engag[ing] gage in the risky business of problematizing their own practice, seeking evidence to evaluate in order to judge where change is needed, and then to act on their decisions" (Pedder & James, 2012, p. 7). In other words, it is crucial that the instructors engage in learning-oriented assessment with respect to their own professional practices so they will be

able to better help learners to promote intentionality and autonomy in learning (Pedder and James, 2012). This was indeed the aspect that appeared to be missing in the assessment practices of some of the instructors interviewed in the study that should be promoted for the education of autonomous lifelong learners.

Socio-emotional Dimension

On the other hand, at the socio-emotional dimension, the results of the study suggested that the majority of the participating instructors implemented test fairness in terms of scoring the learners according to either their performance only on summative tests or to both their summative scores and their performance throughout the term. Only a few (PTEFEL 2, and PTEFEL 4), however, mentioned the significance of sharing assessment criteria with the learners in determining test fairness. In fact, although almost all the participating instructors mentioned the percentage of midterm, final, and class activity scores in determining the students' final scores, only a few clearly specified their assessment criteria, especially for the tasks that involved subjective scoring. Nevertheless, transparency in assessment is a very consequential factor in determining both test fairness and the educational value of the tests. In fact, as Messick (1994) argues,

> If assessment itself is to be a worthwhile educational experience serving to motivate and direct learning, then the problem and tasks posed should be meaningful to the students. That is, not only should students know what is being assessed, but the criteria and standards that constitute good performance should be clear to them in terms of both how the performance is to be scored and what steps might be taken or what directions moved in to improve performance...a number of things can be

done...to improve somewhat the transparency and meaningfulness of structured and semi-structured tasks. (p. 16)

The findings, therefore, have significant implications both for teacher education programs and for future research in this respect.

Conclusions

To conclude the study, the section discusses its implications for teacher education programs and future research on AL. The results primarily highlight the significant elements that need to be emphasized in teacher education programs. First, given the significance of the teachers' emotional inclinations toward various aspects of assessment, it is crucial for teacher education programs to maneuver on this aspect so that teachers can realize the importance of assessment in improving learning. In fact, it is essential for teachers to realize that assessment is not merely a means for measuring the students' learning, but perhaps more significantly, one of the best means to support learning. Such a view could ultimately be conveyed to the learners so that they will not regard formative assessment as extra and non-essential work (Yorke, 2003), but as a significant step toward becoming life-long learners.

On the other hand, considering the three elements, which Carless et al. (2006) found consequential for assessment to contribute to learning (i.e., assessment tasks, student engagement in self-and/or peer assessment, and feedback as feed-forward), it is further necessary that teacher education programs familiarize teachers with different assessment strategies, which they can use to integrate assessment with their instruction in different courses. Further important is that teachers get familiarized with different types of feedback (i.e., descriptive, evaluative, and supportive (Hatti & Timperly, 2007)) so that they can feed the students' learning forward.

Another significant element to be emphasized in teacher education programs is for the teachers to develop and use assessment rubrics, which they can share with the learners, particularly for the exams that are subjectively scored. The factor is significant not only in enhancing transparency in assessment and thus test fairness, but as formerly discussed, is essential in determining the educational value of the test and guiding their learning (e.g., Bennet, 2011; Messick, 1994).

The present study further provides insights for future research on assessment literacy. Primarily, future research could use the same model used in this study to examine the assessment literacy of the stakeholders in other contexts. Nevertheless, although the present study used the interview as the main data collection technique, future research could triangulate the results, for instance, by observing teachers' classrooms and/or analyzing their assessment instruments, such as their tests or tasks, assessment rubrics/criteria, as well as their feedbacks more closely to better identify their strengths and weaknesses. Besides, in studying the socio-emotional aspect of assessment, the current investigation focused merely on the teachers' consideration of test fairness and their protection of the students' privacy; other researchers could further consider other factors, such as their regard for the consequential validity of their assessments, as well as for the influence of assessment on the students' engagement and on teacher-student relationships.

References

Afsahi, E., Heidari, T., (2017). Iranian EFL teacher's assessment literacy and inclination towards the use of alternative assessment: *Journal of Applied Linguistics and Language Research* (vol. 4, pp. 283-290).

Ahmadi. A., Mirshojaee.B., (2016). Iranian English language teachers' assessment literacy: The case of public schools and language institute teachers. *The Iranian EFL Journal* (vol. 12, pp. 6-32).

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). Introduction to Research in Education. Cengage Learning.

Baker, B. A. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision-makers. *Canadian Journal of Applied Linguistics*, *19*(1), 63-3. https://journals.lib.unb.ca/index.php/CJAL/article/view/23033

Benson, P. (2006). Autonomy in language teaching and learning. *Language Teaching*, *49*(1), 21– 40. https://doi.org/10.1017/s0261444806003958

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139-48.

Boud, D. et al. (2010) Assessment 2020: Seven propositions for assessment reform in Higher Education, Sydney: Australian Learning and Teaching Council. Online. Available: <http://www.assessmentfutures.com> (accessed 26/05/19).

Bratchell, N. (1989). Cluster analysis. Chemometrics and Intelligent Laboratory. System, *6*, 105–125.

Brookhart, S. (2013). Educational assessment knowledge and skills for teachers. Educational Measurement: Issues and Practice, 30, 3-12.

Brown, J. D., & Bailey, K. (2008). Language testing courses:What are they in 2007? Language Testing, 25(3), 349-383.https://doi.org/10.1177/0265532208090157.

Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 0265532221992274.

Carless, D., Joughin, G., &Mok, M. (2006). Learning-oriented assessment: principles and practice. *Assessment and evaluation in Higher Education*, *31*(4), 395-398.

Carless, D., &Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*(2), 205-249.

Cumming, A. (2013). Assessing integrated writing tasks: Promises and Perils. *Language Assessment Quarterly*, *10*(1), 1-8.

Davies, A. (2008). Textbook trends in teaching language testing. Language Testing, 25(3), 327–347 https://doi.org/10.1177/ 0265532208090156. DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419–438.

Deluca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251-272.

Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, *36*(3), 347–368. https://doi. org/10.1177/0265532219826390

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, *21*(3), 354-375.

Fulcher, G. (2012). Assessment literacy for language classroom, Language Assessment Quarterly, 9(2), 113-132.

Gardner, J. (2012). Assessment for Learning: A Compelling Conceptualization. In Gardner, J. (Ed.). *Assessment and Learning*. SAGE Publications.

Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a foreign language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice, 21*(3), 286-305.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Jeong, H. (2013). Defining assessment literacy: is it different for language testers and non-language testers? *Language Testing*, *30*(3), 345–362. https://doi.org/10.1177/0265532213480334.

Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the

language assessment literacy survey. *Language Assessment Quarterly*, *17*(1), 100–120. https://doi.org/10.1080/15434303.2019.1674855

Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, *61*, 112-122.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, *32*(2), 169–197. https://doi.org/10.1177/0265532214554321

Lee, J., & Butler, Y. G. (2020). Reconceptualizing language assessment literacy: Where are language learners? *TESOL Quarterly*, *54*(4), 1098–1111. https://doi.org/10.1002/tesq.576

Malone, M. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, *30*(3), 329-344.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, *23*(2), 13-23.

Pedder, D. & James, M. (2012). Professional Learning as a Condition for Assessment for Learning. In J. Gardner, (Ed.). *Assessment and Learning*c (pp.34-48). London: SAGE Publications.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128-138.

Popham, W. J. (2004). Why Assessment Illiteracy Is Professional Suicide. Educational Leadership, 62(1), 82.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, *48*(1), 4-11.

Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on student learning. *Language Testing in Asia*, *5*(1), 1–15. https://doi.org/10.1186/s40468-015-0019-z

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, *30*(3), 309-327.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and practice*, *18*(1), 23-27.

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing 30*(3), 403-412.

Vlanti, S. (2012). Assessment practices in the English language classroom of Greek junior high school. *Research Papers in Language Teaching and Learning*, *3*(1), 92-122. https://rpltl.eap. gr/images/2012/RPLTL-03-01-fulltext.pdf

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374-402.

Volante, L., & Fazio, X. (2007). Exploring Teacher Candidates' Assessment Literacy: Implications for Teacher Education Reform and Professional Development. *Canadian Journal of Education, 30*(3), 749-770.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149-162.

Yim, O., & Ramdeen, K. T. (2015). Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *The quantitative methods for psychology,11*(1), 8-21.

Yorke, M. (2003) Formative assessment in higher education: Moves towards theory and enhancement of pedagogic practice, *Higher Education, 45*, 477–501.

Appendix A

Information about professors participating in the study

Professors	Academic rank	Years of academic experience	Courses taught
PTEFL 1	Professor	Over 10	• **BA**: Language Testing • **MA**: SPSS, Research Methodology • **PhD**: Research Methodology
PTEFL 2	Associate professor	Over 15	• **BA**: Reading Comprehension, Grammar and Writing, Phonetics and Phonology, Linguistics • **MA**: Advanced Writing, Applied Linguistics • **PhD**:Second Language Acquisition
PTEFL3	Associate professor	Over 15	• BA: reading comprehension, listening and speaking, letter writing, essay writing, language teaching, and language testing • MA: Language Teaching Theories and Principles, Language Skills, Advanced Writing, Language Testing • PhD:Teacher Education, Advanced Language Testing, Critique of Language Teaching Theories
PTEFL 4	Associate professor	Over 10	• **BA**: Reading Comprehension, Reading and Writing, Oral presentation of short stories, Linguistics, Essay writing • **MA**: SPSS, Advanced Writing, Applied Linguistics, Language Testing • **PhD**: Advanced Language Testing

Professors	Academic rank	Years of academic experience	Courses taught
PTEFL 5	Assistant professor	Over 8	• **BA**:Reading Comprehension, Oral presentation of short stories, Translation of Islamic Texts • **MA**:Theories of Language Teaching and Learning, Materials Development • **PhD**:Second Language Acquisition
PTEFL 6	Assistant professor	Over 5	• **BA**: Reading Comprehension, Grammar and Writing, Listening and Speaking, Oral Presentation of Stories, Translation of Islamic Texts, Language Teaching Methodology • **MA**:Principles of Language Teaching, Materials Development
PTEFL 7	Assistant professor	Over 6 years	• BA: Reading Comprehension, Listening and Speaking, Oral Presentation of Stories, Use of Language Expressions in Translation, Linguistics, General English • MA: Psychology of Language, Sociology of Language, • PhD: Sociology of Language
PLit 1	Assistant professor	Over 8 years	• BA: Novel I & II, Literary Play, Samples of Simple English Poetry, Survey of English Literature • MA: Shakespeare's Plays, 18th century Literature, Renaissance, A Survey of English Literature • PhD: Shakespeare's Plays,

Professors	Academic rank	Years of academic experience	Courses taught
			Contemporary Dramatic Literature, Modern Plays
PLit 2	Assistant professor	Over 5 years	• BA: Literary Criticism I & II, Samples of simple English Poetry, An Introduction to English Literature, Samples of Simple English Poetry, Play I, Literary Translation, Short Story, Research Methodology, A Study of Islamic Translation Works, Selected Texts of Literary Prose, General English • MA: Literary Criticism, Research Methodology, Contemporary English Poetry, • PhD: Contemporary Dramatic Literature, Modern Plays • Literature & Cinema, Comparative Literature; Comparative Literature of Iran & West
PLing 1	Associate professor	Over 10	• BA: Grammar &Writing, A Study of Islamic Translated Works, Specialized English, Theoretical Principles of Translation, Understanding Semantics, Foundations of Applied Linguistics • MA: Discourse Analysis & Persian Linguistics in Practice, Persian Linguistics' Principles, Critical Discourse Analysis, Persian Linguistics Schools of Thought • PhD: Discourse Analysis

Professors		Academic rank	Years of academic experience	Courses taught
PLing 2	Assistant Professor	Over 3 years	• BA: General English, Specialized English for Studying Social Science Texts. • MA: Learning Persian Language, Contemporary Persian Language Stylistics, Figurative Linguistics, Morphology, Semantic Principles, Syntactic Analysis, Construction of Persian Languages, Research Methodology in Persian Language • PhD: Natural Language Processing	

Appendix B

Learners' information participated in the study

Learners	Academic Degree	Major
STEFL 1	PhD Candidate	Teaching English as a Foreign Language
STEFL 2	PhD Candidate	Teaching English as a Foreign Language
STEFL 3	PhD Candidate	Teaching English as a Foreign Language
STEFL 4	PhD Candidate	Teaching English as a Foreign Language
STEFL 5	PhD Candidate	Teaching English as a Foreign Language
STEFL 6	PhD Candidate	Teaching English as a Foreign Language
STEFL 7	PhD Candidate	Teaching English as a Foreign Language
STEFL 8	PhD Candidate	Teaching English as a Foreign Language
STEFL 9	PhD Candidate	Teaching English as a Foreign Language
STEFL 10	PhD Candidate	Teaching English as a Foreign Language
STEFL 11	PhD Candidate	Teaching English as a Foreign Language
STEFL 12	PhD Candidate	Teaching English as a Foreign Language
STEFL 13	PhD Candidate	Teaching English as a Foreign Language
STEFL 14	PhD Candidate	Teaching English as a Foreign Language
STEFL 15	MA Student	Teaching English as a Foreign Language
STEFL 16	MA Student	Teaching English as a Foreign Language
STEFL 17	MA Student	Teaching English as a Foreign Language
STEFL 18	MA Student	Teaching English as a Foreign Language
STEFL 19	MA Student	Teaching English as a Foreign Language
STEFL 20	MA Student	Teaching English as a Foreign Language
STEFL 21	MA Student	Teaching English as a Foreign Language
STEFL 22	MA Student	Teaching English as a Foreign Language
SLit 1	MA Student	English Literature

Learners	**Academic Degree**	**Major**
SLit 2	MA Student	English Literature
SLit 3	MA Student	English Literature
SLit 4	MA Student	English Literature
SLit 5	MA Student	English Literature
SLit 6	MA Student	English Literature
SLit 7	MA Student	English Literature
SLit 8	MA Student	English Literature
SLit 9	MA Student	English Literature
SLit 10	MA Student	English Literature
SLit 11	MA Student	English Literature
SLit 12	MA Student	English Literature
SLing 1	MA Student	Linguistics
SLing 8	MA Student	Linguistics
SLing 9	MA Student	Linguistics
SLing 10	MA Student	Linguistics
SLing 11	MA Student	Linguistics
SLing 12	MA Student	Linguistics

Appendix C

Assessment Literacy Dimensions	Questions of the interview
Conceptual dimension	1. In general, what is your purpose in assessing the students? 2. What instruments do you use to assess the students (e.g., formal test, observations, and presentations)? 3. What type(s) of assessment (e.g., end of the term, continuous) is/are preferred in your field of study? 4. What different methods do you use to create rubrics for grading tests? Are you familiar with any methods which can be used? 5. How do you assess students' performance (e.g., papers, tests, presentations)? 6. In your opinion is it important for teachers to pass teacher training courses? What are other ways through which teachers can gain knowledge about assessment?
Praxiological dimension	7. How do you find the important elements to consider in your assessments? 8. Are you familiar with different types of feedback? (If yes which one?) What type of feedback do you find the most useful in facilitating students' learning? 9. In your opinion does it have any benefit to involve students in the process of assessment? What different strategies can you use for this aim? 10. How has your course changed or developed over the years? What changes have occurred? 11. What difficulties do you have for assessing students and curriculum design of the course? How do you overcome these difficulties?
Socio-emotional dimension	12. How do you interpret evidence generated from assessment? (Do you consider factors such as the students' performance during the term, cheating during the exam, etc.?) 13. How do you report the results to other stakeholders such as students?