EFL Teachers Assessment Literacy Needs Inventory: A Case of Fulcher's Assessment Literacy Framework

Mohammad Reza Khodashenas iD

PhD Candidate of TEFL, Islamic Azad University,
Torbat-e Heydarieh Branch, Torbat-e Heydarieh, Iran

Hossein Khodabakhshzadeh* iD

Assistant Professor of TEFL, Islamic Azad University,
Torbat-e Heydarieh Branch, Torbat-e Heydarieh, Iran

Purya Baghaei iD

Assistant Professor of TEFL, Islamic Azad University,
Mashhad Branch, Mashhad, Iran

Khalil Motallebzadeh iD

Associate Professor of TEFL, Islamic Azad University,
Torbat-e Heydarieh Branch, Torbat-e Heydarieh, Iran

Abstract

Language assessment literacy has been addressed in a wealth of research. Although many studies have attempted to measure teachers' assessment literacy, there is still a gap that prompted us to investigate the area from the EFL teachers' assessment literacy needs perspective. To accomplish the purpose, in line with the changes in classroom assessment over the past decades, this study was an attempt to develop and validate an inventory on Teachers Assessment Literacy Needs (TALNs). As the first stage, a set of items was generated through an extensive review of the relevant studies. In the quantitative phase, the developed inventory was administered to 159 English as a foreign language teachers selected through convenience sampling. An inventory construction and validation framework, consisting of exploratory analyses, was used to examine the construct validity of the proposed inventory. The results indicated that the inventory can be best explained by four components, which are *knowledge of language assessment literacy, consequences of language assessment literacy, processes of language assessment literacy, and teachers' expectations of Continuing Professional Development (CPD) programs*. The TALNs inventory developed in this study aimed to help practitioners and researchers to investigate teachers' needs in assessment literacy. Fulcher's (2012) assessment literacy framework was drawn on as the analytic model guiding the study.

*Keywords***:** Assessment literacy, Needs analysis, Fulcher's assessment literacy framework

*Corresponding author's email: hkhodabakhshzade@gmail.com

INTRODUCTION

Accountability and assessment literacy are now served as rudimental features for all teachers and play vital roles in teacher education programs (Xu & Brown, 2016). According to Scarino (2013), the effectiveness of a language program is highly dependent on deep understanding, clear awareness, and careful implementation of assessment techniques. The implementation of diverse assessment strategies to evaluate and enhance student performance has been a focus of attention in the field of English Language Teaching. Considering the concept of education, assessment is highly advantageous. Not only does it reflect teachers' success in teaching but also it shows learners' progress and improvement in the classroom setting. Moreover, according to Öz and Atay (2017), assessment helps teachers to "recognize what is wrong, what is right, and what parts need to be changed, improved, or omitted" (p.26). Integrating assessment and instruction to supporting, monitoring, and reporting students' learning and demonstrating educational standards is recommended to all teachers throughout the world (DeLuca, LaPointe-McEwan, & Luhanga, 2016; Gotch & French, 2014). In an effort to enhance teachers' classroom assessment methods, many researchers (e.g., DeLuca, LaPointe-McEwan, & Luhanga, 2016; Gotch & French, 2014; Mertler & Campbell, 2005) developed different tools and instruments to investigate and monitor teachers' assessment literacy. In their systematic review of assessment literacy measures, Gotch and French (2014) discovered that there is little psychometric data to support these measures, and that existing instruments lack representativeness and relevance of content in light of developments in the assessment area. These findings are not surprising considering that most of assessment literacy instruments are based on early 1990s assessment standards (i.e., Standards for Teacher Competency in Educational Assessment of Students [STCEAS], American Federation of Teachers [AFT], National Council on Measurement in Education [NCME], and National Education Association [NEA], 1990; Gotch & French, 2014;

DeLuca, LaPointe-McEwan, & Luhanga, 2016).

Using the 1990s standards as a guideline, instruments such as the Teachers Assessment Literacy Questionnaire (TALQ) (Plake, Impara, & Fager,1993) and Classroom Assessment Literacy Inventory (CALI) (Mertler, 2003) were developed to investigate teachers' assessment literacy levels. These assessment literacy instruments were designed to evaluate teachers' knowledge of the language skills and to highlight their strengths and weaknesses in assessment literacy. Although the strong and weak areas identified in these studies differed by various samples, the overall agreement was that teacher assessment knowledge was often inadequate in comparison to standards and expectations. Brookhart (2011) identified that the 1990s assessment standards no longer account properly for the diversity of assessment activities or the assessment expertise required by teachers in current educational landscape and their assessment needs.

LITERATURE REVIEW

Previous Instruments of Teachers' Assessment Literacy

Several studies (e.g., DeLuca, LaPointe-McEwan, & Luhanga, 2016; Mertler & Campbell, 2005; Gotch & French, 2014; Plake, et al.,1993) have been conducted to investigate teachers and students' perceptions of assessment literacy. Most of these studies were quantitative and based on the original 1990s Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME, & NEA, 1990).

In their study, Mertler and Campbell (2005) established an Assessment Literacy Inventory (ALI) aiming at helping teachers and school administrators establish a reliable and valid process of grading students. Moreover, they intended to help teachers take the advantage of this inventory in terms of professional development and in-class assessment.

Considering the concept of language assessment literacy from the viewpoint of teachers, Rezaeifard and Tabatabaei (2018) investigated 52 Iranian EFL teachers' perceptions of assessment literacy. They used

Mertler's (2003) Classroom Assessment Literacy Inventory (CALI) as the main instrument of their study. After analyzing the data, they showed that majority of the participants were at the low level of assessment literacy. Furthermore, in their mixed methods study on 16 in-service and preservice Iranian EFL teachers, Dehqan and Asadian Sorkhi (2020) revealed that years ofɇteaching experience played a vital role in demonstrating teachers' knowledge of assessment. They argued that in-service teachers were more literate in assessment compared to pre-service ones. Moreover, they asserted that teachers were not interested in implementing assessment literacy skills in their classes. Therefore, they suggested that both practical and theoretical concepts of assessment literacy be incorporated in teachers' education programs.

Considering students' perceptions of assessment literacy in the foreign language context of Iran, Brown, Pishghadam and Sadafian (2014) used the Students' Conceptions of Assessment (SCoA) inventory as the instrument of their study for examining their conceptions of assessment. Their findings showed that all the 760 Iranian university students who participated in their study had both positive and negative attitudes toward assessment. In other words, they claimed that although assessment might improve both learning and teaching, it might hinder learning development. This conclusion is partially supported by Tong and Adamson's (2015) study, conducted in the foreign language-teaching context of Hong Kong. They revealed that most of the students agreed that feedback helped their learning, but they were not satisfied with their teacher's feedback. They also had a negative feeling toward the concept of assessment.

Considering teachers' assessment competency and skills, Plake et al. (1993) conducted a significant research study in which they investigated the assessment skills of 555 teachers and 268 administrators from 45 different states in the USA. The Teacher Competencies Assessment Questionnaire (TCAQ), a 35-item instrument, was designed in the first phase of the study to assess the seven competency criteria. The instrument was assessed by a 10-member NCME panel to determine construct validity before it was pilot-

tested with 70 instructors to get reliability estimates (Crocker & Algina, 2006). As the first study to investigate teacher assessment competency, the findings revealed considerable gaps in teachers' comprehension and implementation of assessment as shown by a sixty-six percent average score across the 35 items. Participants lacked skill in reading, integrating, and conveying assessment data, in particular. Later, they used their findings for continuing professional development programs. Based on Plake et al.'s (1993) study, O'Sullivan and Johnson (1993) employed the TCAQ with 51 graduate students engaged in a teacher assessment course. The course introduced students to performance-based assignments that were tied to the Standards (AFT, NCME, & NEA, 1990). Participants were invited to complete the Classroom Assessment Tasks (CAT) survey in addition to the pre- and post-TCAQ administrations to examine alignment between the Standards and the performance assessment tasks. Responses to Classroom Assessment Tasks indicated a significant alignment of performance tasks with the Standards, providing further validity evidence for the TCAQ. In a similar study, A revised version of the TCAQ was administered to 220 undergraduate students engaged in a pre-service measuring course by Campbell, Murphy, and Holt (2002). They discovered that teacher candidates' proficiency varied across the seven Standards based on their investigation. As a result, these researchers found that teacher candidates lacked crucial parts of competency when they entered the teaching profession.

Moreover, Mertler and Campbell (2004, 2005) collaborated on reconceptualizing the TCAQ into the Assessment Literacy Inventory (ALI). Their goal was to contextualize the items by reorganizing them into scenario-based questions, reflecting a more realistic approach to the Standards (AFT, NCME, & NEA, 1990). The ALI consisted of seven scenarios, each of which was tied to one of the Standards and was accompanied by a set of five multiple-choice. Like O'Sullivan and Johnson (1993), Mertler and Campbell (2004, 2005) administered the ALI to instructors participating in a measurement course to assess student learning

in reference to the Standards.

Moreover, considering the concept of critical language assessment, Tajeddin, Khatib, and Mahdavi (2022) have recently developed an inventory to assess EFL teachers' Critical Language Assessment Literacy (CLAL). The CLAL scale consisted of five factors which are (a)teachers' knowledge of assessment objectives, scopes, and types; (b) assessment use consequences; (c) fairness; (d) assessment policies; and (e) national policy and ideology.

Therefore, considering all the relevant studies, it was found that developing and validating a reliable scale to fulfill teachers' language assessment literacy is highly needed.

PURPOSE OF THE STUDY

Assessment is now served as a vital component of the language teaching process. It enables teachers to improve or change their instructional practices and also helps them evaluate students' progress and achievement in learning (Harris, Irving, & Peterson, 2008). Assessment literacy in foreign/second language learning and teaching is crucially important since it enables language teachers to understand, analyze, and utilize the information to improve their instruction (Falsgraf, 2005; Scarino, 2013). Furthermore, knowledge of assessment literacy helps language teachers choose the most effective and appropriate instruments to assess students' learning and progress (Siegel & Wissehr, 2011).

Despite the increasing and considerable importance of language assessment literacy, teachers' needs in the assessment literacy landscape has remained unexplored in many educational contexts. Moreover, since most of the teachers are involved in the process of decision-making and spend much of their professional time on developing and designing assessment-related tasks and activities, it is not still satisfactory (Brookhart, 2011, DeLuca & Klinger, 2010; Popham, 2009; Galluzzo, 2005; Zhang & Burry-Stock, 1997). Therefore, the aim of this study is to develop a valid and reliable

instrument that is representative of teachers' needs in assessment literacy, specifically the Iranian EFL high school teachers. In particular, an instrument that answers teachers' current needs within the existing educational accountability system and accounts for the numerous dimensions of assessment literacy beyond just addressing assessment purposes is required.

Particularly, this study drew on the Fulcher's assessment literacy framework (2012) as a basis for delineating EFL teachers' assessment literacy needs.

METHOD

Participants

Since Iran is a geographically vast country, it is not possible to collect data from every province and municipality. Therefore, the convenience sampling strategies in which the subjects are selected because of their convenient accessibility were employed for data collection purposes. Using Krejcie and Morgan's (1970) sample size table, the participants of the study were 159 Iranian EFL high school teachers working at public schools in different cities. The major of most of the participants was English language teaching, and a few of them had studied linguistics, translation, and English literature majors. All had more than five years of teaching experience. They also had different university degrees (BA, MA, or PhD).

The Instrument-Developing Procedure

To create a complete assessment literacy inventory that is representative of EFL teachers' current assessment literacy needs, the researchers adopted a multistep development method. To be more specific, the researchers (a) conducted a document analysis of prior and current assessment standards to aid them in early item construction, and (b) gathered validity data to support the intended interpretations and applications of the instrument. The 2014

Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and NCME, 2014) describe five sources of validity evidence (content, response processes, internal structure, relationship to other variables, and consequences). This article presents evidence of validity based on content and internal structure (construct).

According to Dörnyei (2007) the process of developing a standard questionnaire is a challenging procedure which requires some stages which are (a)initial item development, (b) initial piloting of the items, (c) final piloting and item analysis.

Initial Item Development

The first stage of a questionnaire development is collecting as many potential items for each section and creating a collection of items called the "item pool". In doing so, the researchers used two different sources: (a) reviewing the current literature on language assessment literacy including similar questionnaires and Fulcher's framework of language assessment literacy (2012), and borrowing some proper items from those published questionnaires that were properly acknowledged; and (b) interviewing EFL teachers and asking their needs and challenges in assessment literacy.

Since the Fulcher's language assessment literacy framework (2012) was the basis of this study, the item pool consisted of different types of items originated from the Fulcher's framework, and some items were based on EFL teachers' expectations of Continuing Professional Development (CPD) programs about assessment literacy (97 items). The items were about EFL teachers' knowledge and principles of assessment, familiarity with test processes, skills, and abilities to place knowledge in real situations, and their expectations of Continuing Professional Development (CPD) programs.

In total, in creating the item pool, the form of the items, the wording of the items, and the types of responses that the questionnaire is designed to induce were taken into account. Furthermore, all items were based on a five-

point Likert scale, in which the participants had to determine their needs in assessment literacy by indicating the extent to which they agree with each statement using (1) not at all, (2) very little, (3) little, (4) moderate, (5) a lot for all items.

Evidence of Validity Based on Content: Expert-panel Review

The degree to which the measurement encompasses most of the dimensions of the concept under research is referred to as content validity; hence, an instrument is deemed valid if it considers all the associated features of the concept under research. Therefore, the second stage of developing a questionnaire refers to the initial piloting of the item pool for the purpose of reducing the large list of items gathered from the previous stage to the intended final number. To do so the researchers asked a panel of experts including EFL head teachers, university instructors, and experts in assessment literacy to go through the items and provide feedback. They were asked to check its face and content validity and if necessary, change, add, or remove some items.

As Morgado et al., (2017) claimed, expert judges are well-versed in the topic of interest and/or scale development. Moreover, target population judges are potential scale users. Eleven university instructors and eighteen EFL head teachers who were experts in assessment literacy went through the items, and based on expert feedback, 21 items were removed, and the total number of items decreased to 76.

Evidence Based on Internal Structure: Pilot Testing

The third stage is to do the final piloting and item analysis. Based on the feedback received from the panel of experts, the researchers put together a near-final version of the questionnaire that seemed satisfactory and did not have any glitches. Subsequently, the researchers created an online version of the questionnaire through using Google Docs. The link of the questionnaire was sent to the target group through social media networks. As the

participants completed the questionnaire and submitted it, their responses were automatically loaded into a database on the web server, from which they were downloaded onto Microsoft Excel. Finally, 159 questionnaires were used for data analysis.

The item analysis is the final stage of developing a questionnaire. To fine-tune and finalize the questionnaire, the researchers subjected the responses of the pilot group (EFL high school teachers) to statistical analysis and checked the missing responses and possible signs that the instructions were not understood correctly, and the range of responses elicited by each item, the internal consistency of multi-item scales, and factor analysis were calculated. Based on the results of reliability and factor analysis, the researchers excluded the items that did not work properly and selected the best items related to the purpose of the study.

A heterogeneous sample, i.e., a sample that both reflects and captures the range of the target population was selected for the purpose of piloting.

RESULTS

Construct Validity Analysis

The construct validity of a questionnaire may be verified using factor analysis (Bornstedt, 1977; Ratray & Jones, 2007). A questionnaire has construct validity when all the items represent the underlying construct. Based on the relationships between variables (in this study, questionnaire items), exploratory factor analysis identifies the constructs - i.e., factors - that underpin a dataset (Field, 2009; Rietveld & Van Hout, 2011; Tabachnik & Fidell, 2007). The underlying constructs are supposed to be the ones that explain the greatest fraction of the variation shared by the variables. Factor analysis, unlike the frequently used principal component analysis, does not assume that all variance within a dataset is shared (Costello & Osborne, 2005; Field, 2009; Rietveld & Van Hout, 2011; Tabachnik & Fidell, 2007). Therefore, factor analysis is assumed to be a more reliable questionnaire evaluation method than principal component analysis (Costello & Osborne,

2005).

Initially, the factorability of the 76 TALNs items was examined. Several well-recognized criteria for the factorability of a correlation were used. Firstly, it was observed that most items have some correlations with each other, suggesting reasonable factorability. Secondly, a large enough sample size is required to undertake a credible factor analysis (Costello & Osborne, 2005; Field, 2009; Tabachnik & Fidell, 2007). In order to determine whether the sample size is large enough, the Kaiser-Meyer-Okin's measure of sampling adequacy (KMO) was calculated. According to Field (2009, p.647), the KMO "represents the ratio of the squared correlation of variables to the squared partial correlation of variables". Table 1 presents the estimated KMO and Bartlett's Test of Sphericity of the present study.

Table 1. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.730
Bartlett's Test of Sphericity	Approx. Chi-Square	1.654E4
	Df	2850
	Sig.	.000

Based on Field (2009), when the KMO is near 0, it is difficult to extract a factor. On the other hand, when the KMO is close to 1, a component or factors may most likely be retrieved since the opposing pattern is observable. Thus, KMO "values between 0.5 and 0.7 are average, values between 0.7 and 0.8 are acceptable, values between 0.8 and 0.9 are excellent, and values beyond 0.9 are exceptional" Field (2009, p.647).

As a result, as Table 1 shows, the KMO of the present study is 0.730, which is near 1, and it is suitable to justify the sample adequacy to do the factor analysis (Kaiser, 1974; Pallant, 2020; Field, 2009; Tabachnick & Fidell, 2007). Therefore, the sample size was large enough. Moreover, the

communalities were all above 0.3, and Bartlett's test of sphericity was significant ($p < .05$). This means that the variables are correlated highly enough to provide a reasonable basis for factor analysis. Given these overall indicators, factor analysis was deemed to be suitable with all 76 items.

Seventy-six items relating to EFL teachers' assessment literacy needs were factor analyzed using Exploratory Factor Analysis (EFA) with Varimax rotation. The analysis yielded four components explaining a total of 53.545% of the variance for the entire set of variables. Component 1 was labeled knowledge and basics of Language Assessment Literacy (LAL) due to the high loadings by the items related to this issue. The second component derived was labeled Practical effects of Language Assessment Literacy (LAL) on real situations. This component was labeled as such due to the high loadings by the items related to the consequences of assessment on different issues. Due to the high loadings by the items related to the testing processes and principles, component 3 was labeled principles and processes of Language Assessment Literacy (LAL). The label of each component was based on the Fulcher's assessment literacy framework (2012). And finally, the fourth component was labeled needs of Continuing Professional Development (CPD) Programs. This component was labeled as such due to the high loadings by the items related to the EFL teachers' expectations of the CPD programs. After rotation, to improve clarity, variables with loadings lower than 0.3 were considered to have a nonsignificant impact on a factor; therefore, they were omitted (Field, 2009). Moreover, items located in three or four components were also omitted (11 items were omitted). Table 2 displays the items and factor loadings for the rotated factors.

Table 2. Rotated component matrix

Item	As an EFL teacher I need to know more about……………	1	2	3	4
		Components			
57	the use and interpretation of both descriptive and inferential statistics	.906			
30	the knowledge of the process of conducting item analysis	.841			
15	the use of advanced statistics (e.g., Classical True Score theory, Generalizability theory, Item Response theory, SEM)	.828			
53	the history of language testing (pre-scientific, psychometric, structuralist, sociolinguistic-pragmatic)	.828			
71	research methods in setting up experiments in testing (e.g., quantitative, qualitative, and mixed-methods approaches)	.823			
54	the basic concepts of language testing and assessment (e.g., tests, measurement, evaluation, test use, test type, test format)	.821			
29	the knowledge of the process of conducting test analysis	.816			
75	knowing scales of measurement (e.g., nominal, ordinal, interval and ratio scale)	.794			
2	test validity and its different forms (e.g., predictive, concurrent, content, construct, face, response)	.788			
3	test reliability and its different forms (e.g., test-retest, parallel forms, split-haves, Kuder-Richardson formulae, Cronbach's alpha, scorer reliability)	.769			
16	the use of more modern statistical tests (e.g., Multilevel modelling, Autoregressive SEM models, Latent growth curve modelling, Time series approaches, Event history analysis)	.738			
35	doing pre-test (item facility, item discrimination, choice distribution)	.677			
68	theories of testing (traditional testing, discrete-point testing, integrative testing, communicative testing)	.663			
7	testing models and frameworks	.648			

18	features of developing a good test	.636	
19	authenticity in test developing	.587	
9	different types of test score interpretations (norm-referenced and criterion-referenced interpretation)	.587	
17	the principles of rubric development	.574	
26	interactiveness in testing (interaction between test takers' characteristics and test tasks)	.551	
28	accountability (obligation of teachers to accept responsibility for students' performance)	.540	
70	test critique (critical evaluation of tests)	.536	
27	learners' preparation to take a test	.491	
55	different types of tests and designs of assessments for all four language skills (i.e., reading, writing, speaking, listening)		.488
6	the differences between testing, assessment and measurement	.459	
8	different types of tests and assessments and their usages (objective versus subjective, essay type versus multiple choice)	.456	
44	the practical effects of assessment on students' performance	.889	
52	the use of tests scores and interpretations in educational programs	.848	
45	the social consequences of tests	.831	
49	the psychological consequences of tests (e.g., memory improvement, students' learning style, …)	.819	
43	the practical effects of assessment literacy on teachers' teaching strategies	.789	
50	the responsibility of test takers	.783	
48	the educational consequences of tests (e.g., educational decisions, reforming the curriculum, …..)	.778	
51	the use and effects of tests on educational programs	.748	
46	the political consequences of tests (e.g., educational policies, ….)	.730	
47	the economical consequences of tests	.704	
37	the effects of using different platforms of online assessment on educational programs (e.g., testmoze, google doc, Monta,….)	.637	
38	the effects of different types of tests on learning and	.586	

	teaching			
63	various platforms for online assessment			.517
34	the process of developing and using personal response assessments (e.g., checklists, journals, videotapes, audiotapes, self-assessment, teacher observation, portfolios, conferences, diaries)		.499	
59	the principles of developing a good test			.430
66	different types of tests and their functions and effects	.413		
69	the use of alternative assessment	.386		
64	the effects of using various computer software programs for test construction, test analysis and test scoring on educational programs	.359		
72	the effect of test taking strategies on learning and teaching	.316		
21	the process of developing a good test and test specifications		.789	
22	the principles of educational measurement		.773	
23	the principles of using tests in society		.698	
25	the effect of tests on teaching/learning (washback)		.664	
5	the design of assessments for productive skills (speaking and writing)			.529
20	test bias and analyzing it in test designs (e.g., cultural background, ethicality, sex, native language, background knowledge)		.512	
11	ethical issues in assessment		.502	
32	the process of administrating oral/written exams			.483
60	the responsibility of test takers and test givers			.435
33	the process of developing and using constructed-response assessments (e.g., fill in the blank, short answer)		.402	
41	the process of test administration			.671
56	different types and use test scores and their interpretation in educational programs			.625
40	the functions of tests (achievement, proficiency, aptitude, selection, placement, diagnosis)			.623
39	different interpretation of tests			.582

76	administrating and scoring oral and written exams	.582
42	the process of writing test specifications	.507
65	the process of making assessment real and personal	.497
31	the process of administrating and scoring computer-based testing	.410
73	providing test security	.391
58	the ethical issues in assessment	.389
4	the design of assessments for receptive skills (reading and listening)	.355

After rotation, the first part of the questionnaire which refers to the EFL teachers' needs of knowing more about Language Assessment Literacy (LAL) basics and history, accounted for 24 items. The following table (Table 3) represents this part.

Table 3. The items related to factor1(knowledge of LAL)

Item	As an EFL teacher how much training do you need on …. ……….?	Loaded Factor
1	the use and interpretation of both descriptive and inferential statistics	.906
2	the knowledge of the process of conducting item analysis	.841
3	the use of advanced statistics (e.g., Classical True Score theory, Generalizability theory, Item Response theory, SEM)	.828
4	the history of language testing (pre-scientific, psychometric, structuralist, sociolinguistic-pragmatic)	.828
5	research methods in setting up experiments in testing (e.g., quantitative, qualitative, and mixed-methods approaches)	.823
6	the basic concepts of language testing and assessment (e.g., tests, measurement, evaluation, test use, test type, test format)	.821
7	the knowledge of the process of conducting test analysis	.816
8	knowing scales of measurement (e.g., nominal, ordinal, interval and ratio scale)	.794
9	test validity and its different forms (e.g., predictive, concurrent, content, construct, face, response)	.788
10	test reliability and its different forms (e.g., test-retest, parallel forms, split-haves, Kuder-Richardson formulae, Cronbach's alpha, scorer reliability)	.769
11	the use of more modern statistical tests (e.g., Multilevel modelling, Autoregressive SEM models, Latent growth curve modelling, Time series	.738

	approaches, Event history analysis)	
12	doing pre-test (item facility, item discrimination, choice distribution)	.677
13	theories of testing (traditional testing, discrete-point testing, integrative testing, communicative testing)	.663
14	testing models and frameworks	.648
15	features of developing a good test	.636
16	authenticity in test developing	.587
17	different types of test score interpretations (norm-referenced and criterion-referenced interpretation)	.587
18	the principles of rubric development	.574
19	interactiveness in testing (interaction between test takers' characteristics and test tasks)	.551
20	accountability (obligation of teachers to accept responsibility for students' performance)	.540
21	test critique (critical evaluation of tests)	.536
22	learners' preparation to take a test	.491
23	the differences between testing, assessment and measurement	.459
24	different types of tests and assessments and their usages (objective versus subjective, essay type versus multiple choice)	.456

The second part of the questionnaire which refers to the EFL teachers' needs of knowing more about the effects of Language Assessment Literacy (LAL) on real life situations accounted for 16 items. Table 4 shows the items of this part.

Table 4. The items of factor 2 (effects of LAL)

Item	As an EFL teacher how much do you need to know about …………?	Loaded Factor
1	the practical effects of assessment on students' performance	.889
2	the use of tests scores and interpretations in educational programs	.848
3	the social consequences of tests	.831
4	the psychological consequences of tests (e.g., memory improvement, students' learning style, …)	.819
5	the practical effects of assessment literacy on teachers' teaching strategies	.789
6	the responsibility of test takers	.783
7	the educational consequences of tests (e.g., educational decisions, reforming the curriculum, …..)	.778
8	the use and effects of tests on educational programs	.748

9	the political consequences of tests (e.g., educational policies, ….)	.730
10	the economical consequences of tests	.704
11	the effects of using different platforms of online assessment on educational programs (e.g., testmoze, google doc, Monta,….)	.637
12	the effects of different types of tests on learning and teaching	.586
13	different types of tests and their functions and effects	.413
14	the use of alternative assessments	.386
15	the effects of using various computer software programs for test construction, test analysis and test scoring on educational programs	.359
16	the effect of test taking strategies on learning and teaching	.316

The third part of the questionnaire, which refers to the EFL teachers' needs of knowing more about the principles and processes of Language Assessment Literacy (LAL), accounted for 8 items. Table 5 shows the items of this part.

Table 5. The items of factor 3 (processes of LAL)

Item	As an EFL teacher how much do you need to know about……………………?	Loaded Factor
1	the process of developing a good test and test specifications	.789
2	the principles of educational measurement	.773
3	the principles of using tests in society	.698
4	the effect of tests on teaching/learning (washback)	.664
5	test bias and analyzing it in test designs (e.g., cultural background, ethicality, sex, native language, background knowledge)	0512
6	the process of developing and using personal response assessments (e.g., checklists, journals, videotapes, audiotapes, self-assessment, teacher observation, portfolios, conferences, diaries)	.499
7	ethical issues in assessment	.502
8	the process of developing and using constructed-response assessments (e.g., fill in the blank, short answer)	.402

The fourth part of the questionnaire which refers to the EFL teachers' Language Assessment Literacy (LAL) expectations of CPD programs accounted for 17 items. Table 6 shows the items of this part.

Table 6. The items of factor 4 (CPD programs)

Item	As an EFL teacher I need to participate in CPD programs to know more about………	Loaded Factor
1	different types of tests and designs of assessments for all four language skills (i.e., reading, writing, speaking, listening)	.488
2	various platforms for online assessment	.517
3	the principles of developing a good test	.430
4	the design of assessments for productive skills (speaking and writing)	.529
5	the process of administrating oral/written exams	.483
6	the responsibility of test takers and test givers	.435
7	The process of test administration	.671
8	different types and use test scores and their interpretation in educational programs	.625
9	the functions of tests (achievement, proficiency, aptitude, selection, placement, diagnosis)	.623
10	different interpretation of tests	.582
11	administrating and scoring oral and written exams	.582
12	the process of writing test specifications	.507
13	the process of making assessment real and personal	.497
14	the process of administrating and scoring computer-based testing	.410
15	providing test security	.391
16	the ethical issues in assessment	.389
17	the design of assessments for receptive skills (reading and listening)	.355

Reliability Analysis

After doing Exploratory Factor Analysis (EFA) and removing the insignificant items and categorizing the items in their appropriate group, the internal consistency of the questionnaire should be assessed (Field, 2009). Cronbach's Alpha was used to estimate the reliability of four parts of the questionnaire separately and totally. Cronbach's Alpha is a coefficient used to rate the internal consistency (homogeneity) or correlation of the items in a questionnaire together. If a questionnaire enjoys strong internal consistency, most measurement experts (e.g., Field, 2009; Garson, 2010; Cortina, 1993) agree that it should show only moderate correlation among items. According to Field (2009), a questionnaire with an α of 0.8 (or more) is considered

reliable. The reliability indices (table 7) reveals that all parts of the questionnaire enjoy high level of internal consistency.

Table 7. Cronbach's Alpha reliability coefficient

	Reliability Coefficient (α)
Part 1: EFL teachers' language assessment literacy knowledge	.916
Part2: Principles and processes of language assessment literacy	.862
Part 3: Practical aspects of language assessment literacy	.906
Part 4: EFL teachers' expectations of CPD programs	.854
Total	.901

DISCUSSION

Despite the fact that language assessment literacy is considered essential for teacher development (DeLuca, LaPointe-McEwan, & Luhanga, 2016; Gotch & French, 2014), no validated instrument has been designed for investigating teachers' assessment literacy needs. With that in mind, this study was an attempt to develop an inventory to reflect the EFL teachers' needs in assessment literacy. To fulfill this purpose, exploratory analysis was used to examine the construct validity of the proposed inventory. Moreover, the Fulcher's assessment literacy framework (2012) was drawn on as the analytic model guiding the study. The EFA results indicated that the inventory can be best explained by four components, three of them, namely *language assessment literacy knowledge, principles and processes of language assessment literacy and practical aspects of language assessment literacy*, are based on the Fulcher's framework. The fourth component refers to the *teachers' expectations from CPD programs*.

The 24 items in the *language assessment literacy knowledge* factor clearly show the significance of knowing the basics of language assessment literacy, empowering teachers for classroom assessment. It is believed that

knowing the basics of language assessment literacy can empower teachers to be not only knowledgeable but also creative (Crocker & Algina, 2006) if taught thoroughly and meticulously.

The 16 items in the *practical aspects of language assessment literacy* factor obviously represent the consequences of assessment literacy on learners' real-life situations. Teachers' knowledge of social, educational, psychological consequences of assessment can empower them to be more precise in making decisions about the learners' future. Also, knowing these practical features help teachers contextualize the assessment activities (Mertler & Campbell, 2004, 2005). Thus, this is not surprising that in the context of L2 education, teachers need to acquire the knowledge of how to assess learners as well as how to make decisions about their future (Richards & Farrel, 2005).

The eight items in the *principles and process of language assessment* literacy factor emphasize the role of knowing the process of developing different types of assessment tasks in teacher education. Familiarity with test methods, techniques, processes, and awareness of test principles and practices including ethics help teachers be creator of new motivating assessment contexts (Jeong, 2013).

The 17 items in the *Expectations of CPD programs* factor reflect teachers' expectations from CPD programs about language assessment literacy. These items, also, highlight the importance of CPD programs in empowering teachers with newly-developed concepts in assessment literacy. Moreover, teachers' CPD programs not only do equip teachers with both learning and learner-centered assessment tasks but also help them expand their understanding of assessment technical knowledge (Behzadi, Golshan, & Sayadian , 2019).

CONCLUSION AND IMPLICATIONS

Although earlier inventories and surveys of teacher assessment literacy sparked much of this research, numerous academics have pointed out that

these measures are now out of date (e.g., Brookhart, 2011; DeLuca et al., 2016; Gotch & French, 2014), and none of them reflect teachers' assessment literacy needs and expectations. In the absence of any instrument investigating teachers' needs in assessment literacy, however, it has not been possible to quantify this construct in its operational terms. Thus, the present study was conducted to design and validate an instrument unique to EFL context.

The TALNs developed in this study is a new inventory for research and professional development in the field of teacher assessment literacy. It also gives directions to practitioners and researchers in the realm of language assessment literacy to provide teachers with professional development programs based on their needs and expectations. This study provides initial validity and reliability data to support the TALNs as a helpful indication of teachers' assessment literacy needs and their expectations from CPD programs. The TALNs data can spark crucial studies regarding teachers' current assessment literacy levels and needs, as well as a database for targeted professional learning.

As a result, we provide the TALNs in this paper as an instrument for use and development by researchers and educational practitioners in the service of improving teachers' language assessment literacy and perceiving their needs.

This study was an initial step towards Iranian EFL teachers' language assessment literacy needs. Further research is needed to make it more comprehensive and general. To do this, other EFL contexts such as language institutions, universities, and even other areas of education can be considered for future studies. Considering triangulation, also, researchers can benefit from other data gathering tools such as think aloud, interview, and observation. Additionally, the present study drew on the Fulcher's language assessment literacy framework (2012). Other studies can focus on other assessment frameworks as their analytic guiding model. Finally, other statistical analyses such as Confirmatory Factor Analysis (CFA) or Rasch model can be implemented for establishing the factors and assigning

items.

Disclosure statement
No potential conflict of interest was reported by the authors.

ORCID

Mohammad Reza Khodashenas   iD   http://orcid.org/0000-0002-6574-0654
Hossein Khodabakhshzadeh   iD   http://orcid.org/0000-0001-6129-6667
Purya Baghaei   iD   http://orcid.org/0000-0002-5765-0413
Khalil Motallebzadeh   iD   http://orcid.org/0000-0001-8382-8978

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: National Council on Measurement in Education.

Behzadi, A., Golshan, M., & Sayadian, S. (2019). Validating a continuing professional development scale among Iranian EFL teachers. *Journal of Modern Research in English Language Studies 6*(3), 105-129.

Bornstedt, G.W. (1977). Reliability and validity in attitude measurement. In G. F. Summers (Ed.), *Attitude measurement*. London: Kershaw Publishing Company.

Brookhart (2011). Educational assessment and knowledge skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3-12.

Brown, G. T., Pishghadam, R., & Sadafian, S. S. (2014). Iranian university students' conceptions of assessment: Using assessment to self-improve. *Assessment Matters*, *6*(1), 5-33.

Campbell, C., Murphy, J. A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. In

Annual meeting of the mid-western educational research association, Columbus, OH (pp. 17-18).

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104.

Costello, A.B., & Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment*, *Research and Evaluation*, *10*(1), 1-9.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.

Dehqan, M., & Asadian Sorkhi, S. R. (2020). Pre-service and in-service teachers' knowledge and practice of assessment literacy: a dweller in an Ivory Tower. *Issues in language teaching, 9*(2), 347-375.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438.

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

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Falsgraf, C. (2005). *Why a national assessment summit? New visions in action*. National Assessment Summit. Meeting conducted in Alexanderia, Va.

Field, A. (2009). *Discovering statistics using SPSS statistics*. UK: Sage publication.

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

Galluzzo, G. R. (2005). Performance assessment and renewing teacher education. *Clearing House*, *78*(4), 142-145.

Garson, D. (2010). *Factor analysis: topics in multivariate analysis*. http://www.chass.ncsu.edu/garson/pa756/factor-htm.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, *33*(2), 14-18.

Harris, L., Irving, S., & Peterson, E. (2008). *Secondary teachers' conceptions of the purpose of assessment and feedback*. [Paper presentation]. Annual conference of the Australian Association for Research in Education,

Brisbane, Australia.

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, *30*(3), 345-362.

Joint Committee on Standards for Educational Evaluation. (2015). *Classroom assessment standards: Practices for PK-12 teachers.* http://www.jcsee.org/the-classroom-assessment-standards-new-standards.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement, 30*(3), 607-610.

Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?* [Paper presentation]. Annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

Mertler, C. A., & Campbell, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, *33*, 49-64.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. [Paper presentation]. The annual meeting of the American Educational Research Association, Montreal, QC, Canada.

Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., & Ferreira, M. E. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*(3). https://doi:10.1186/s41155-016-0057-1.

O'Sullivan, R. S., & Johnson, R. L. (1993). *Using performance assessments to measure teachers' competence in classroom assessment.* [Paper presentation]. The annual meeting of the American Educational Research Association, Atlanta, GA.

Öz. S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: perceptions and practices. *International Association of Research in Foreign Language Education and Applied Linguistics ELT Research Journal*, *6*(1), 25-44.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. London: Routledge.

Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: A

national survey. *Educational Measurement: Issues and Practice*, *12*(4), 10-39. https://doi:10.1111/j.1745-3992.1993. tb 00548.x.

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

Rattray, J.C., & Jones, M.C. (2007), Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, *16*(2), 234-243.

Rezaeifard, Z., & Tabatabaei, O. (2018). Investigating assessment literacy of EFL teachers in Iran. *Journal of Applied Linguistics and Language Research*, *5*(3), 91-100.

Richards, J. C., and Farrell. T. S. C (2005). *Professional Development for Language Teachers: Strategies for Teacher Learning.* Cambridge: Cambridge University Press.

Rietveld, T., & Van Hout, R. (2011). *Statistical techniques for the study of language and language behaviour*. Berlin: De Gruyter Mouton.

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

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education, 22*(4), 371-391.

Tabachnik, B.G., & Fidell, L.S. (2007) *Using multivariate statistics*. New York, NY. Pearson.

Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*. https://doi.org/10.1177/02655322211057040.

Tong, S.Y. A, & Adamson, B. (2015). Student voices in school-based assessment. *Australian Journal of Teacher Education, 40*(2), 2.

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

Zhang, Z., & Burry-Stock, J. A. (1997). *Assessment Practices Inventory: A multivariate analysis of teachers' perceived assessment competency*. [Paper presentation]. Annual meeting of the American Educational Research Association, Chicago, IL.