

RESEARCH

Open Access



A phenomenographic study on language assessment literacy: hearing from Iranian university teachers

Afsheen Rezai^{1,2*}, Gudarz Alibakhshi³, Sajjad Farokhipour⁴ and Mowla Miri³

* Correspondence: Afsheen.rezai@abru.ac.ir

¹Department of Teaching English and Linguistics, Ayatollah Ozma Burojerdi University, Borujerd City, Iran

²Teaching English and Linguistics Department, Faculty of Literature and Humanities, Ayatollah Ozma Burojerdi University, Lorestan Province, Burojerdi City, Iran
Full list of author information is available at the end of the article

Abstract

This study aims to disclose the Iranian university teachers' perceptions of the fundamentals of language assessment literacy (LAL). To this aim, using purposive sampling, eighteen university teachers from two Iranian universities were invited to participate in semi-structured interviews. Their viewpoints were audio-recorded, transcribed, and subjected to a phenomenographic analysis. Findings yielded two overarching LAL domains: *knowledge* (e.g., having an acceptable level of digital LAL, satisfying ethical requirements, benefiting more from performance assessment, considering students' individual differences, making assessment valid, assuring that tests are reliable, and having an acceptable level of pedagogical content knowledge) and *skills* (e.g., involving students in assessment, using alternative assessment methods, employing logically traditional assessment methods, informing students about test results, administering tests in standardized ways, using valid grading procedures, and bringing positive wash-back effects). After discussing the results, the study concludes with proposing a range of implications for different testing stakeholders and highlighting some avenues for further research.

Keywords: Language assessment literacy, Higher education context, University teachers, Phenomenographic analysis

Introduction

Assessment is construed as one of the most prominent facets of instructional contexts since it can have a considerable impact on the quality of teaching and hence learning. One of the branches of assessment which has gained noticeable momentum by the seminal work of Black and William (1998) is "assessment for learning." In the succeeding years, more studies validated that assessment should be considered as an influencing factor to boost deep learning (Coombe et al., 2020; Elwood & Klenowski, 2002), raise second language (L2) learners' motivation (Brookhart & Bronowicz, 2003), improve L2 learners' self-concept (Black & Wiliam, 2010), and increase L2 learners' understanding of quality assessment (Smith et al., 2013). As Malone (2013) argues, "strong, properly implemented assessment provides teachers, students, and all testing stakeholders with important information about student performance and about the

extent to which learning objectives have been attained in the classroom” (p. 330). There is, indeed, a reciprocal relationship between teaching and assessment such that “assessment informs and improves teaching and vice versa” (Malone, 2013, p. 330). However, this interrelated linkage between teaching and assessment cannot be established and strengthened unless L2 teachers are equipped with sufficient assessment knowledge. This undeniable need for assessment knowledge provoked Stiggins (1991, 1995) to propose the very construct of “assessment literacy.”

In L2 education, assessment literacy is concerned with the understanding of testing stakeholders (with special focus on teachers and test-makers) of measurement principles and procedures to systematically design, appropriately implement, and fairly rate tests (Inbar-Lourie, 2008; Stiggins, 2001; Taylor, 2009). Despite the lack of common consensus on LAL concept (Fulcher, 2012), in light of Standards for Teacher Competence in Educational Assessment of Students by the American Federation of Teachers (AFT), it can be argued that an L2 teacher with a high level of LAL is specialized in “selecting and developing assessments for the classroom, administering and scoring tests, using scores to aid instruction, communicating results to stakeholders, and being aware of the inappropriate and unethical uses of tests” (, 1990, p. 74).

Given the pivotal roles that L2 teachers are expected to play in classroom assessment, such as collecting consistent and accurate information about L2 learners’ progress, using appropriately this information to modify their instructional practices, and offering effective feedback on the performance of their students, it should be assured that L2 teachers possess required LAL (Gebriel, 2017; Shepard, 2000; Sultana, 2019). In other words, due to these growing responsibilities and increasing expectations, L2 teachers need the required assessment knowledge and skills to pave the way for the success of this highly complex, demanding, vital, and multifaceted process of L2 learning. The very same necessity also applies to higher education context since, as Stiggins (1995) notes, university teachers need to “know the difference between sound and unsound assessment” (p. 240).

The nature of university teachers’ perceptions of LAL in the higher education context is of central importance. Along with Brown et al. (2011), it can be argued that assessment objectives may not be met owing to “lack of teacher cooperation, knowledge, or belief in the proposed new usage of assessment” (p. 75). For both theoretical and practical reasons, it is very crucial to examine how LAL is perceived by university teachers. In this way, university authorities can see if university teachers are equipped with an acceptable level of LAL. If not so, they can take the required steps to amend the shortcomings and boost LAL by running pre-service and in-service teacher training courses. Hence, the present research explored perceptions of university teachers about LAL in the Iranian EFL higher education context to provide valuable insights for different testing stakeholders.

Review of the literature

Dimensions of LAL

Over the years, a plethora of attempts have been made by leading scholars to provide a comprehensive model of LAL. For example, Boyles (2005) presents a collection of LAL competencies that are required for foreign language teachers in the USA context. These

competencies cover a range of abilities, including the ability to diagnose proper assessment methods, use various tools of assessment, analyze and interpret tests results, give answer appropriately to the test results and their meanings, and situate the assessment results properly in their teaching.

In a similar vein, a framework of fundamental competencies of LAL was suggested by Inbar-Lourie (2008). In his framework, he considers LAL “as a body of knowledge and research grounded in theory and epistemological beliefs, and connected to other bodies of knowledge in education, linguistics and applied linguistics” (p. 396) instead of taking it “as a collection of assessment tools and forms of analysis” (p. 396). The core competencies in his framework include the social role of language assessment, contemporary views about the essence of language competence, classroom and external testing practices, and the effects of language assessments on different testing stakeholders (Inbar-Lourie, 2008).

Informed by the previous works (Coombe et al., 2012; Davies, 2008; Fulcher, 2010, 2012; Inbar-Lourie, 2013; Malone, 2013; Hill & McNamara, 2011; Scarino, 2013; Taylor, 2009), Giraldo (2018) presents a comprehensive framework to illuminate different dimensions of LAL. His framework entails three central components. The first component is *knowledge* with three dimensions that deals with theoretical considerations, including *awareness of the basic tenets of applied linguistics*, *awareness of theories and concepts of assessment*, and *awareness of assessment context*. The second component is *skills* with five dimensions, namely *instructional skills*, *design skills for language assessments*, *skills in educational measurements*, and *technological skills*. The third component is *language assessment principles* that refers to familiarity with and actions toward critical issues in language testing practices. Based on this component, for example, teachers need to treat all students or test-takers with respect or make testing practices democratic and people-oriented by considering students’ voices and concerns in the test processes. It should be, nevertheless, noted that this framework suffers from two major drawbacks. Firstly, it is essentially theoretical-oriented and has not been built on empirical findings. Secondly, the components of LAL can be simply fallen into two broad categories: knowledge and skills. The components of language assessment principles can be put under the other two categories (knowledge and skills).

Research on teachers’ perceptions of LAL

There has been an increasing interest in LAL among researchers and practitioners over the last decades (Davies, 2008; Fulcher, 2010, 2012; Malone, 2013) which has led to a large body of research across different settings (Bøhn & Tsagari, 2021; Coombe et al., 2012; Homayounzadeh & Razmjoo, 2021; Janatifar & Marandi, 2018; Popham, 2009; Shahzamani & Tahririan, 2021; Stiggins, 1995; Vogt & Tsagari, 2014; Wang et al., 2006; Watmani et al., 2020; Xu & Brown, 2016; Yan & Fan, 2021; Zhou et al., 2020). Due to space limit, only some are reviewed to lay the groundwork for the current study.

In the Chinese context, Xu and Brown (2016) explored EFL teachers’ ($n=891$) conceptions about LAL using the adapted version of the teacher assessment language questionnaire (TALQ). They found that the participants suffer from insufficient assessment literacy levels. Furthermore, Lam (2019) investigated Hong Kong secondary school teachers’ ($n=66$) assessment literacy to measure L2 writing in a qualitative study. The findings revealed that although most of the participants had required assessment literacy,

some could not even make a distinction between “assessment of learning” and “assessment for learning.” Exploring teacher educators’ understandings of LAL in Norway was the focus of a study carried out by Bøhn and Tsagari (2021), indicating that for the participants, LAL includes four primary competences, namely disciplinary, assessment-specific, pedagogical, and collaboration. More recently, Yan and Fan (2021) conducted a qualitative study to explore the contextual and experiential factors affecting language testers, EFL teachers, and graduate students’ LAL profile in the Chinese context. Their results disclosed that the LAL profile differed for the three groups at individual and group levels. At the individual level, LAL development processes were different for each participant even though they shared some common patterns. At the group level, the language testers and graduate students enjoyed a higher level of LAL than the EFL teachers.

In Iranian contexts, Janatifar and Marandi (2018) carried out a study to verify the components of LAL through a modified version of Fulcher’s (2012) LAL survey. The responses of high school teachers ($n=280$) indicated that in the EFL context of Iran, LAL includes four major components, including test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects, and reliability and validity. In addition, the literacy assessment of EFL teachers ($n=88$) was compared with non-EFL teachers ($n=112$) in a recent study by Watmani et al. (2020). Using the teacher assessment literacy scale of Plake et al. (1993), the participants’ literacy assessment was measured. The findings were indicative of a huge lack of required literacy assessment for the participants along with the different perceptions of literacy assessment among the EFL and non-EFL teachers. Additionally, Homayounzadeh and Razmjoo (2021) conducted a study to measure LAL of instructors ($n=12$) and graduate and post-graduate students ($n=46$). They used Pastore and Andrade’s (2019) LAL framework which includes conceptual, praxiological, and socio-emotional dimensions. Their findings disclosed that there were discrepancies between the instructors’ and students’ LAL. For example, there existed differences between the two groups regarding how ethical requirements should be exercised at the socio-emotional dimension. Recently, Shahzamani and Tahririan (2021) carried out a study to gauge ESP instructors’ ($n=21$), content instructors’ ($n=8$), and ELT instructors’ ($n=13$) LAL to assess reading comprehension with respect to formative assessment. Their findings documented that the three groups had largely the same levels of LAL to assess reading comprehension.

As it can be implied from the above-mentioned studies, scarce attention has been given to university teachers’ perceptions of the fundamentals of LAL in the Iranian higher education context. In addition, to the best knowledge of the researchers, no qualitative study has already addressed teachers’ perceptions of LAL in the EFL context of Iran. Hence, the present qualitative study aims to further our understanding of LAL’s fundamentals by investigating university teachers’ perceptions in the Iranian higher education context. The hope is that this study’s results can improve the quality of assessment practices in the classroom by making a positive shift in university teachers’ conceptualizations of LAL.

Context of the study

This study was conducted in the academic context of Iran, namely Lorestan University and Ayatollah Burojerdi University, both of which are state universities. In general,

Iranian universities fall into three categories: state universities, non-governmental sectors (Azad), and distance education (Payam-e-Noor University) (Rasian, 2009). Given the fact that the present study was carried out at two state universities, their description is in order. The state universities are under the management of the Ministry of Science, Research, and Technology and dependent on the government's funding. They, in fact, represent the main higher education system to which the students that obtain a good rank on the National Entrance Examination (Konkur) can be admitted. Studying at the state universities is nearly tuition-free for students and university teachers with a robust professional resume and outstanding professional competence can be employed at these universities. The university teachers need to be present on the university campus 4 days a week. Their working hour is forty and their job duties include research, teaching, and academic services. As university teachers become qualified as associate professors and full professors, they allocate more time to research and assign less time to teaching. Furthermore, they are fully free to choose and deliver the course materials, opt for their own method of pedagogy, and decide on the type of assessment. It should also be noted that tertiary education in Iran is divided into four levels: Associated degree (A.A.), Bachelor of Arts (B.A.), Master of Arts (M.A.), and Doctoral of Philosophy (Ph.D.). Every academic year is run in two semesters which start in September and end in June (Rasian, 2009).

Method

Research design

LAL is construed as a multi-faceted and complex phenomenon which could be interpreted differently by various individual in diverse contexts. To capture such diversity, a phenomenographic approach was employed to collect the required data in this study. Phenomenography is an appropriate qualitative approach to disclose a group of people's perceptions of a particular phenomenon (Cresswell & Poth, 2018). According to Marton (1986), phenomenography is considered "as having the purpose of the qualitatively different ways in which people experience, conceptualize, perceive, and understand various aspects of, and various phenomena in, the world around them" (p. 31). It aims, as Marton (1986) adds, to "discover and systematize ways of thinking that synthesize how people interpret different aspects of reality" (p. 180).

Participants

A sample of 18 university teachers specialized in applied linguistics ($n=8$), linguistics ($n=6$), and the English literature ($n=4$) were selected using purposive sampling. The reason for using purposive sampling was that it allows researchers to identify, select, and categorize rich data cases in qualitative paradigm (Miles & Huberman, 1994). It should be noted that the data collected process continued until data saturation occurred; that is, the data happened to be repetitive and nothing new was shared by the participants. The participants included 10 male university teachers (55.55%) and 8 female ones (44.45%). Their teaching experiences ranged from 1 to 10 (27.7%), 11 to 20 (44%), to 21 to 30 (27.7%). The participants' demographic information is presented in Table 1.

Table 1 Demographic information of the participants

Participant	Gender	Level	Major	Teaching experience
UT1	F	Assi. Prof.	Linguistics	8
UT2	M	Assi. Prof.	English Literature	21
UT3	M	Assi. Prof.	Applied Linguistics	11
UT4	M	Assi. Prof.	Linguistics	9
UT5	F	Assi. Prof.	Applied Linguistics	17
UT6	F	Assi. Prof.	Linguistics	4
UT7	M	Asso. Prof.	English Literature	19
UT8	M	Assi. Prof.	Applied Linguistics	24
UT9	M	Assi. Prof.	English Literature	14
UT10	F	Assi. Prof.	Applied Linguistics	26
UT11	M	Asso. Prof.	Applied Linguistics	12
UT12	F	Assi. Prof.	Linguistics	6
UT13	F	Assi. Prof.	Applied Linguistics	29
UT14	M	Asso. Prof.	Linguistics	28
UT15	F	Assi. Prof.	Applied Linguistics	18
UT16	F	Assi. Prof.	English Literature	5
UT17	M	Asso. Prof.	Linguistics	14
UT18	M	Assi. Prof.	Applied Linguistics	15

UT university teacher

It is worthy to note that some ethical considerations were observed throughout the study. Prior to running the semi-structured interviews, the researchers informed the participants about the goals of the current study. The participants were assured that their privacy would be warranted such that their identities and their responses to the questions would be kept confidential during the semi-structured interviews. They signed an informed consent form in Persian. The researchers tried to create a positive and warm climate during the semi-structured interviews to let the participants express their perceptions with ease. Moreover, the participants were told that they could withdraw from the study whenever they wished. Finally, the researchers tried their best to present the participants' perceptions as accurately as possible by avoiding any biases and misinterpretations.

Instruments and data collection procedures

To collect the required data, a phenomenographic semi-structured interview (Marton & Booth, 1997) was administered wherein the participants were invited to reflect upon their perceptions of LAL. Initially, an interview checklist was developed by undertaking the following steps. Firstly, a thorough investigation of the literature pertaining to LAL was accomplished (e.g., Brookhart, 2013; Eyal, 2012; Firoozi et al., 2019; Fulcher, 2012; Giraldo, 2018, 2021; Inbar-Lourie, 2013; Pastore & Andrade, 2019, Taylor, 2013; Zhou et al., 2020). In a sense, the researchers examined if there was any theory, framework, and instrument that might have already been used to measure the relevant constructs (Dörnyei, 2007). Based on the prior literature, an initial draft of the broad concepts of LAL was prepared. Afterward, these broad concepts were analyzed and broken down

into smaller segments in the form of questions. Then, to appraise the validity of the checklist, two university professors specialized in second language assessment at Tehran University were invited to comment on its content and language. In accordance with their comments, the vague and problematic items were either amended or discarded from the checklist. In principle, the checklist revolved around some fundamental concepts of LAL, including involvement of students in assessment, alternative assessment methods, traditional assessment methods, digital language assessment literacy, ethical requirements, standardized test administration, reporting test results, performance assessment, students' individual differences, grading procedures, test validity, test reliability, wash-back effects, and university teachers' pedagogical content knowledge.

Having designed and validated the checklist, the semi-structure interviews were administered. They began with an open-ended question: What does language assessment literacy mean to you? Then, the first researcher continued the interviews with the rest of the questions in the checklist, yet when needed, he also raised additional, relevant questions to elicit more expanded and elaborate responses. The checklist, in fact, helped the researchers to run the interviews more systematically (Patton, 1990). The semi-structured interviews were run in the participants' mother tongue (in Persian) to allow the participants express their ideas with greater ease and avoid the barriers that the second language might put up in their way. The participants' words were recorded by a voice recorder to be meticulously transcribed and analyzed later. To increase variations among the participants' perceptions, as noted by Durden (2018), the first researcher ran the semi-structured interviews in diverse temporal and spatial contexts. The underlying reason was not to let the time and place of the interviews affect the participants' conceptions about LAL. It should be noted that each interview lasted around 45–60 min.

Data analysis

The collected data was analyzed according to the multi-stage approach proposed by Marton (1986). This phenomenographic approach includes six distinct stages detailed below. At the first stage, prior to analyzing the data, the researchers tried to identify the preconceived conceptions. For this purpose, the researchers meticulously went through the collected data and wrote down an initial draft of the preconceived perceptions. Then, having identified the primary conceptions about LAL, the researchers focused on the participants' actual words and extracted the pre-categories from the participants' actual words. They tried best to be as faithful as to the conceptions presented in the participants' actual words. At the next stage, the researchers compared and contrasted a single participant's words with other participants' ones. During this stage, the researchers tried to spot the regularities embedded in the participants' words and they created increasingly larger pre-categories. In addition, they sought to identify the connections among the diverse pre-categories. As suggested by Sjostrom and Dahlgren (2002), at the fourth stage, the pre-categories were analyzed based on three principles: frequency (i.e., the number of times a conception related to LAL was mentioned by the participants), position (i.e., determining the relevance of a given conception to the other conceptions, and verifying how much it is relevant), and pregnancy

(i.e., identifying the conceptions related to LAL that received the most attention by the participants). At the fifth stage, the researchers selected the excerpts that mostly represented the categories. During this stage, the researchers took special care to stick to the intended meanings of the conceptions expressed by the participants. At the sixth stage, the researchers analyzed the pre-categories in different meaning pools. To establish the final categories, the researchers examined these meaning pools first individually and then cross-sectionally. As Åkerlind (2017) stresses, this strict multiple-stage approach was followed to guarantee the quality and transferability of the results.

It should be stressed that some measures were undertaken to enhance the reliability and validity of the findings. Regarding the findings' reliability, the data were coded concurrently by two researchers. Their codings' inter-reliability was 0.87 which was acceptable for the current study's purposes. When a disagreement raised, they solved it by discussing the coding. Concerning the accuracy and the credibility of the findings, member checking strategy and respondent validation were used. As Lincoln and Guba (1985) note, member checking is a common strategy to establish findings' credibility. To this aim, all the extracted constructs and concepts were reviewed by three analysts. Then, a copy of them were given to five of the participants to approve, add more comments, and assure their meanings. This technique allowed the respondents to check out their intentions, rectify mismatches, and include additional information. Based on their initial feedback, some modifications were made and they were asked to check the modifications. The participants confirmed that the extracted constructs and concepts represented their intended conceptions.

Results and discussion

Having analyzed the collected data, the researchers gained two overarching categories in LAL: *knowledge* (e.g., having an acceptable level of digital LAL, satisfying ethical requirements, benefiting more from performance assessment, considering students' individual differences, making assessment valid, assuring that tests are reliable, and having an acceptable level of pedagogical content knowledge) and *skills* (e.g., involving students in assessment, using alternative assessment methods, employing logically traditional assessment methods, informing students about test results, administering tests in standardized ways, using valid grading procedures, and bringing positive wash-back effects). They are detailed below (Fig. 1).

Knowledge

Having an acceptable level of digital LAL

Digital language assessment literacy (DLAL) was the first theme catching the participants' attention. Digital assessment literacy is defined as "the role of the teacher as an assessor in a technology-rich environment" (Eyal, 2012, p. 37). Concerning the significance of DLAL, one of the participants commented:

1. Despite the paramount importance of digital language assessment literacy, most university teachers are not equipped with it. For example, I myself don't know how to use the options of a platform like Adobe Connect to administer an online test. (UT17)

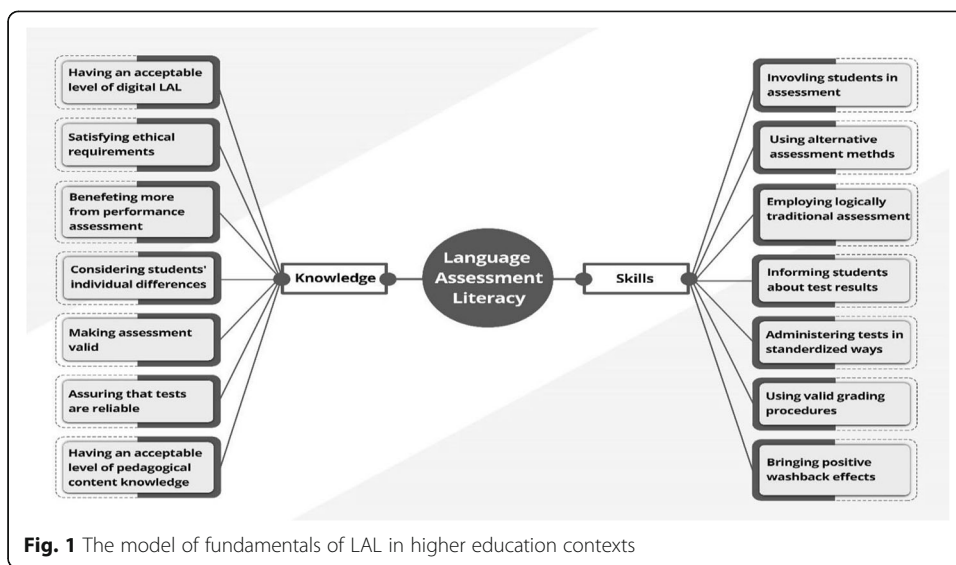


Fig. 1 The model of fundamentals of LAL in higher education contexts

Corroborating with the previous statement, another participant stated:

2. With the non-sustaining spread of online classes across the country, we need to make our digital assessment literacy update. To be honest, we all have difficulties assessing our students' learning in online classes. We most often just give some questions via Whats'app and students usually answer them on an open-book format. (UT11)

The statements above show that DLAL is perceived as an integral part of LAL. In alignment with Eyal (2012), the study's findings evidence that teachers should be equipped with the required literacy to assess students' performance by using modern digital technologies. The study's findings may be explained from this view that the increasing interest in using modern digital technologies in language assessment programs is associated with efficiency and availability they provide for testing stakeholders (Chappelle, 2008). Furthermore, these results may be discussed from the perspective that as there exist constant advances in technology and the primacy of computer-assisted language learning (CALL), university teachers need to be updated in this domain (Winke & Fei, 2008). Also, the findings support Eyal (2012), arguing that teachers should possess the required level of literacy to be efficient assessors in the digital era.

Satisfying ethical requirements

"Satisfying ethical requirements" was another recurring theme elicited from the data. The university teachers maintained that assessment practices should meet ethical requirements. In this regard, one of the participants remarked:

3. Students should perceive testing practices as fair. For this, I try to clarify the grading criteria and announce the objectives of tests, in advance. In this way, my students can become motivated to demonstrate their abilities better. (UT17)

In line with the previous statement, another participant stressed the importance of the ethical requirements in assessment practices as follows:

4. I never let my students' scores be polluted by irrelevant constructs, such as cheating and gender bias. I mean that I do not allow the quality of my assessment practices to be adversely affected by test-takers' gender. Or, I do not permit my students to cheat during test administrations to increase their scores unfairly. (UT12)

These statements documented the significance of considering ethical requirements in assessment practices. This could be explained by the point when tests are perceived as unfair by students, they may make students feel angry, confused, powerless, embarrassed, and frustrated (Buttner, 2004; Chory et al., 2017; Horan et al., 2010). Another possible explanation could be students' perceptions of un/fairness of testing practices which may cause students to show positive and negative affective reactions, as well as behavioral responses (Rasooli et al., 2019). Additionally, in alignment with Fan et al. (2020), the study's results can be discussed from this perspective that if assessment practices do not consider the ethical requirements, the decisions taken based on test results may not be very accurate. In turn, these incorrect decisions may adversely affect students' life and destiny. This study's findings show support to the previous studies, revealing that ethics in assessment practices has an immense effect on student learning (Holmgren & Bolkan, 2014), amount of students' engagement (Berti et al., 2010), and students' motivation to continue learning (Chory-Assad, 2002).

Benefiting more from performance assessment

Another theme which was frequently highlighted by the participants was "benefiting more from performance assessment." In this respect, one of the university teachers noted:

5. Students' activities over the course are of paramount importance. In my final evaluation, students' class attendance, students' class participation, students' responses to my oral questions are important. (UT13)

Besides, one of the respondents underscored the significance of performance assessment as follows:

6. A major part of students' final scores is the tasks that they do during the course. For example, every week I encourage my students to read passages, summarize them in their own words, and present them in the classroom. Or, the stage is ready for students who like to prepare a reading and present it in the classroom. (UT10)

The extracts evidence that knowing how to implement performance-based assessment is essential for university teachers. The findings can be discussed from the perspective that performance-based assessment such as doing a project may give a better

picture of the processes involved in student learning (Backman, 2002). Furthermore, along with Bachman and Palmer (2010), it may be argued that performance-based assessment is more useful because it permits students to apply their knowledge and skills to solve real-life problems. This argument receives support from Brindley (1994), who considers performance assessment as a way to appraise “both knowledge and ability for use” (p. 75). For example, as the findings demonstrated, valuing students’ participation in instruction can be so facilitative for students to use their L2 competence to do a task. Additionally, in alignment with Backman (2002), it could be noted that performance assessment should be practiced since it measures complex abilities that traditional assessment practices are not able to measure easily. That is, performance assessment exposes students to more cognitively and affectively challenging tasks compared to the traditional assessment. The study’s findings are in consistent with Giraldo (2018, 2021), asserting that LAL includes knowledge of the advantages and disadvantages of implementing performance assessment.

Considering students’ individual differences

“Considering students’ individual differences” was the other theme emerged from the respondents’ words. In this regard, a university teacher remarked:

7. Testing practices would be effective if students’ individual differences, such as learning styles, cultural background, age, and first language be taken into account. For one thing, some students are good at answering close-ended items while other students may feel more at ease with open-ended items. (UT6)

Additionally, another participant complained about ignoring students’ individual differences in assessment practices in the classroom. He commented:

8. Since in Iran, students come from different cultural and ethnic backgrounds, tests should be value-free in terms of culture and ethnicity. For example, I remember that some of my students were irritated in the last semester due to some value-laden questions in the final exam. (UT14)

The participants’ words disclosed that university teachers should be aware of students’ individual differences in assessment practices. The study’s findings may be explained by this well-recognized fact that student learning is highly affected by individual characteristics such as interests, learning styles, and motivation; also, student performance is likely to be influenced by such individual characteristics during test taking (Tomlinson & Moon, 2013; Wang et al., 2006). Furthermore, in consistent with Afl (2009), the results may receive support from the view that since different students with different cognitive and affective individual differences approach tests differently, test results are likely not to be the same for all test-takers. Another possible explanation for the study’s results, as Tomlinson & Moon, 2013 and Moon (2016) note, is that to provide enough opportunities for students to demonstrate their knowledge and understanding of materials, differentiated assessment should be practiced. Moreover, the study’s findings may trace back to this view that when assessment is differentiated for

varying students' individual differences, it is more equitable (Moon, 2016; Tierney, 2013). The study's findings accord with those of Crosthwaite et al. (2015), reporting that students with different cognitive and affective individual differences consider teachers' assessment differently. For example, teachers' assessment of class participation was fair for students who valued class participation while it was unfair for the students who were more interested in working alone.

Making assessment valid

The other recurrent theme emerging from the data was "making assessment valid." To stress the significance of validity, one of the participants asserted:

9. I make my best effort to assure that my tests are visually appealing, their contents represent the course syllabus, and they met the principles of communicative approaches. This makes the test results positively affect students' life and destiny. (UT15)

In addition, one the participants lamented that validity of tests usually remains unnoticed in assessment practices.

10. Though validity guarantees test quality, most of the testing practices lack the required validity. The possible reason for this problem is that test validation is highly complex, demanding, and, of course, expensive for teachers. (UT18)

The statements above highlight the significance of validity in assessment practices. One possible reason for the study's findings is that if requirements of test validation are not satisfied, tests cannot measure what they are intended to measure (Fulcher & Davidson, 2007; McNamara, 2006). Furthermore, study's results support this long-standing view that test validation incorporates three essential elements, including content, construct, and criterion (Bachman & Palmer, 2010; Hamp-Lyons & Lynch, 1998). In alignment with Messick (1994, 1996), the study's findings may be explained from the viewpoint that validity gives meaning to test scores; that is, validity evidence shows that test performance is closely linked to performance outside the class. Moreover, the study's results may be explained from this view that, as McNamara (2006) asserts, validity increases the degree to which correct interpretations and decisions can be made about students' life. Likewise, as the study's demonstrated, despite the utmost importance of test validation, it is not widely practiced in classroom assessment. The underlying reason for this problem is that measuring test validity is complex, demanding, time-consuming, and expensive for university teachers (Bachman, 1988, Bachman, 1990). In brief, validity is the backbone of any assessment practices and should be given enough attention by university teachers.

Assuring that tests are reliable

"Assuring that tests are reliable" was the other theme that gained remarkable attention by the participants. In this respect, one of the participants stated that:

11. Reliability and consistency of test results should be acceptable. For this, for example, the number of items should be large enough, they need to cover all the contents of course, and they should be of different kinds from multiple-choice item to essay items. (UT11)

In congruence with the precedent statement, the participants stressed that an integral part of LAL is knowledge of basic statistics.

12. Having knowledge of basic statistics is important for a teacher. I mean, if we aim to have reliable tests, we should know how to estimate reliability through the available statistical formulae. (UT3)

These statements are indicative of the significance of test reliability in assessment practices. One possible reason for the study's findings, as Bachman, 1988, Bachman, 1990) highlights, is that reliability is very essential to assure that our tests can consistently measure the intended characteristics. Moreover, the study's results could be explained from this perspective that the results of reliable tests are of great value in terms of time-saving for university teachers. That is, for busy university teachers, when a test enjoys a high level of reliability, it can be assumed that test results have not been adversely affected by students' temporary psychological and physical state, environmental factors, test forms, and multiple raters (Bachman & Palmer, 2010; Fulcher & Davidson, 2007). Additionally, in alignment with McNamara (2006), the study's results may be discussed from this respect that reliable tests generate dependable, repeatable, and consistent information about students' abilities. In consequence, this information makes way for more meaningful interpretations of test results and correct decisions about students' life and destiny.

Having an acceptable level of pedagogical content knowledge

The last theme in the domain of knowledge was "having an acceptable level of pedagogical content knowledge." Pedagogical content knowledge, as defined by Rodabaugh (1994), is "a teacher's internalized procedural knowledge of how particular topics, problems and issues can be organized, adapted and presented to learners with diverse interests and capabilities" (p. 12). Concerning the importance of this factor, one of the participants asserted:

13. Though it seems irrelevant, pedagogical content knowledge is an integral part of LAL, I believe. My reason for this is that when teaching is not done effectively, the testing practices lose their meanings and effects. So, familiarity with the basics of applied linguistics is quite essential for university teachers. (UT11)

This quote indicated that university teachers should be equipped with sufficient pedagogical content knowledge. The study's findings may be explained from this respect that by having an acceptable level of pedagogical content knowledge, university teachers know teaching methodologies, subject knowledge, and how to adapt them to learning purposes. As this is done well, testing practices can be found meaningful by

students (Rasooli et al., 2018). Additionally, in alignment with Rasooli et al. (2018), it could be argued that when university teachers are not equipped with the required pedagogical content knowledge, they are likely to fail to explain and deliver the learning contents adequately. This, in turn, adversely affects students' opportunities to learn and show their abilities. The study's findings are in accordance with the previous studies (Chory, 2007; Lankiewicz, 2014; Rodabaugh, 1994), reporting that students' perception of un/fairness of assessment practices is highly affected by pedagogical content knowledge of teachers. Likewise, in line with the study's findings, Bempechat et al. (2013) found that students regarded teachers' testing practices as unfair and ineffective due to the failure of their teachers to illuminate the assignments in-depth, allot time to address students' confusions, and cover the lessons patiently.

Skills

Involving students in assessment

The first theme related to *Skills* domain was "involving students in assessment" (Fig. 1). As its name suggests, the participants maintained that university teachers should engage students in testing practices. In support of this, one of the participants noted:

14. When I involve my students in testing practices, I feel that they become more motivated to be in class. When their voices are heard, they feel more self-confident and their self-efficacy increases. (UT2)

In accordance with the precedent statement, another participant approached the issue as follows:

15. Well, we need to incorporate our students in our assessment practices, since this makes them monitor and self-regulate their learning and get a better picture of their actual abilities. (UT5)

The above statements disclosed that students should be engaged in assessment practices. One possible reason for the study's findings, as Nicol and Macfarlane (2006) note, could lie in the point that when students are involved in assessing their abilities, their self-regulation competence increases. According to Zimmerman and Schunk (2011), self-regulated individuals are capable of monitoring, controlling, and regulating their own cognitions, affect, and behaviors; thus, they are in a better position to reach their goals. Additionally, along with Asghar (2010), the study's results may be explained from this view that involvement of students in testing practices may positively affect their motivation and self-efficacy by helping students believe that they have the required abilities to do testing tasks. Furthermore, the study's results may be justified from this respect that students may control their anxiety (Hayes & Embretson, 2013) and disturbing emotions better (Tapia & Panadero, 2010) in a testing context where they see themselves as an integral part of it. The study's findings provide support to those of Pintrich and De Groot (1990), indicating that involving students in assessment practices brings about noticeable positive benefits for student learning and achievement. Likewise, the study's results show support to the previous studies (Sivan, 2000; Tillema et al., 2011),

reporting that an effective strategy to develop quality assessment is incorporating students' voices and concerns in the design, implementation, and scoring of tests.

Using alternative assessment methods

"Using alternative assessment methods" was another theme that gained noticeable attention by the respondents. In this respect, one of the participants expressed disdain with not using alternative assessment, such as self-assessment, peer-assessment, and portfolio assessment in the Iranian higher education contexts. His exact words are reported below:

16. Though using alternative assessment methods may be very advantageous, they are rarely exercised by university teachers. This reservation may be due to the implementation difficulties with such assessment methods. For example, administering peer-assessment is really time-consuming and using portfolio assessment demands diverse skills. (UT7)

Additionally, the participants pinpointed when university teachers are skilled enough to implement alternative assessment methods, their assessment practices are more advantageous for student learning. For this, a participant commented:

17. Well, sometimes I try to use alternative assessment methods such as self-assessment. I've found that self-assessment helps students reflect upon their performance. And based on their reflection, they can self-regulate better their future learning by rectifying their errors. (UT1)

The university teachers' words represented that LAL encompasses the required skills to implement alternative assessment methods. One possible explanation for the findings may be associated with the increasing interest in learner-centered pedagogy opposed to teacher-centered pedagogy over the last decades (Aryadoust, 2016; Topping, 2009). In this regard, the alternative assessments, such as peer-assessment and self-assessment have been introduced and practiced as viable alternatives to consider assessment for learning (Rosman et al., 2015). Moreover, in alignment with Black and William (1998), the study's results may be explained from this perspective that alternative assessment methods can generate valuable information that can be used to offer feedback tailored to students' needs and interests. This, in turn, may promote student learning. Also, the study's findings can be discussed more with reference to self-regulation (Pintrich, 2000). Alternative assessment methods, most specially self-assessment, as Andrade and Brookhart (2014) stress, make students' self-regulation competence increase. As a consequence, the students' awareness on tasks' goals would be raised and they can check their progress toward them. The study's findings are partially in consistent with those of Heidi and Du (2007), revealing that their participants acknowledged the benefits of self-assessment in helping them find out the assignments' expectations, identify their weaknesses, and plan to rectify them. Additionally, the study's findings lend credence to the studies done by Black and William (1998). In brief, their results revealed that owing to the implementation of peer-assessment, the

learning quality increased and both low-achiever and high-achiever students could benefit from the instructions.

Employing logically traditional assessment methods

The other theme extracted from the collected data was “employing logically traditional assessment methods.” The participants pinpointed that traditional assessment methods should not be frowned upon completely when it comes classroom assessment.

18. It is not fair to say that traditional assessment methods should be completely discarded. For example, one of their biggest advantages is their practicality. I can implement them in diverse contexts for different purposes. (UT5)

Furthermore, another university teacher confirmed the previous perception, maintaining that:

19. This is a reality that most university teachers use traditional assessment methods, such as multiple-choice, true-false, matching, and gap-fill questions. Well, one underlying reason for such extensive interest is that they can easily compare students' performance. (UT4)

As the participants' words evidence, the university teachers stressed that there are some undeniable advantages with traditional assessment methods if used appropriately. Easy, fast, and economical design, implementation, and grading seem to be the reasons for such strong advocate of the traditional assessment methods (Bachman & Palmer, 2010; Bailey, 1998). Another possible reason of this may be related to the easy analysis of their results. That is, as Bachman, 1990) argues, teachers have fewer difficulties analyzing and comparing students' scores over time and across a large, diverse group of students. Another possible explanation for the findings, as Dikli (2003) points out, is that traditional assessment methods are easy to score and let teachers determine which areas their students have excelled in and which they are struggling with. In light of the findings, as Bailey (1998) and Dikli (2003) discuss, the tendency of the university teachers toward traditional assessment methods lies in the fact that they “look like” tests, and hence, there is less resistance against them on the part of students.

Informing test-takers about test results

“Informing test-takers about test results” was another recurring theme extracted from the participants' responses. To confirm the significance of informing test-takers about test results, one of the participants noted:

20. I don't know why my colleagues do not often turn back students' test sheets with enough feedback. As I do it, not only do my students understand that tests are important but also they see their mistakes and can modify their inter-language systems. (UT10)

Another relevant issue related to this theme was keeping students' scores confidential. In this regard, one of the university teachers quoted:

21. I think it is really unethical to announce students' scores before the class. This may damage self-efficacy and self-confidence of the students who have taken a low score. In consequence, they may lose their motivation to continue learning. (UT12)

As it can be implied from the participants' responses, university students need to be informed about tests' results and receive enough feedback about their performance, and their scores should be kept confidential. One possible explanation for the study's findings is that a large segment of teachers' career duties is to extract, diagnose, and offer appropriate feedback to student learning in a constant, ongoing, communicative manner (Cowie & Bell, 1999). What is more, as Moss and Brookhart (2009) highlight, the feedback given to students' errors opens valuable opportunities for the students to modify or consolidate their learning. Furthermore, along with Tunstall and Gipps (1996), the study's results can be explained from this respect that keeping students informed about test results can increase their motivation. That is, when students receive positive feedback ability from their teachers and see their own scores, they may feel a sense of achievement and increase their future effort. Moreover, the study's results can be justified from this view that fair assessment requires keeping students' scores confidential (Rasooli et al., 2018), since publicizing students' scores may hurt the self-efficacy of those students who have not got high scores (Green et al., 2007). The study's findings are in accordance with the previous studies (Carless, 2006; Lizzio & Wilson, 2008). They reported that the presence of effective feedback was conceived as an overarching dimension of quality and fair testing practices. To close, university teachers with high level of LAL are aware of the value of informing students of their own test results but not publicizing them.

Administering tests in standardized ways

"Administering tests in standardized ways" was another theme that was repeatedly highlighted by the participants.

22. I try my best to administer tests in a standardized way. For example, prior to running tests, I consult the date with my students, I explain the staff they need to bring in test sessions, I clearly determine the location wherein tests will be run, and I inform them about the time duration of tests. (UT13)

Another respondent also expressed disdain for the lack of standardized administrations in the end-of-semester exams. She remarked:

23. Test administrations at universities are rarely standardized. For example, the test setting is usually noisy, proctors are not well trained to treat appropriately test-takers, there is not enough accommodation for students who are left-handed or disabled, and there is no break during the test administration. (UT8)

The statements above indicate that a major part of LAL is related to the required skills to administer tests in standardized ways. In alignment with DeLuca et al. (2016) and Abedi et al. (2004), the study's findings are discussed from this view that standardized administrations aim to present students with equal opportunities to show what they have learned and what they can do with it. That is, there should be fairness and equity with test administrations and accommodations so that no a single group of students receive special advantages. As Haladyna et al. (1991) assert, another reason of the study's findings is that one of the major sources of students' score pollution is the actual administration of tests. For example, when students are interfered with responses, they are ridiculed or scorned by proctors and other test-takers on test days, or they get anxious and tired due to the uncomfortable test settings, their scores may be adversely polluted. The study's findings provide support to Abedi et al. (2004) who found that students may feel a sense of unfairness when test administration is not standardized by advantaging some students. Additionally, the study's findings are in keeping with those of Siegel (2007, 2014), showing that while accommodation and standardized test administrations, on the surface, are considered as an integral part of LAL, the practices need more attention to respond differentially to students' special learning needs.

Using valid grading procedures

"Using valid grading procedures" was another theme gaining remarkable attention by the participants. The participants stressed that the grading procedures should be transparent, consistent, and logical.

24. I believe that the grading criteria should be clear and fair for all students. They should know how their performance would be graded and what parts of the materials have more weigh in the final evaluation. In this way, students' scores are got less polluted by irrelevant factors. (UT14)

Corroborating with the previous statement, another respondent was of the following opinion:

25. When the grading procedures are not vague and we communicate about them, students can manage their time and energy in a better way and demonstrate their abilities well. (UT4)

These remarks show that an integral part of LAL is using valid grading procedures. Grading procedures, according to Tata (1999), should be transparent, consistent, and unbiased. That is, grading procedures are perceived as fair by being consistent, using accurate information, and maintaining an impartial process. The study's findings may be discussed from this respect that university teachers are expected to apply valid grading procedures consistently and impartially to all test-takers. This is due to the fact that by implementing valid grading procedure students' scores would not be polluted by irrelevant constructs (Gordon & Fay, 2010). Another possible explanation for the study's results is that by being informed about the grading procedures, students can manage

their studies to get better scores. In turn, as Tata (1999) puts it, high grades offer both immediate benefits for students like intrinsic motivation and approval of parents, as well as long-term benefits like admission to graduate school and preferred employment. The study's findings are consistent with the previous studies (Gordon & Fay, 2010; Rasooli et al., 2018; Tata, 1999), reporting that grading procedures have a great impact on students' perceptions of test fairness, teachers' assessment quality, and student learning.

Bringing positive wash-back effects

The last theme highlighted by the participants was "bringing positive wash-back effect." That is, the participants maintained that university teachers need to have the required level of LAL to implement the tests in such a way that they the tests have positive effects on test-takers' lives and education.

26. An important part of LAL, I think, is the ability to design tests that can bring positive wash-back effects. For example, when they are developed based on the tenets of communicative approaches, students are encouraged to spend more time improving their communicative competence. Because they know that to pass the tests, they need to handle communicative tasks. (UT16)

In alignment with the previous statement, another respondent commented:

27. Given the effects of testing practices on my teaching, I usually pay particular attention to the design, administration, and grading of my tests. By making my test quality, my students know that they should put more time and energy into their studies. (UT1)

These words indicate that university teachers need to develop their tests in a way that they create positive wash-back effects. One possible reason for these findings is the reality that instruction and assessment go hand-in-hand (Bachman & Palmer, 2010). That is, along with Lantolf and Peohner (2014), if the instruction is going to be effective, it must entail assessment, and simultaneously, fair assessment is not possible without considering instruction. Additionally, coupled with Green and Johnson (2010), the study's findings may be explained from this perspective that testing practices are administered to achieve two purposes: assessment of learning and assessment for learning. Test results, on one hand, are summative and can be used to make high stake decisions like college admission and they, on the other hand, are diagnostic and formative used to inform instruction (Fan et al., 2020). Moreover, the study's findings are indicative of this view that, as Tsagari (2009) notes, the examination is a strong instrumental motivation for students to learn in the examination preparation. The study's findings receive support from the previous studies (Cheng et al., 2011; Tsagari, 2009). They found that students had a tendency to select activities intended for test orientation or test-specific coaching to prepare for a particular examination. To close, as Brown (1997) points out, "if you want to change student learning then change the methods of assessment" (p. 7).

Conclusion

The present study intended to examine the fundamentals of LAL from the Iranian university teachers' perceptions. Findings yielded two overarching LAL domains: *knowledge* (e.g., having an acceptable level of digital LAL, satisfying ethical requirements, benefiting more from performance assessment, considering students' individual differences, making assessment valid, assuring that tests are reliable, and having an acceptable level of pedagogical content knowledge) and *skills* (e.g., involving students in assessment, using alternative assessment methods, employing logically traditional assessment methods, informing students about test results, administering tests in standardized ways, using valid grading procedures, and bringing positive wash-back effects). As the study's findings demonstrated, it can be concluded that university teachers should be equipped with required LAL to make the way for quality education.

In light of the study's findings, some implications are presented. The first implication is that university students' voices should be heard and they should be given the right to question test processes. By doing so, university students feel treated fairly and empowered when they see that their voices and concerns are influential in their life and destiny (Hamid & Hoang, 2016). Another implication is that alternative assessment methods, such as peer-assessment, self-assessment, and portfolio assessments should be more practiced. The consequences are likely to be positive if they are implemented with consideration of the contextual factors and students are supported and trained how to use them (Brown & Harris, 2014; McDonald, 2013), discuss and agree on grading criteria (Brookhart, 2013), and have experience with the subject (Morrison et al., 2014). The next implication is that university teachers' assessment can be perceived as effective, efficient, and equitable, if they accommodate students' individual differences (e.g., interests, readiness levels, age, gender, culture, first language, learning styles) (Harris & Brown, 2013; Rasooli et al., 2018). The following implication is that university teachers should design and implement tests to create positive wash-back effects. To meet this aim, they can implement more performance-based assessment, for example, by having students participate in class discussions or do term projects. A further implication of the findings is that university officials should be aware of the central importance of LAL and take up the required steps to boost it by running pre-service and in-service teacher training programs. Indeed, the study's findings, along with Xu and Brown (2016), call for joint attempts by university officials, teacher educators, and university teachers for making up consistent literacy assessment programs throughout university teacher professional life. The final implication is for the researchers in LAL field. They should be on the alert that LAL components may change over time as views on L2 teaching change.

While the current study's findings provide a clear picture of the fundamentals of LAL in the Iranian higher education context, they should be interpreted in light of the limitations imposed on the study. The present study probed university teachers' perceptions of LAL components in classroom contexts. To reach a more comprehensive conceptualization, future research can explore the university students' conceptions about LAL components. Furthermore, while this study provided some valuable insights into the effects of pedagogical content knowledge on the efficacy of university teachers' assessment practices, interested researchers can investigate how university teachers' pedagogical content knowledge can affect students' perceptions of quality assessment.

In addition, in alignment with the findings, future research can empirically explore the kind and amount of wash-back to classroom assessment with the application of LAL components. Additional research is also required to identify grading criteria considered as efficient by both university teachers and students. Future studies can also research digital assessment literacy from the perceptions of both university teachers and university students. Last but not least, more research is required to explore university teachers' and students' perceptions of the effects of informing students about test results on their learning.

Appendix

Interview checklist

1. What does language assessment literacy mean to you?
2. In your opinion, who is a literate assessment university teacher?
3. Do you think that you should involve your students in CA?
4. Do you think that ethical requirements should be met in CA?
5. What is your opinion about traditional assessment methods such as multiple-choice, true-false, and so on?
6. What is the importance of making tests valid in CA?
7. Why should university teachers make their tests as reliable as possible?
8. Do you think that a university teacher should make the grading valid?
9. Can we consider the skills of administering tests in a standardized way as a part of LAL?
10. Do you agree that bringing a positive wash-back effect is a component of LAL?
11. Can we say that pedagogical content knowledge is an integral part of LAL?
12. Why should we regard digital language assessment literacy as a part of LAL?
13. Should a university teacher take the individual difference of test-taker into account?
14. Do you agree that a university teacher should inform the students about test results?
15. Why should accept that using alternative assessment methods is an essential part of LAL?
16. Can we say that literate assessment university teachers use more performance assessment?
17. Is there anything else that you may want to add?

Abbreviations

LAL: Language assessment literacy; CA: Classroom assessment; DLAL: Digital language assessment literacy; UT: University teacher

Authors' contributions

The study was planned and implemented by Dr. Rezai, as the major author, along with the cooperation of Dr. Alibakhshi, Dr. Farokhipour, and Dr. Miri in the data collection and analysis phases as well as revising the manuscript. In the end, the four authors read and confirmed the final manuscript.

Funding

No grant, fund, or other supports was received by the authors.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The authors affirmed that during formulating and completing the current study, all ethical requirements were considered and satisfied.

Competing interests

There is no competing interest.

Author details

¹Department of Teaching English and Linguistics, Ayatollah Ozma Burojerdi University, Borujerd City, Iran. ²Teaching English and Linguistics Department, Faculty of Literature and Humanities, Ayatollah Ozma Burojerdi University, Lorestan Province, Burojerd City, Iran. ³Department of English Language and Literature, Allameh Tabataba'i University, Tehran, Iran. ⁴Department of English Language and Literature, Shahid Mahalati Higher Education Center, Iran.

Received: 7 July 2021 Accepted: 23 September 2021

Published online: 03 December 2021

References

- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1–28. <https://doi.org/10.3102/00346543074001001>.
- Afl. (2009). *Position paper on assessment for learning*. Third International Conference on Assessment for learning. NZ: Dunedin Retrieved from www.fairtest.org/position-paper-assessment-learning.
- Åkerlind, G. S. (2017). What future for phenomenographic research? On continuity and development in the phenomenography and variation theory research tradition. *Scandinavian Journal of Educational Research*, 60(3), 1–10. <https://doi.org/10.1080/00313831.2017.1324899>.
- American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), & National Education Association (NEA). (1990). *Standards for teacher competence in educational assessment of students*. Buros Center for Testing.
- Andrade, H., & Brookhart, S. (2014). *Toward a theory of assessment as the regulation of learning*. Symposium presentation at the annual meeting of the American Educational Research Association. Philadelphia, PA.
- Aryadoust, V. (2016). Gender and academic major bias in peer-assessment of oral presentations. *Language Assessment Quarterly*, 13(1), 1–24. <https://doi.org/10.1080/15434303.2015.1133626>.
- Asghar, A. M. (2010). Reciprocal peer coaching and its use as a formative assessment strategy for first-year students. *Assessment & Evaluation in Higher Education*, 35(4), 403–417. <https://doi.org/10.1080/02602930902862834>.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Bachman, L. F. (1988). Problems in examining the validity of the ACTFL Oral Proficiency Interview. *Studies in Second Language Acquisition*, 10(2), 149–164. <https://doi.org/10.1017/S0272263100007282>.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Backman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 1–24. <https://doi.org/10.1191/0265532202lt240oa>.
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Heinle & Heinle.
- Bempechat, J., Ronfard, S., Mirny, A., Li, J., & Holloway, S. D. (2013). She always gives grades lower than one deserves: A qualitative study of Russian adolescents' perceptions of fairness in the classroom. *Journal of Ethnographic & Qualitative Research*, 7, 169–187.
- Berti, C., Molinari, L., & Speltini, G. (2010). Classroom justice and psychological engagement: Students' and teachers' representations. *Social Psychology of Education*, 13, 541–556. <https://doi.org/10.1007/s11218-010-9128-9>.
- Black, P., & William, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. <https://doi.org/10.1177/003172171009200119>.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222–233. <https://doi.org/10.17507/jltr.1202.02>.
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National assessment summit papers* (pp. 11–15). Iowa State University.
- Brindley, G. (1994). Task-centred assessment in language learning: The promise and the challenge. In N. Bird, P. Falvey, A. Tsui, D. Allison, & A. McNeill (Eds.), *Language and learning: Papers presented at the Annual International Language in Education Conference, Hong Kong, 1993* (pp. 73–94). Hong Kong Education Department.
- Brookhart, S. M. (2013). The public understanding of assessment in educational reform in the United States. *Oxford Review of Education*, 39(1), 52–71. <https://doi.org/10.1080/03054985.2013.764751>.
- Brookhart, S. M., & Bronowicz, D. L. (2003). I don't like writing. It makes my fingers hurt: Students talk about their classroom assessments. *Assessment in Education: Principles, Policy & Practice*, 10(2), 221–242. <https://doi.org/10.1080/096959403000121298>.
- Brown, G. (1997). *Assessing student learning in higher education*. Routledge.
- Brown, G. T., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, 2(1), 22–30. <https://doi.org/10.14786/flr.v2i1.24>.
- Brown, K. W., West, A. M., Loverich, T. M., & Biegel, G. M. (2011). Assessing adolescent mindfulness: Validation of an adapted Mindfulness Attention Awareness Scale in adolescent normative and psychiatric populations. *Psychological Assessment*, 23(4), 1023–1033. <https://doi.org/10.1037/a0021338>.
- Buttner, E. H. (2004, 28). How do we dis students? A model of (dis)respectful business instructor behavior. *Journal of Management Education*, 319–334. <https://doi.org/10.1177/1052562903252656>.

- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233. <https://doi.org/10.1080/03075070600572132>.
- Chapelle, C. A. (2008). Utilizing technology in language assessment. In N. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 2263–2274). Springer, DOI: https://doi.org/10.1007/978-0-387-30424-3_172.
- Cheng, L. Y., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221–249. <https://doi.org/10.1177/0265532210384253>.
- Chory, R. (2007). Enhancing student perceptions of fairness: The relationship between instructor credibility and classroom justice. *Communication Education*, 56(1), 89–105. <https://doi.org/10.1080/03634520600994300>.
- Chory, R., Horan, S. M., & Houser, M. L. (2017). Justice in the higher education classroom: Students' perceptions of unfairness and responses to instructors. *Innovative Higher Education*, 42, 321–336. <https://doi.org/10.1007/s10755-017-9388-9>.
- Chory-Assad, R. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly*, 50(1), 58–77. <https://doi.org/10.1080/01463370209385646>.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyneff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge University Press.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3), 1–16. <https://doi.org/10.1186/s40468-020-00101-6>.
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101–116. <https://doi.org/10.1080/09695949993026>.
- Cresswell, J. W., & Poth, Ch. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed.). SAGE publication.
- Crosthwaite, P., Bailey, D., & Meeker, A. (2015). Assessing in-class participation for EFL: Considerations of effectiveness and fairness for different learning styles. *Language Testing in Asia*, 5(1), 1–19. <https://doi.org/10.1186/s40468-015-0017-1>.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/026532208090156>.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. In *Educational Assessment, Evaluation and Accountability*, (pp. 1–22). <https://doi.org/10.1007/s11092-015-9233-6>.
- Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *Turkish Online Journal of Educational Technology-TOJET*, 2(3), 13–19.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Durden, G. (2018). Accounting for the context in phenomenography-variation theory: Evidence of English graduates' conceptions of price. *International Journal of Educational Research*, 87, 12–21. <https://doi.org/10.1016/j.ijer.2017.11.005>.
- Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, 27(3), 243–256. <https://doi.org/10.1080/02602930220138606>.
- Eyal, L. (2012). Digital assessment literacy-the core role of the teacher in a digital environment. *Educational Technology & Society*, 15(2), 37–49. <http://www.jstor.org/stable/jeductechsoci.15.2.37>.
- Fan, X., Johnson, R., Liu, X., & Gao, R. (2020). College students' views of ethical issues in classroom assessment in Chinese higher education. *Studies in Higher Education*, DOI: <https://doi.org/10.1080/03075079.2020.1732908>, 1, 15
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(2), 1–14. <https://doi.org/10.1186/s40468-019-0078-7>.
- Fulcher, G. (2010). *Practical language testing*. Hodder Education.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- Gebril, A. (2017). Language teachers' conceptions of assessment: An Egyptian perspective. *Teacher Development*, 21(1), 81–100. <https://doi.org/10.1080/13664530.2016.1218364>.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>.
- Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development*, 23(2), 265–279. <https://doi.org/10.15446/profile.v23n2.90533>.
- Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching*, 58(3), 93–98. <https://doi.org/10.1080/87567550903418586>.
- Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23(7), 999–1011. <https://doi.org/10.1016/j.tate.2006.04.042>.
- Green, S. K., & Johnson, R. L. (2010). *Assessment is essential*. McGraw-Hill Higher Education.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Research*, 20(2), 1–7. DOI: <https://doi.org/10.3102/0013189X020005002> 1991, 5, 2
- Hamid, M. O., & Hoang, N. T. H. (2016). Humanizing language testing. *The Electronic Journal for English as a Second Language*, 22(1), 1–20.
- Hamp-Lyons, L., & Lynch, B. K. (1998). Perspectives on validity: A historical analysis of language testing conference abstracts. In A. J. Kennan (Ed.), *Validation in language assessment* (pp. 253–276). Lawrence Erlbaum Associates, Inc.
- Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101–111. <https://doi.org/10.1016/j.tate.2013.07.008>.
- Hayes, H., & Embretson, S. E. (2013). The impact of personality and test conditions on mathematical test performance. *Applied Measurement in Education*, 26(2), 77–88. <https://doi.org/10.1080/08957347.2013.765432>.
- Heidi, A., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Educational Administration & Policy Studies Faculty Scholarship*, 32(2), 159–181. https://scholarsarchive.library.albany.edu/eaps_fac_scholar/1.
- Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. <https://doi.org/10.1177/0265532211428317>.

- Holmgren, J., & Bolkan, S. (2014). Instructor responses to rhetorical dissent: Student perceptions of justice and classroom outcomes. *Communication Education*, 63(1), 17–40. <https://doi.org/10.1080/03634523.2013.833644>.
- Homayounzadeh, Z., & Razmjoo, S. A. (2021). Examining 'assessment literacy in practice' in an Iranian context: Does it differ for instructors and learners? *Journal of Teaching Language Skills*, 40(2), 1–45. <https://doi.org/10.22099/jtls.2021.40269.2978>.
- Horan, S. M., Chory, R., & Goodboy, A. (2010). Understanding students' classroom justice experiences and responses. *Communication Education*, 59(4), 453–474. <https://doi.org/10.1080/03634523.2010.487282>.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 328–402. <https://doi.org/10.1177/0265532208090158>.
- Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307. <https://doi.org/10.1177/0265532213480126>.
- Janatifar, M., & Marandi, S. S. (2018). Iranian EFL teachers' language assessment literacy under an assessing lens. *Applied Research on English Language*, 7(3), 361–382. <https://doi.org/10.22108/are.2018.108818.1221>.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. <https://doi.org/10.1016/j.system.2019.01.006>.
- Lankiewicz, H. (2014). Teacher interpersonal communication abilities in the classroom with regard to perceived classroom justice and teacher credibility. In M. Pawlak, J. Bielak, & A. Mystkowska-Wiertelak (Eds.), *Classroom-oriented research* (pp. 101–120). Springer International Publishing, DOI: https://doi.org/10.1007/978-3-319-00188-3_7.
- Lantolf, J. P., & Peohner M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 Education*. Routledge, DOI: <https://doi.org/10.4324/9780203813850>.
- Lincolin, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage, Naturalistic inquiry.
- Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263–275. <https://doi.org/10.1080/02602930701292548>.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. <https://doi.org/10.1177/0265532213480129>.
- Marton, F. (1986). Phenomenography: A research approach to investigating different understandings of reality. *Journal of Thought*, 21(3), 28–49.
- Marton, F., & Booth, S. (1997). *Learning and awareness*. Lawrence Erlbaum.
- McDonald, B. (2013). Mentoring and tutoring your students through self-assessment. *Innovations in Education and Teaching International*, 50(1), 62–71. <https://doi.org/10.1080/14703297.2012.746516>.
- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), 31–51. https://doi.org/10.1207/s15434311laq0301_3.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. <https://doi.org/10.3102/0013189X023002013>.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>.
- Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.
- Moon, T. R. (2016). Differentiated instruction and assessment: An approach to classroom assessment in conditions of student diversity. In G. T. L., Brown, & L. R. Harris, *Handbook of human and social conditions in assessment* (Ed.) (pp. 284–301). Routledge.
- Morrison, C. A., Ross, L. P., Sample, L., & Butler, A. (2014). Relationship between performance on the NBME comprehensive clinical science self-assessment and USMLE Step 2 Clinical Knowledge for USMGs and IMGs. *Teaching and Learning in Medicine*, 26(4), 373–378. <https://doi.org/10.1080/10401334.2014.945033>.
- Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Association for Supervision & Curriculum Development.
- Nicol, D. J., & Macfarlane, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*, (2nd ed.). Sage.
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92(3), 544–555. <https://doi.org/10.1037/0022-0663.92.3.544>.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33–40. <https://doi.org/10.1037/0022-0663.82.1.1714.1201>.
- Plake, B. S., Impara, J. C., & Fager, J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–39. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>.
- Rasian, Z. (2009). Higher education governance in developing countries, challenges, and recommendations: Iran as a case study. *Nonpartisan Education Review/Essays*, 5(3), 1–18.
- Rasooli, A., DeLuca, C., Rasegh, A., & Fathi, S. (2019). Students' critical incidents of fairness in classroom assessment: An empirical study. *Social Psychology of Education*, 22(3), 701–722. <https://doi.org/10.1007/s11218-019-09491-9>.
- Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation*, 56, 164–181. <https://doi.org/10.1016/j.stueduc.2017.12.008>.
- Rodabaugh, R. C. (1994). *College students' perceptions of unfairness in the classroom*. Retrieved from <http://digitalcommons.unl.edu/podimproveacad/319>.
- Rosman, T., Mayer, A. K., & Krampen, G. (2015). Combining self-assessments and achievement tests in information literacy assessment: Empirical results and recommendations for practice. *Assessment and Evaluation in Higher Education*, 40(5), 740–754. <https://doi.org/10.1080/02602938.2014.950554>.

- 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. <https://doi.org/10.1177/0265532213480128>.
- Shahzamani, M., & Tahiriari, M. H. (2021). Iranian medical ESP practitioners' reading comprehension assessment literacy: Perceptions and practices. *Iranian Journal of English for Academic Purposes*, 10(1), 1–15.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.1002/9780470690048.ch10>.
- Siegel, M. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching*, 44(6), 864–881. <https://doi.org/10.1002/tea.20176>.
- Siegel, M. (2014). Developing pre-service teachers' expertise in equitable assessment for English learners. *Journal of Science Teacher Education*, 25(3), 289–308. <https://doi.org/10.1007/s10972-013-9365-9>.
- Sivan, A. (2000). The implementation of peer assessment: An action research approach. *Assessment in Education: Principles, Policy & Practice*, 7(2), 193–213. <https://doi.org/10.1080/713613328>.
- Sjostrom, B., & Dahlgren, L. O. (2002). Nursing theory and concept development or analysis: Applying phenomenography in nursing research. *Journal of Advanced Nursing*, 40(3), 339–345. <https://doi.org/10.1046/j.1365-2648.2002.02375.x>.
- Smith, C. D., Warsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students' assessment literacy. *Assessment and Evaluation in Higher Education*, 38(1), 44–60. <https://doi.org/10.1080/02602938.2011.598636>.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5–15. <https://doi.org/10.1111/j.1745-3992.2001.tb00065.x>.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1–14. <https://doi.org/10.1186/s40468-019-0077-8>.
- Tapia, J. A., & Panadero, E. (2010). Effects of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje*, 33(3), 385–397. <https://doi.org/10.1174/021037010792215145>.
- Tata, J. (1999). Grade distributions, grading procedures, and students' evaluations of instructors: A justice perspective. *The Journal of Psychology*, 133(3), 263–271. <https://doi.org/10.1080/00223989909599739>.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>.
- Taylor, L. (2013). Communicating the theory practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>.
- Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125–144). SAGE, DOI: <https://doi.org/10.4135/9781452218649.n8>.
- Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning- A review of research studies. *Studies in Educational Evaluation*, 37, 25–34. <https://doi.org/10.1016/j.stueduc.2011.03.004>.
- Tomlinson, C. A., & Moon, T. R. (2013). Differentiation and classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 415–430). SAGE, DOI: <https://doi.org/10.4135/9781452218649.n23>.
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20–27. <https://doi.org/10.1080/00405840802575769>.
- Tsagari, D. (2009). *The complexity of test washback*. Peter Lang.
- Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389–404. <https://doi.org/10.1080/0141192960220402>.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>.
- Wang, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. *Journal of Computer Assisted Learning*, 22(3), 207–217. <https://doi.org/10.1111/j.1365-2729.2006.00166.x>.
- Watmani, R., Asadollahfam, H., & Behin, B. (2020). Demystifying language assessment literacy among high school teachers of English as a foreign language in Iran: Implications for teacher education reforms. *International Journal of Language Testing*, 10(2), 129–144.
- Winke, P., & Fei, F. (2008). Computer-assisted language assessment. In N. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 1442–1453). Springer, DOI: https://doi.org/10.1007/978-0-387-30424-3_110.
- Xu, Y., & Brown, G. T. (2016). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133–158.
- Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>.
- Zhou, J., Zhao, K., & Dawson, P. (2020). How first-year students perceive and experience assessment of academic literacies. *Assessment and Evaluation in Higher Education*, 45(2), 266–287. <https://doi.org/10.1080/02602938.2019.1637513>.
- Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–12). Routledge.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.