



Language Teaching Research Quarterly

2022, Vol. 29, 102–119



EFL Teachers' Assessment Literacy: A Contextualized Measure of Assessment Theories and Skills

Ebrahim Mohammadkhah¹, Gholam Reza Kiany^{2*}, Zia Tajeddin³, Parvaneh ShayesteFar⁴

^{1,2,3}Department of English Language Teaching, Tarbiat Modares University, Iran

⁴Department of English Language Teaching, Farhangian Teacher Education University, Tehran, Iran

Received 21 December 2021 Accepted 30 April 2022

Abstract

Teachers from different sectors of education have always been involved in the process of assessing learners' academic performance. In the language teaching discipline, this has highlighted the significance of language assessment literacy (LAL), reflecting its significance to language teachers across various settings. In the wake of recent theoretical burgeoning, development of a contextualized framework that proves helpful in conceptualization and measurement of English as a foreign language (EFL) teachers' LAL is required. To this aim, the present study gained insights from Fulcher's (2012) LAL framework into validation and examination of a LAL scale that could be considered for its applicability and usefulness as a LAL measure and help EFL teachers self-evaluate their LAL levels (in particular, their LAL theoretical and practical components). For this purpose, a mixed-methods approach was used to synthesize interview data from six national and international experts with questionnaire data obtained from Iranian EFL teachers (N=173) to gauge their assessment theories and skills and bear evidence in support of the reliability, validity and utilization of the measure put forth. The study rounds out with the significance of teachers' LAL in teachers' pedagogy, with finally presenting theoretical and practical implications of the obtained evidence for educational research and practices.

Keywords: *Assessment Literacy, Language Assessment Literacy, Teacher Professional Development, Pedagogical Content Knowledge, Teacher Self-assessment*

Introduction

Language assessment literacy (LAL), whose origin can be traced back to general education (Popham, 2010; Stiggins, 1991), refers to the knowledge, skills and principles that stakeholders involved in language assessment activities should possess in order to fulfill their educational responsibilities, especially to perform assessment tasks (Coombe et al., 2020; Inbar-Lourie, 2017; Latif, 2021; Rauf & McCallum, 2020; Tajeddin et al., 2022). Based on the presumption that nearly a half of the guidelines time is committed to different assessment-related assignments (Stiggins & Conklin, 1992), teachers' LAL can work as a measuring stick in the assessment of their proficient viability and quality (Garrett & Steinberg, 2015). There's presently a progressing discussion on what is implied by LAL and how best to assess it to meet the changing evaluation needs of the 21st century instructors, in both local and worldwide settings (Fulcher, 2012; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014). Several LAL literature reviews have been detailed within the accessible writing on LAL assessment. However, these surveys have been designed for different teachers with different LAL needs, contexts, purposes and activities. As such, attempts were made in this study to develop a LAL survey measure to be feasible and intelligible for local English language teachers in the Ministry of Education (ME) context in Iran.

In spite of the great significance of LAL for instructors, a thorough analysis of the viable and hypothetical perspectives on the LAL of instructors with assorted appraisal needs, such as EFL instructors, has not been developed as of today. Studies on evaluation of proficiency of EFL instructors present that the findings are to be relevantly applied in terms of their information of the appraisal hypothesis, and standards so as to aptly hone on classroom practices (e.g., Fulcher, 2012; Inbar-Lourie, 2013), yet, a great many of them do not gauge the appropriate level of proficiency in evaluating their learners. More rigorous studies are hence needed with regard to how to gauge teachers' LAL, especially with the peculiarities and particularities of global standards, more specifically in the EFL context of Iran where nearly all instructors are non-native speakers and their LAL needs have received scant attention. As Fulcher (2012) postulated that an extended definition of LAL ought to incorporate not only the principles and practices but also necessities of the appraisal settings, the present study, thus, endeavored both to develop and validate EFL teachers' LAL through a contextualized measure which could be utilized as a demonstrative and diagnostic instrument for surveying purposes in this regard. As the debate over the content of LAL and who counts as a stakeholder was rather academically unsettled, promising efforts were made to put forth a sketchy understanding of LAL and its components. This was specifically made after the American Federation for Teachers' delineation of 'Standards for Teacher Competencies in Educational Assessment of Students' in 1990, suggesting several domains of teacher assessment literacy including *choosing and developing assessment methods, administering, scoring, and interpreting assessment results, using assessment results for decision making and grading, communicating assessment results, and recognizing unethical, inappropriate assessment use*. Although this framework was not focally designed for LAL purposes, it has informed research and conceptualization of LAL over years (see Giraldo, 2018; Giraldo & Murcia, 2019; Inbar-Lourie, 2013). These domains were then propounded by many researchers (e.g., Brindley, 2001; Inbar-Lourie, 2013; Kremmel &

Harding, 2020; Malone, 2013; Taylor, 2013; Xu & Brown, 2016). In a more comprehensive view, Fulcher (2012) mentions assessment core domains in detail covering abilities and skills needed to “design, develop, maintain or evaluate large scale standardized and/or classroom based test, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice” (p. 125). Additionally, he emphasized the “ability to place knowledge, skills, processes, principles and concepts within historical, social, political and philosophical framework” (Fulcher, 2012, p. 125) as well as the knowledge of different assessment purposes which he summarized in these questions: “What is this test/assessment for?” and “What decisions are to be made on the basis of the evidence generated?” (Fulcher, 2017, p. 465). Notwithstanding the findings emerging in these areas over the recent years, significant gains are still missing with regard to local EFL contexts where there often exists a widening gap between the idealized, state-mandated assessment purposes and the realized practices of local LAL. This implies that assessment agents such as EFL teachers (whether pre- or in-service) may not possess an adequate knowledge base required for accomplishing various assessment-related tasks in the current changing times.

In general, despite its paramount importance for student learning, quality teaching, and professional responsibilities (Berry et al., 2019; Black & Wiliam, 1998; Fulcher, 2012, 2020; Scarino, 2013), LAL has not been fully defined as a vital component of EFL teacher knowledge, in particular, teacher pedagogical content knowledge (PCK). This issue would give LAL rather a spare-tool view for EFL teachers that may become prone to deviations over time. Since research on teachers’ LAL is still inchoate, there seems to be a meaningful gap between contemporary assessment theories that uphold assessment culture, principles, and skills, and their practical manifestations in the EFL field. In other words, there is the question of (a) what LAL conceptually entails and (b) whether such hypothesized dimensions hold true in second language (L2) assessment contexts, such as EFL contexts, where English language teachers are to update their assessment knowledge base to be accountable and responsive. This calls for exploring the LAL construct as well as measuring the extent EFL teachers possess it. Given this understanding, the present study aimed to develop a measure for exploring the main LAL dimensions in the Iranian EFL community, and the extent to which the dimensions are empirically valid. Since there are few scales developed for such a local context, the study can be of theoretical and practical significance to EFL teachers and other involved stakeholders in that it would add to local understandings of assessment in other EFL contexts and, as pointed out by Fulcher (2012), would help teachers perceive their LAL levels and their training needs in the rapidly changing assessment times.

Review of the Literature

Teachers’ Language Assessment Literacy (LAL)

Following expansion of assessment literacy to different fields and domains (e.g., technology literacy, math literacy, ...), the addition of LAL was to be expected and LAL came into its own (e.g., Scarino, 2013; Taylor, 2013; Vogt & Tsagari 2014). Interest in LAL can thus be traced back to recent decades when different definitions were given by researchers in the field. For

instance, Davies (2008) was among the first who defined LAL concepts and components in terms of knowledge of language and context, the principles of language assessment and measurement, and training in assessment skills. In O'Loughlin's (2013) perspective, the LAL concept is viewed as "a range of skills related to test production, test score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within society" (p. 363). In the same vein, in Taylor's (2009) account, the term language assessment literate is to be applied to the those who are equipped with a sufficient and adequate understanding of LAL principles, and the ability to realize these principles into assessment practice, choose or develop appropriate assessment tasks, gather and evaluate assessment data and make decisions about their students' status and performance outcomes.

Inbar-Lourie (2008) has, however, taken up a critical perspective in language assessment and described the concept as the "capacity to ask and answer critical questions about the purpose of assessment, about the fitness of the tool being used about testing conditions, and about what is going to happen on the basis of the results" (p. 389). Later, in a more thorough classification of LAL, she addressed the following items as basics of LAL (Inbar-Lourie, 2013, pp. 27-41):

- Understanding social dimension of language assessment, including social role of assessment, social responsibility of language tests, social and political forces and consequences of assessment means and tests,
 - Knowledge on how to develop, administer, analyze, and report assessment results to the intended stakeholders,
 - Understanding of large-scale test data,
 - Proficiency in *Language Classroom Assessment*,
 - Mastering language acquisition and learning theories and relating to them in the assessment process,
 - Matching assessment with *language* teaching approaches. Knowledge about current language teaching approaches and pedagogies,
 - Awareness of the dilemmas that underlie assessment: formative vs. summative; internal external; validity and reliability issues particularly with reference to authentic language use; and
 - Individualized nature of LAL, the product of the knowledge, experience, perceptions, and beliefs that language teachers bring to the teaching and assessment process (based on Scarino, 2013).

Likewise, Vogt and Tsagari (2014) defined LAL as "the ability to design, develop and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade and score assessments on the basis of theoretical knowledge" (p. 377). Concentrating on LAL components, Pill and Harding (2013) identified LAL as comprising competencies that teachers require in order to "understand, evaluate and create language tests and analyze test data" (p. 382). Core components of teacher LAL have been recently studied in a wider scope. For instance, Kremmel and Harding (2020) implemented an empirical analysis of Taylor's (2013) framework utilizing survey data (N=1086) of participants from various parts of the world with the aim of gauging their LAL needs. In so doing, they gained access to supporting

evidence in favor of Taylor's LAL model containing *knowledge of theories; principles and concepts; technical skills; language pedagogy; impacts and social values; scoring; and personal beliefs and attitudes*. Underpinning all of these LAL frameworks is, consequently, a collection of core components, merits and perceptions.

As of today, of the most consummate definitions of teacher assessment literacy has been put forth by Fulcher (2012) who scaled teacher LAL along three core components

- knowledge, skills and abilities (i.e., *practical knowledge*),
- processes, principles and concepts (i.e., *theoretical and procedural knowledge*), and
- origins, reasons and impacts (i.e., *historical, social, political and philosophical knowledge*).

Emphasizing that not all of these core components are essential for all stakeholders, Fulcher (2012) highlighted the significant role of teachers' practical knowledge in providing the foundation of their assessment literacy rather than a mere reliance on their theoretical and principled understandings. Such a broader and comprehensive model of LAL has widely been adopted by many researchers in different countries (Tzagari & Vogt 2017). There is evidence that these research attempts have focused on LAL through different projects and studies including surveys; item writings, validation and reliability interpretations; portfolio, peer- and self-assessment, and feedback provision; classroom tests, statistics, scoring and score interpretations; ethics and fairness in assessment; and stakeholders' needs profile analyses (see for instance, Inbar-Lourie, 2008; Pill & Harding, 2013; Vogt & Tzagari, 2014; Xu & Brown, 2016). This shows that research into teacher LAL has not been limited only to theoretical conceptualizations of assessment competencies or components but aimed at the realization and identification of these components in practice.

Evolved after Fulcher (2012, 2015) and Taylor's (2013) suggestions, an area which has, therefore, dominated LAL research, to date, is explaining how an assessment literate teacher might be or what s/he needs to possess to be an assessment literate. This resulted in more concentration on the development of questionnaires, surveys and tests over the recent years. Harding and Kremmel (2016), Kremmel and Harding (2020), and Xu and Brown (2016), for instance, have concentrated on the development and empirical examination of more comprehensive instruments. However, the question of assessment needs and levels of particular groups of language teachers, such as the EFL teachers and teacher trainers, are yet hypothetical. In addition, the urgent need for possessing such competencies and skills with regard to its appropriate and effective implementation renders it essential to develop LAL scales and better yet, upgrade the available measures to account for the peculiarities of the EFL assessment in the contexts of Iran.

Therefore, informed by the available theoretical and empirical evidence, in particular Fulcher's (2012) elaborate classification of LAL providing insights into the LAL needs of teachers in specific contexts, the current study endeavored to develop a working instrument capable of adequately measuring the EFL teachers' LAL teaching in Iran's ME context. It is debated that language teaching in an EFL context, like that of Iran, conjures up the contextual particularities that can enormously shape and re-shape the teachers' assessment knowledge and

practices, and by extension their literacy (Latif, 2021; Tajeddin et al., 2022). In addition to such a development, utility and appropriacy of this contextually informed instrument were given a more thorough examination against the data obtained from the EFL context. Given the study objectives, the following research question was explored:

What specific constructs underlie a contextualized LAL model hypothesized for the Iranian high school teachers?

Method

Sample

Notwithstanding the localized scope and nature of the study, we solicited for the views of some international scholars who had then conducted big assessment projects in different contexts, local or global. Therefore, to obtain more comprehensive views and perspectives, both national and international English language teaching (ELT) scholars from different universities (Cambridge University, the Centre for English Language Communication of the National University of Singapore, Tarbiat Modares University, Kharazmi University, and Farhangian Teacher Education University, Iran) as well as 22 provincial Head Teachers (14 females, 8 males) in Iran were invited to participate in the study. Besides their qualifications and voluntary consent, they were also selected based on their specializations in different language assessment areas ranging from washback studies and test design, assessment and teacher education, linguistics, and virtual assessment skills, with a range of teaching experience from 3 to 24 years.

Regarding the scale validation, convenience sampling procedures were utilized in order to collect the data from English language teachers, via on-line distributions and in-person. A total of 110 paper surveys were first distributed among teachers (107 returned). This was, subsequently, followed by online forms (due to the COVID-19 pandemic condition) sent to more than 100 teachers employed in educational sectors (e.g., ME state-run and private schools) of different geographical regions, including Tehran, Azerbaijan, Isfahan, Shiraz, and Alborz. The final number of respondents (N=173) consisted of 125 (72%) female and 48 (28%) male teachers, with the age range of 22-55. Regarding the level of education, 64 (37%) had a B.A. degree, 76 (44%) held an M.A., and 33 (19%) teachers were Ph.D. holders. As to assessment training courses or program, 45 (26.5%) teachers reported receiving 'no training', 70 (41%) teachers with 'pre-service training', 32 (19%) with 'in-service training', and 26 (15%) teachers admitted receiving 'pre- and in-service training'. With regard to their fields of expertise, 12 (7%) had their specialization in English Literature, 6 (4%) in Linguistics, 10 (6%) in Translation, and 145 (83%) in English language teaching.

Design

This study adopted a mixed-methods design for the development of a self-assessment scale intended to measure Iranian EFL teachers' LAL levels. To this end, the data were first collected from a series of qualitative interviews. The constructed scale was then subjected to exploratory and confirmatory factor analyses to quantitatively examine its factor structures and the subcomponents. Main phases of the study are shown in Table 1.

Table 1

Design and Stages of the Study

Phase	Data collection	Data analysis
Qualitative phase	Semi-structured interview with ELT professors and head-teachers (22)	Thematic analysis
Quantitative phase	Questionnaire administration (172 teachers)	Exploratory and confirmatory factor analyses

Instruments and Procedures

The instruments used in this study were as follows:

ELT Professors and Head Teachers' Interview Guides

The aim of the interview was to delve more into the concept of English language assessment and, further, identify its components. The interview guide included seven main questions soliciting for interviewees' perceptions and views on teacher LAL, criteria for assessing it, LAL belief/theory-performance route, ways to improve it, convergence between higher-order policies on language assessment and teachers' professionalization in LAL, LAL-based qualifications, and factors affecting implementation of LAL in different assessment contexts.

Prior to the interviews, one language test designer and one teacher reviewed the content and face validity of the questions. Semi-structured individual interviews were conducted, each lasting 15 - 45 minutes, and analyzed via thematic analysis. Tape recordings and simultaneous note-taking were kept during the interviews. Four university professors preferred to send their answers in written form, however. Despite the difference in the format of the participants' responses, we conducted detailed inductive analyses of the data in several stages to compensate for the possible impact of the response mode on the theme extraction and development process. The interviews were conducted in English as the participants preferred. The participants were assigned pseudonyms to maintain their anonymity.

Language Assessment Literacy (LAL) Scale

To develop the self-reporting LAL instrument, a comprehensive review of the relevant literature was made to gain a better picture of LAL models such as Fulcher (2012), Xu and Brown (2016) and Taylor's (2013) models. More than 80 categories and behaviors (e.g., *teachers' scoring closed-response questions, teachers' item writing, etc.*) were collected. This initial source of data was then combined with the recurring categories emerged from the interview analyses, resulting in the following commonalities and overarching components that finally comprised the scale:

- (1) *Assessment in language pedagogy* (i.e., as a constituent of teachers' PCK competency) (14 items),
- (2) *Assessment purposes and principles* (6 items),
- (3) *Technical skills* (13 items), and
- (4) *Scoring and interpretations* (6 items).

Six nationally-recognized language assessment experts and two teachers checked and confirmed the content and face validity of the 39-item Likert-type instrument (see Appendix for the final version of the survey instrument). Participants were then asked to select out of the four choices provided for each of the items. Take, items 37 and 18, respectively, for example, asking ‘how skilled/competent you are in *knowing how to statistically evaluate language assessments regarding their reliability or validity?* And ‘*developing quality items and tasks for particular assessment purposes?*’

Results

Interviews: ELT Professors and Head Teachers’ Views

The interview data were iteratively scrutinized (by listening) and carefully transcribed. An ‘inductive analysis procedure’ was followed for data analysis. According to this procedure, the transcribed data were first broken down into separate chunks, examined with regard to their underlying themes, and finally categorized into several categories. Specifically, the theme development process includes main stages of (1) Initialization (transcribing, coding, identifying meaning units, abstracting and reflecting), (2) Construction (classifying, comparing, labeling, translating and transliterating extracted codes), (3) Rectifying (relating themes to established knowledge), and (4) Finalizing (finalizing relevant themes).

As to the units underlying the LAL, 190 themes were identified and subsequently subsumed under 7 categories. Results showed that “knowledge of peer and self-assessment” was the most frequently-preferred category among the responses with 29 hits.

Table 2

Frequency of the Categories of Language Assessment Literacy

Themes	Total Frequency
Disciplinary knowledge and PCK	27
Purpose, Method, and Content of assessment	28
Knowledge of grading	25
Knowledge of feedback	28
Knowledge of peer and self-assessment	29
Knowledge of interpreting and communicating assessment results	28
Ethics of assessment	25
Total	190

To shed more light on the minutiae of the interviewees, a more elaborate understanding of their perceptions and beliefs is provided. Almost all respondents believed that ‘*Disciplinary knowledge and pedagogical content knowledge*’ (with a frequency of 27) is the sieve through which other components get refined and fine-tuned. Of note, Peyman (university professor 1) mentioned the following:

...if a teacher, whether a language teacher at high school or university, has a very shallow *knowledge of principles of language learning and teaching*, and if they do not know what the status quo bears for them, I think the rest of the criteria would end up in failure". (Peyman)

Of interest is the fact that almost all the interviewees made it clear that *'Ethics in Assessment'* is important, yet there was a common thread running through all their comments: the fact that ethics has always been too soft an issue and rather subjective. Rather a dissatisfaction was reported by the interviewees' views as follows

...I believe fairness in assessment will be recognized as an important category in assessment because we are constantly making decisions, so it is a must for us to be as fair as possible. Nevertheless, I am not sure where the boundaries are and how not to trespass this territory. (Kiarash)

Yet in another part of the interview, one of the head teachers made complaints regarding lack of proper training on how to select the right method of assessment relevant to the category of *'knowledge of purpose, method, and content of assessment'*. She was rather ambivalent about whom to blame, whether training programs at universities or in-service education programs. She said that

... English teachers are *not sure what to test, and how to test it* if they aim for any material only interior to the classroom. (Soheila)

The LAL Surveys: the EFL Teachers' Perspectives

The final version of the LAL scale was administered to 210 English language teachers; 173 were returned back (a response rate of 82%). Because the scale items had been written to relate to four components resulting from the outcome of content and thematic analyses conducted in the qualitative phase, confirmatory factor analyses (CFA) with AMOS (version 18; Arbuckle, 2009) were performed to corroborate the hypothesized factor structure of the scale. The results of the structural measurement of the scale provided empirical support for the existence of four subscales. Of 39 items, 13 revealed very weak loadings on their factors. Examining the normed Chi-square (CMIN/DF) and other Goodness-of-Fit Indices (GFIs) showed a significantly fit model with CMIN/DF=1.94 (acceptable range of >1 & <3), Comparative Fit Index (CFI)=.90, Goodness-of-Fit Index (GFI)=.89, and RMEAS=.06 ($<.08$). As a result, 'Language Assessment Literacy Scale' stands corroborated with a rather good-fitting model.

Outliers, data points that exerted excessive leverage, and normality were also checked throughout the data (Mardia's statistics). Also, no skewness and kurtosis were found, with the results ranging from -.671 to .024, and -.527 to 1.41 for the variables, respectively.

Table 3 depicts the classification of measurement model parameters obtained through CFA-AMOS.

Table 3

Parameter Estimates for the Measurement Model of the LAL Scale

			Estimate	S.E	C.R	P	Loadings (CFA estimates)
Q3	<---	LA_inPedag	1.000				
Q6	<---	LA_inPedag	.867	.127	6.833	***	par_1
Q7	<---	LA_inPedag	.959	.136	7.042	***	par_2
Q28	<---	LA_inPedag	1.129	.149	7.594	***	par_3
Q29	<---	LA_inPedag	1.208	.149	8.114	***	par_4
Q30	<---	LA_inPedag	1.259	.150	8.379	***	par_5
Q31	<---	LA_inPedag	1.207	.150	8.049	***	par_6
Q37	<---	LA_inPedag	.960	.142	6.762	***	par_7
Q40	<---	ATheory._P.P	1.000				
Q41	<---	ATheory._P.P	.980	.126	7.802	***	par_8
Q42	<---	ATheory._P.P	.949	.129	7.335	***	par_9
Q43	<---	ATheory._P.P	.987	.128	7.686	***	par_10
Q44	<---	ATheory._P.P	.842	.130	6.485	***	par_11
Q12	<---	Technical_Skills	1.000				
Q13	<---	Technical_Skills	1.013	.124	8.196	***	par_12
Q15	<---	Technical_Skills	1.052	.138	7.614	***	par_13
Q19	<---	Technical_Skills	1.026	.128	8.007	***	par_14
Q20	<---	Technical_Skills	1.044	.127	8.228	***	par_15
Q21	<---	Technical_Skills	1.035	.131	7.925	***	par_16
Q22	<---	Technical_Skills	1.175	.139	8.476	***	par_17
Q23	<---	Technical_Skills	.975	.134	7.248	***	par_18
Q18	<---	Technical_Skills	1.142	.130	8.796	***	par_19
Q32	<---	Scoring_Interpret.	1.000				
Q33	<---	Scoring_Interpret.	1.197	.121	9.861	***	par_20
Q34	<---	Scoring_Interpret.	1.111	.118	9.440	***	par_21
Q35	<---	Scoring_Interpret.	.990	.122	8.081	***	par_22

Note: LAinPedag=language assessment in pedagogy; Atheory P.P=assessment theory, purposes and principles; Scoring_Interpret=scoring and interpretations; ***=p-value<.01.

The loading analysis of the items and p- values of the modified model could support a four-component model of LAL. Of the remaining 26 items, no item violated the adequacy criteria at the CFA stage (CR>1.96, p-value<.05, error variance, or SE≤1.0); subsequently, it did not stand to reason to discard the 26 items mainly owing to their revelatory loading (p<.05). As such, Table 4 provides insights into the factors represented in the final format of the survey.

Table 4*Main Factors of the LAL Scale*

Item Number	Main Factors			
	1	2	3	4
Q3	.62			
Q6	.59			
Q7	.63			
Q28	.69			
Q29	.75			
Q30	.78			
Q31	.75			
Q37	.60			
Q40		.65		
Q41		.68		
Q42		.66		
Q43		.70		
Q44		.56		
Q12			.67	
Q13			.69	
Q15			.64	
Q19			.68	
Q20			.69	
Q21			.67	
Q22			.73	
Q23			.61	
Q18			.75	
Q32				.68
Q33				.86
Q34				.79
Q35				.68

As shown in Table 4, the loadings of all 26 items are above .60, implying that these items could efficiently measure the teachers' LAL. The highest loadings were reported to be .86, .79 and .78 for item 33 (*scoring and interpreting factor*: I am skilled/competent in communicating assessment purposes and results to students), item 34 (*scoring and interpreting factor*: I am skilled/competent in communicating assessment purposes and results to school/institute administrative staff), and item 30 (*assessment in language pedagogy factor*: I am skilled/competent in training my students how to peer-asses each other's performance). The

results also indicated that the lowest loadings were on item 44 (*assessment principle of fairness: I am skilled/competent in accommodating candidates with disabilities or other learning impairments*) and item 6 (*assessment in language pedagogy factor: I am skilled/competent in knowing how assessments can influence the design of a language course or curriculum*).

The internal consistency of the total scale was .96 for the scale in its totality, and .86, .78, .88 and .83 for Assessment in language pedagogy ($\bar{x}=2.84$), Assessment purposes and principles ($\bar{x}=2.71$), Technical skills ($\bar{x}=2.80$) and Scoring and interpretation ($\bar{x}=2.81$), respectively, showing satisfactory reliability indices.

Discussion

The present study used a mixed-methods approach to provide a clear picture of a survey intended to measure the Iranian EFL teachers' literacy of assessment. A synopsis of available literature on LAL models and interview themes proposed a LAL profile that resulted in the development of a self-assessment survey. The survey instrument asked English language teachers to self-appraise their LAL levels and the extent they use such knowledge as a base for performing their assessment-related responsibilities and tasks.

The developed scale was carefully examined for its factor structure through CFA processes empirically narrowing all the items down to four distinct factors: (a) assessment in language pedagogy; (b) assessment theories and principles; (c) scoring and interpretations; and (d) technical skills. The results are consistent with Fulcher's (2012) 'assessment knowledge, skills and abilities' and 'assessment principles, concepts and processes'. The results are also in line with Taylor's (2013) 'language pedagogy', 'knowledge of theory', 'technical skills', 'principles and concepts', and 'scores and decision making' components. The present findings, furthermore, corroborate the main components of the teachers' assessment knowledge base hypothesized by Xu and Brown's (2016) model, especially 'disciplinary knowledge and PCK', 'knowledge of assessment purposes, content and methods', 'knowledge of grading', and 'knowledge of assessment interpretation and communication while considering ethics', although with reduction in the number of the components. The factor structure of the current LAL scale was corroborated by the present data set, indicating the applicability of the current tool for appraising and diagnosing the EFL teachers' assessment knowledge base.

Data analyses also revealed different loadings of the factors. Regarding the highly loaded items, the teachers evaluated themselves as competent in communicating the results with students and institutional administrative staff, and having the students engage in peer assessment. This finding attests to the significance of teachers' verbal and nonverbal literacy abilities to negotiate test results, as Popham (2010) and Gilardo (2021) argued. Moreover, because in the Iranian public sector (i.e., ME) assessment is a clearly-defined path for student advancement, it seems that the teachers have developed the literacy (particularly principles) required for professionalism in delivering assessment. An immediately relevant point here is that the teachers have developed knowledge of macro-structural principles of delivering assessment as well as micro-structural particularities of enacting assessment at the classroom level. Scarino (2013) and Inbar-Lourie (2017), among others, have emphasized that developing such micro- and macro-

level literacy is essential for teachers in order to become fully aware of contextual idiosyncrasies of effective assessment, which seems to be the case with our EFL teachers.

On the other hand, the low-loaded factors pertained to course design and fairness items. In her discussion of LAL, Inbar-Lourie (2013) argued that language teachers' literacy demands developing a vernacular specific to language teachers in accounting for both in-class and higher-order issues. As course design and fairness are higher-order issues that require developing a knowledge base involving theoretical advocacy, it seems that the teachers have rarely encountered situations that raise issues of course design (which is done usually by higher-order decision makers in the Iranian context) and fairness. Relatedly, contribution of teacher assessment to course/curriculum design is not the focal topic in the Iranian educational system, and more noticeably in tertiary education. Given the significance of socio-moral issues in language assessment (see Chan & Luo, 2020; Giraldo, 2021), these findings call for incorporating into pre-service and in-service teacher education program modules related to teachers' assessment literacy with a specific focus on curriculum-oriented assessment.

In its generality, the present study corroborates the LAL themes and item. For instance, the particular items of 'strengthening disciplinary competence', 'the ability to introduce peer- and self-assessment to students', 'the skill of conducting formative assessment', as well as 'the ability to communicate the assessment results under the principles of ethics and fairness' go in lockstep with the similar items put forth by Kremmel and Harding's (2020) and Bøhn and Tsagari's (2021) studies. The findings support the conception that 'teachers' LAL is part of their PCK' in that teachers need to know *how* to assess the *what/content* of school curriculum.

More specifically, Fulcher's supposition that language teachers should be much aware of assessment principles and practices was empirically supported. Nonetheless, a slight difference was observed in political and historical items that were avoided by some respondents, though social aspects such as equity and fairness in assessment were considered by them. Our survey shows that our EFL teachers and head teachers are not only aware of a need for an adequate LAL repertoire but also a need for improving and updating it. A synthesis of the survey results with qualitative outcomes is indicative of our participants' understanding and clear conception of the undeniable role of language assessment and testing in society. This local EFL community showed their needs to acquire more of the 'what' to design as well as 'when' and 'how' to design, deliver and score a test in its social context. This implies a change in the community's understanding of a more balanced LAL which, as strongly argued by Fulcher (2012), integrates both procedural and practical approaches to language assessment within a much wider social context.

When aspects of teacher practical, procedural and theoretical assessment knowledge were explored from the lens of the new conceptions of language assessment which call for a variety of assessment skills and techniques as well as an updated knowledge base, helping teachers acquire LAL in order to implement quality assessments in their teaching context becomes a prominent issue. However, a prerequisite for this is the teachers' awareness of their own assessment literacy levels. The LAL instrument of the present study, with its unique constructs, helps refine both its

conceptualization and consideration as a contextually situated knowledge base in need of improvement and further use and at the same time, a solid base to inform assessment practices.

Conclusions

In the learning-oriented assessment view, ascertaining and enhancing learners' learning process requires teachers to "have a range of required skills and competencies at their disposal" (Fulcher, 2012, p.124) and be able to employ them appropriately in various assessment practices. In practical terms, teachers are reported not to be able to effectively and appropriately design, develop or evaluate large-scale or classroom tests without adequate possession of assessment abilities and skills, and knowledge of assessment processes, principles and concepts. This clearly implies that teachers' inadequate assessment literacy undermines their ability to assess and enhance students' learning. In this vein, language teachers, novice or experienced ones need not only to be aware of their language assessment practices, procedures and performance while teaching but also to reflect on them pedagogically. Therefore, improving teachers' LAL as part of their professional development empowers them to handle their assessment tasks and enhances their teaching and students' learning. This study aimed to provide a clear picture of what components a contextualized LAL model of the Iranian EFL teachers comprises. To this aim, a LAL scale was developed and empirically validated through participants' response data indicating the component structure of the EFL teachers' LAL and the feasibility of the instrument. The results of the study indicate that the present contextualized LAL scale comprises four distinct components: Assessment technical skills, Assessment purposes and principles, Assessment in language pedagogy, and Scoring and interpretations of assessment results. These findings, though with some distinctions, are much in line with Fulcher (2012) and Taylor (2013) such that teachers' language assessment conceptions, knowledge and beliefs have impacts on their assessment-related practices and performance, and that the teachers need to be professionally trained on LAL due to the changing nature of assessment in the rapidly changing world.

The findings, overall, have several implications for the field. First, the current instrument could be used as a contextually-informed scale for diagnostic purposes, for instance, for getting informed about our EFL teachers' LAL levels, the poor areas they identify for themselves, and the needs they perceive they have. Second, the obtained information can also inform appropriate teacher education programs or any professional development workshop in the LAL area. Accordingly, teacher trainers and decision-makers can plan licensure or accreditation LAL training programs and follow-up interventions in pre- and in-service contexts to help the EFL teachers achieve the LAL core competencies and appropriately implement them in practice. Given the dynamic nature of LAL, not only the teachers but also teacher trainers need to constantly update their LAL with the latest research-based insights. Thus, "training and support to engage in assessment and to adapt to learner-centered activities are considered critical" (Fulcher, 2020, p. 44).

The study has limitations that offer a rich basis for future research. The first one concerns the nature of our sample. Due to the social distance limitations during the Covid-19 (at the time of

this study) and closures of many schools and language institutes, it was only possible for us to gather data only from 173 teachers. This affords researchers opportunities for future research using a more representative sample. To develop the scale in more productive ways, there is, therefore, a need for a more sample of teachers from different the EFL contexts of the country (i.e., few contexts holding 'assessment culture' vs. many contexts still with 'examination/test-based culture', especially public schools; ShayesteFar, 2020). Second, as has been argued in the literature of LAL (e.g., Giraldo, 2018; Inbar-Lourie, 2013; Kremmel & Harding, 2020), the voice of various stakeholders should be considered in designing LAL scales. We think that including these stakeholders (such as parents and students) would produce more comprehensive scales. Third, developing initial categories using the experience of more scholar and researchers is more likely to give a more inclusive range of items needed to be included in a LAL scale, which could be done by future research.

References

- Arbuckle, J. L. (2009). *IBM SPSS Amos 18 user's guide*. Chicago, IL: IBM.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113-123. <https://doi.org/10.1093/elt/ccy055>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi: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. <http://dx.doi.org/10.17507/jltr.1202.02>
- Brindley, G. (2001). Outcomes-based assessment in practice: some examples and emerging insights. *Language Testing*, 18(4), 393-407. <https://doi.org/10.1177/026553220101800405>
- Chan, C.K.Y., & Luo, J. (2020). An exploratory study on teacher assessment literacy: Do novice university teachers know how to assess students' written reflection? *Teachers and Teaching*. [10.1080/13540602.2020.1787375](https://doi.org/10.1080/13540602.2020.1787375)
- 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), 2-16. <https://doi.org/10.1186/s40468-020-00101-6>.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.
- 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. (2015). Assessing second language speaking. *Language Teaching*, 48, 198-216. <https://doi.org/10.1017/S0261444814000391>
- Fulcher, G. (2017). The practice of language assessment. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 463-475), New York, NY: Routledge.
- Fulcher, G. (2020). Operationalizing language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8-28). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Garrett, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, 37(2), 224-242. <https://doi.org/10.3102/0162373714537551>
- 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/profle.v20n1.62089>
- Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), 243-259. <https://doi.org/10.14483/22487085.14514>
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413-428). Berlin, Germany: De Gruyter
- Inbar-Lourie, O. (2013). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. <http://dx.doi.org/10.1177/0265532208090158>
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, L. G. Or, & S. May (Eds.), *Language testing and assessment* (pp.275-270). New York, NY: Springer.

- 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>
- Latif, M. W. (2021). Exploring tertiary the EFL practitioners' knowledge base component of assessment literacy: implications for teacher professional development. *Language Testing in Asia*, 11(1), 1-22. <https://doi.org/10.1186/s40468-021-00130-9>
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344. <https://journals.sagepub.com/doi/10.1177/0265532213480129>
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380. <https://doi.org/10.1177%2F0265532213480336>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary enquiry. *Language Testing*, 30(3), 381-402.
- Popham, W. J. (2010). *Everything school leaders need to know about assessment*. Thousand Oaks, CA: Corwin
- Rauf, M., & McCallum, L. (2020). Language assessment literacy: Task analysis in Saudi universities. In L. McCallum & C. Coombe (Eds.), *The assessment of L2 written English across the MENA region* (pp. 13-41): Palgrave Macmillan.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment andin teacher learning. *Language Testing*, 30(3), 309-327. <https://doi.org/10.1177/0265532213480128>
- ShayesteFar, P. (2020). A model of interplay between student English achievement and the joint affective factors in a high-stakes test change context: Model construction and validity. *Educational Assessment, Evaluation and Accountability*, 23(3), 335-371. <https://doi.org/10.1007/s11092-020-09326-8>
- Stiggins, R. (1991). Assessment literacy. *The Phi Delta Kappan*, 72, 534-539. <https://www.jstor.org/stable/i20404445>
- Stiggins, R. J. & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.
- Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of the EFL teachers: Scale construction and validation. *Language Testing*, X, 1-30. <https://doi.org/10.1177/02655322211057040>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36
- 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>
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6, 41-63.
- Vogt, K. & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. DOI: 10.1080/15434303.2014.960046.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>

Appendix: Teacher Language Assessment Literacy (LAL) Scale

(Find the complete version in the link:

https://docs.google.com/forms/d/e/1FAIpQLSfEo5oWy0RTKK_nYrJSX6CBI0ZWBiBOYEMRqoGRcCeAH4g/viewform?usp=sf_link

Dear colleague,

Please specify the extent you perceive you are skilled/competent in the assessment domains described by the following statements.

(Please choose one of the *Not skilled at all=1; slightly skilled=2; skilled=3; highly skilled=4*)

Domains <i>How skilled/competent you are in the following domains?</i>	LAL Levels			
	Not skilled at all 1	Slightly Skilled 2	Skilled 4	Highly skilled 5

Item 3. knowing how major language learning theories and trends (e.g., Cognitivism, Socio-cultural theories, CLA, ...) affect selection of assessments trends (e.g., Integrative, and Functional-pragmatic trends).				
Item 6 knowing how assessments can influence the design of a language course or curriculum.				
Item 7. knowing the consequences of assessments on classroom teaching, learning and materials.				
Item 12. selecting appropriate rating scales depending on the purpose of my assessment.				
Item 13. selecting appropriate items or tasks for a particular assessment purpose.				
Item 15. training others to write good quality items (questions) or tasks for language assessment.				
Item 18. developing quality items and tasks for particular assessment purposes.				
Item 19. using rating scales to score oral skills (e.g., speaking).				
Item 20. using rating scales to score written skills (e.g., writing).				
Item 21. scoring closed-response questions (e.g., Multiple-Choice Questions).				
Item 22. designing rubrics and keys for scoring assessment tasks.				
Item 23. scoring open-ended questions (e.g., short answer questions).				
Item 28. applying technology or digital assessment tools (e.g., mobile apps, web-based- platforms) to give feedback to students.				
Item 29. training my students how to self-assess their performance.				
Item 30. training my students how to peer-assess each other's' performance.				
Item 31. engaging my students in self-monitoring their learning by using assessment information.				
Item 32. communicating assessment purposes and results to students' parents.				
Item 33. communicating assessment purposes and results to students.				
Item 34. communicating assessment purposes and results to school/institute administrative staff.				
Item 35. understanding concepts of technical reports and terminologies (e.g., reliability index, mean, measurement error, standard variation, etc.) of assessments.				
Item 37. knowing how to statistically evaluate language assessments regarding their reliability or validity.				

Item 40. identifying if an item on a test offends or unfairly penalizes students simply because of race, gender, religion or socioeconomic status (i.e., assessment bias).				
Item 41. maintaining equal assessment protocols for all the students involved (e.g., equal time, allocation, ...).				
Item 42. providing individualized learning opportunities to meets students' different needs.				
Item 43. knowing if assessment outcomes are used inappropriately (e.g., unfair pass/fails and selection decisions).				
Item 44. accommodating candidates with disabilities or other learning impairments.				

Thanks a lot for your participation

Acknowledgements

Not applicable.

Funding

Not applicable.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

Rights and Permissions

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. You may view a copy of Creative Commons Attribution 4.0 International License here: <http://creativecommons.org/licenses/by/4.0/>.