Content list available at http://ijltr.urmia.ac.ir

Iranian Journal of Language Teaching Research

Urmia University

Developing a Language Assessment Knowledge Test for EFL Teachers: A Data-driven Approach

Hossein Farhady [a, *], Kobra Tavassoli [b]

[a] *Yeditepe University, Turkey*

[b] *Islamic Azad University – Karaj Branch, Iran*

ABSTRACT

Language assessment knowledge (LAK) is essential for EFL teachers in developing sound assessments and making informed decisions about the progress and achievement of students. It is also essential to have sound measures of teachers' LAK to help them improve their professional accomplishments. To address the issue, this study was designed in multiple stages to develop a test to measure EFL teachers' LAK. First, EFL teachers' needs were identified using an already validated questionnaire (Farhady & Tavassoli, forthcoming). Second, the topics identified through needs analysis were cross-checked with the topics in the assessment textbooks to assure appropriate content coverage. Third, based on the data obtained from needs assessment of 246 EFL teachers and the content of popular language assessment textbooks, a scenario-based LAK test with closed-item formats was developed in six parts, each focusing on one major area of language assessment. Fourth, after having the test reviewed by experts, it was revised and piloted with 50 EFL teachers, necessary revisions were made based on the results of piloting, and the test was finalized and administered to 164 EFL teachers. The data analysis revealed that contrary to teachers' claims in needs assessment, the majority of them had low levels of LAK and they showed willingness to improve their LAK. Further, possible relationships between teachers' LAK and their gender, field of study, educational level, teaching context, and teaching experience were investigated. The findings are presented and implications and applications of the study are explained.

Keywords: assessment knowledge; Data-driven approach; EFL teachers; language assessment knowledge; professional knowledge; teacher education; test development

ARTICLE HISTORY

Received: 27 July 2018 **Revised version received:** 14 Aug. 2018

Accepted: 21 Sept. 2018 **Available online:** 1 Oct. 2018

* Corresponding author: Yeditepe University, Istanbul, Turkey
Email address: hossein.farhady@yeditepe.edu.tr

Introduction

In the past few decades, assessment as an ongoing process of collecting information to evaluate students' performance has been a desirable alternative to traditional one-shot case testing (Brown & Abeywickrama, 2010). This trend has influenced all stakeholders in the assessment process, especially teachers because they are expected to move away from the traditional testing approach and gain necessary knowledge to prepare and use various assessment techniques. Due to such expectations, the teachers' role in the classroom has received considerable attention in the last decades because it was believed that the teacher is the most important factor influencing student learning (Darling-Hammond, 2000; Wallace, 1991). Moreover, many scholars believed that "more can be done to improve education by improving the effectiveness of teachers than by any other single factor" (Wright, Hom, & Sanders, 1997, p.63). Consequently, training teachers and helping them to develop professionally and improve different dimensions of their professional knowledge drew the attention of teacher education institutions (Wallace, 1991). As part of teachers' professional knowledge, assessment knowledge received more attention by different scholars in education such as Popham (2009, 2011) and Stiggins (2005, 2008) since assessment is an indispensable part of any educational endeavor. Likewise, EFL teachers' language assessment knowledge (LAK) has been addressed by many scholars including Farhady and Tavassoli (2017), Inbar-Lourie (2013), Malone (2013), and Taylor (2009).

LAK may be conceptualized as a construct that is rooted in theoretical principles of both language education and language assessment, is shaped by social and cultural contexts, and is filtered by individuals' perceptions of teaching and assessment (Xu, 2015). A reasonable command of LAK is believed to help teachers design, administer, collect, and interpret assessment data to make fair decisions (Inbar-Lourie, 2013).

Despite the importance of teachers' LAK, researchers from language education (Inbar-Lourie, 2008; Malone, 2008; Taylor, 2013) have reported dissatisfaction with the status of EFL teachers' LAK. A glance at research findings reported in the literature reveals that most teachers have claimed that they do not feel comfortable with the application of measurement and evaluation principles in practice. Some scholars even claim that teachers assess students' performance with little or no professional training (Bachman, 2000). Teachers' insufficient command of assessment knowledge is often attributed to the fact that they do not receive effective training on assessment during their pre-service or in-service teacher education programs. Therefore, EFL teachers' LAK has been the focus of many investigations in recent years (e.g., Berry, Munro, & Sheehan, 2016; Eberharter, Kremmel, Holzknecht, & Konrad, 2016; Farhady & Tavassoli, 2015, 2017; Hamp-Lyons, 2017; Harsch, Seyferth, & Brandt, 2017; Janatifar, Marandi, & Babaei, 2017; Jin & Jie, 2017; Lam, 2015; Sellan, 2017; Stabler-Havener, 2018; Tsagari & Vogt, 2015, 2017; Xu & Brown, 2017).

Available Measures of LAK for EFL Teachers

In order to collect information on teachers' level of LAK, an appropriate test or instrument is necessary. Despite the existence of ample research on LAK, a review of the literature reveals that there is not a well-established test to measure EFL teachers' LAK. Only a few items on language assessment were observed in professional knowledge tests for language teachers that are briefly presented below.

There are various types of professional knowledge tests for both pre-service and in-service teachers, including English language teachers. For instance, as Stabler-Havener (2018) stated, pre-service English language teachers who want to receive a teaching certificate in the US should receive a passing score on the *Praxis tests*. The Praxis tests measure the academic skills and content knowledge needed for people entering the teaching profession. This certificate is required by many states and professional organizations in the US (ETS website, 2018). The assessment section of the Praxis tests comprises 18 items out of 120 test items, i.e., 15 percent of the whole test. The few number of items do not seem sufficient to provide a clear picture of the LAK levels of the prospective English language teachers.

Further, to determine the LAK of in-service teachers, the *National Board Certification (NBC)* test is required. NBC is designed to develop, retain, and recognize skilled teachers and to generate ongoing improvement in US schools. It is designed for teachers in various areas including the English language. NBC includes four components: (1) content knowledge, (2) differentiation in instruction, (3) teaching practice and learning environment, and (4) effective and reflective practitioner (National Board for Professional Teaching Standards website, 2018). The first component is a computer-based assessment taken at a testing center whereas the other three components are portfolio-based and submitted electronically. The first component asks the participants to demonstrate their understanding of content knowledge and pedagogical practices for teaching their specific content area. The second component requires the participants to collect and analyze information about their students' strengths and needs and use that information to design appropriate instruction to improve student learning. The third component requires video recordings of interactions between participating teachers and their students. The participants should then provide a written report where they describe, analyze, and reflect on their teaching and interactions with students. The fourth component requires the participants to demonstrate their abilities as an effective and reflective practitioner in improving their students' knowledge, in using assessments to effectively enhance their students' learning, and in their collaboration with peers, colleagues, and parents to advance students' learning (National Board for Professional Teaching Standards website, 2018).

Another popular international test of professional knowledge of EFL teachers is the *Teaching Knowledge Test (TKT)*, which was developed in the UK and first administered in May 2005. Performance on TKT is assumed to manifest teachers' level of professional development. The test seems to be desirable for people who want to demonstrate their teaching knowledge as measured by a globally recognized test. TKT provides an internationally accepted certificate that is interpreted as one's language-teaching ability (British Council website, 2018).

TKT is made up of three core modules that can be taken together or separately in any order. Module 1 is named *language and background to language learning and teaching*, with a focus on description of language, language skills, and background to language teaching and language learning. Module 2 is entitled *lesson planning and use of resources for language teaching*. This module focuses on teachers' ability to plan and prepare a lesson or sequence of lessons, and to select and use resources and materials effectively. Module 3 is termed *managing the teaching and learning process*, with a focus on teachers' and learners' language in the classroom, and classroom management. There are some specialist modules too including *TKT knowledge about language* that attempts to test one's knowledge and understanding of the systems of the English language; *TKT content and language integrated learning* that tests one's understanding of teaching; and *TKT young learners* that is intended to tests one's knowledge of the strategies and skills required to teach English to young learners. The item formats in TKT modules are matching, sequencing, multiple-choice, and odd one out (Cambridge ESOL website, 2018). Although there are a few assessment items in each TKT module, it is still an indication of attention paid to EFL teachers' LAK levels. However, it seems difficult to make any claims about teachers' actual LAK levels on the basis of their performance on a few assessment items.

Despite the international scope of TKT, there is no convincing report on the logic behind the selection of items beyond the expert intuition of the test developers. Nor are there research reports on the effectiveness of this test in the real context of teaching. Therefore, a more empirical and data-driven approach seems necessary to address the issue.

Finally, an attempt was recently made by Khany and Azimi Amoli (2016) who developed and validated a teacher professional development scale in multiple stages in the EFL context of Iran. The test is designed to measure EFL teachers' professional knowledge. The scale covers the essential components of teacher professional development, which are utilized in everyday teaching and learning practices in the classroom. One part of the scale deals with the teachers' assessment knowledge.

Of the available professional knowledge tests and teacher certifications, it seems that only NBC provides a reasonably accurate estimate of English language teachers' LAK levels. However, as with the other tests, it is unlikely that the limited number of items on assessment in NBC would provide adequate data to make accurate decisions about teachers' LAK levels (Stabler-Havener, 2018). In addition, a problem with the available professional knowledge tests is that none of them is data-driven. In other words, none of these tests is based on data collected from stakeholders at earlier stages to help the test development process. Therefore, considering the significance of measuring EFL teachers' LAK and the fact that there is not any sound instrument to do so, this study was designed to develop a data-driven LAK test for EFL teachers.

Test Development Process

Test construction is a multi-dimensional process that often includes the following stages: (1) determining the function and form of the test; (2) planning or determining the content of the test;

(3) preparing different types of items; (4) reviewing the test by experts; (5) pretesting; and (6) determining the psychometric characteristics of the test (Farhady, Jafarpur, & Birjandi, 1994). These stages were also delineated by Bachman and Palmer (1996) under design, operationalization, and administration. *Design* deals with describing the purpose of the test, identifying the tasks, describing the characteristics of the test takers, defining the construct to be measured, developing a plan for evaluating the qualities of the test, and allocating and managing resources. *Operationalization* refers to developing the test tasks, writing instructions, and specifying the scoring method. Finally, *administration* deals with administering the test and collecting feedback, analyzing test scores, and archiving the test tasks and the necessary information.

More recently, Downing (2006, p.5) as well as Lane, Raymond, Haladyna, and Downing (2016, p.4) offered a detailed and systematic approach to test development that includes twelve steps:

1) Overall plan: developing a detailed plan for the whole process of test development;

2) Domain definition: defining the domain to be measured and providing a clear statement of the claims to be made;

3) Content specifications: developing specifications to guide item development, score reporting, etc.;

4) Item development: identifying suitable item formats and developing items;

5) Test design and assembly: creating test forms based on test specifications, and attending to issues about test content, format, scoring rules, scaling, and equating;

6) Test production: producing a clear and precise test form;

7) Test administration: administering the test in a standardized way;

8) Scoring: establishing quality control procedures for scoring responses, and ensuring accurate and consistent scoring;

9) Establishing cut scores: establishing defensible cut scores consistent with the purpose of the test;

10) Reporting test scores: developing understandable test score reports;

11) Test security: establishing policies and procedures for ensuring test security;

12) Test documentation: preparing documentations and technical reports on validity, fairness, and the technical adequacy of the test.

At initial stages of test development, the content of the test is to be determined and the tasks/ items are to be identified and prepared in relation to a specific set of materials. The test development steps mentioned by different scholars indicate that rarely is a test developed based on data collected at early stages. Usually, data collection occurs after a test is developed to check the quality of the items, the tasks, and the whole test. Even a review of recent papers published in

language assessment journals reveals that the few tests which have been developed recently follow the same steps and do no collect data at early stages (e.g., He & Min, 2017; Kim & Elder, 2015).

However, using a data-driven approach to test development, i.e., collecting data at early stages of test construction to feed the test writing process, has not been explored adequately. The professional knowledge tests mentioned above, such as TKT, though popular and widely used in the world, are not data-driven. The data-driven approach in the test construction process is a significant feature of the present test. The test is developed on the basis of collecting information on different stakeholders' ideas and identifying and implementing their needs at early stages. Though needs analysis is a commonly used technique at initial stages of materials development (Richards, 2001), it is not a commonly used technique in test construction stages. Consequently, attempts were made in this study to follow a data-driven approach and feed the information received from the earlier stage of needs analysis of EFL teachers' LAK into developing a LAK test for these teachers. In addition, information obtained from the language assessment literature on what constitutes LAK was utilized to cross-check with the results of needs analysis. The procedures of test development are detailed below.

Method

Despite the existence of ample research on EFL teachers' LAK on the one hand, and the existence of various instruments to check teachers' professional knowledge on the other hand, there is no single test to measure teachers' LAK specifically. Therefore, a comprehensive research agenda in multiple stages was designed in Iran to address the issue. The first stage of the project was identifying the assessment needs of EFL teachers and reported in Farhady and Tavassoli (forthcoming). The second stage, which is the focus of this paper, was developing a data-driven LAK test for EFL teachers. Consequently, the following seemingly simple research question was posed to address this issue:

- How much language assessment knowledge (LAK) do EFL teachers have?

To answer this question, certain steps were taken. The first was identifying EFL teachers' needs for LAK. Using Fulcher's (2012) needs assessment questionnaire, the assessment needs of 246 EFL teachers were identified. The results showed that most of the teachers claimed they were aware of the major topics to be incorporated in a language assessment course and they had the basic knowledge on topics related to language assessment (Farhady & Tavassoli, forthcoming).

In the second part of the study, based on the information obtained from needs assessment, the topics which were identified as essential or important by a significant majority of the participants were selected and cross-checked with the topics in the available language assessment textbooks. The topics were grouped to create more inclusive ones. Then, a scenario-based LAK test was developed in six parts, each part focusing on one major area of language assessment, with closed-item formats of 'matching', 'ordering', and 'multiple-choice'. The major topics included in the LAK test were: Part A, *test types/ functions* (matching items); Part B, *stages of language test design*

(ordering items); Part C, *test characteristics* (matching items); Part D, *key terms/ concepts* (multiple-choice items); Part E, *alternative assessment techniques* (matching items); and Part F, *assessing language components and skills* (multiple-choice items).

Initially, there were 35 items in the test each carrying 1 point. The test was reviewed by experts in the field and it was revised. Next, it was piloted with 50 EFL teachers in Iran. The participant teachers had different ages and teaching experiences: 37 of them were females and 13 were males, almost all of them had BAs and taught at private institutes, and 39 of them had an English-related academic degree whereas 7 had an academic degree in other fields.

Based on the results of item analysis in piloting, all the non-functioning and mal-functioning items were revised or discarded. The test was revised again and finalized with 33 items. Eventually, it was administered to a group of 164 EFL teachers in Iran to check the level of their actual LAK on topics that were deemed significant in needs assessment. Again, information on teachers' age, gender, teaching experience, field of study, educational level, and teaching context was collected through the test to find out the possible relationships between teachers' LAK and these factors. The demographic information of the EFL teachers who participated in the final administration of the LAK test is presented in Table 1.

Table 1
Demographic Information of EFL Teachers in the Final Administration of LAK Test

Variable	Description
Age	19 to 60 years
Gender	120 female, 37 male
Teaching experience	A few months to 32 years
Field of study at university	28 TEFL, 37 English literature, 53 English translation, 21 other fields
Educational level	6 High-school diploma, 5 AA, 115 BA, 25 MA
Teaching context	37 at public schools, 121 at private institutes

Results and Discussion

Since the content of the LAK test was determined by the information obtained from needs assessment and the content of available language assessment textbooks, the test can be claimed to be content valid.

The reliability of the LAK test in the piloting stage estimated through Cronbach's Alpha was .71 that was considered acceptable (Pallant, 2005). Out of the initial 35 items each carrying 1 point, the minimum score was 6 and the maximum score was 28, with the mean and standard deviation of 15.78 and 5.26, respectively. The preliminary investigation showed that contrary to teachers' claims in needs assessment (Farhady & Tavassoli, forthcoming), the majority of them had low levels of LAK as indicated by the low mean of 15.78 out of 35. Further, the high value of standard deviation (5.26) showed that there was a high dispersion of scores among the participants.

The final version of the LAK test with 33 items was administered to a group of 164 EFL teachers (Table 1). The reliability of the finalized LAK test estimated through Cronbach's Alpha was .74, which was a little above the Alpha value in piloting. The descriptive statistics of the participants' scores on the final version of the LAK test are reported in Table 2.

Table 2
Descriptive Statistics of the EFL Teachers' Scores on the Finalized LAK Test

LAK Test Parts	N	Mean	SD
Part A, test types/functions (5 items)	164	2.74	1.39
Part B, stages of language test design (5 items)	164	2.34	1.58
Part C, test characteristics (4 items)	164	1.66	1.22
Part D, key terms/concepts (8 items)	164	3.04	1.67
Part E, alternative assessment techniques (4 items)	164	2.38	1.24
Part F, assessing language components and skills (7 items)	164	3.46	1.71
Total (33 items)	164	15.62	5.40

As shown in Table 2, the mean score of the final version of the test was also low (15.62 out of 33) confirming that the EFL teachers performed poorly on each part as well as on the whole test. Also, the relatively high standard deviation (5.40) indicated a high dispersion of scores on the final version of the test. Further, the minimum score on the whole test was 3, which means there were teachers who could answer only 3 items (out of 33) correctly, and the maximum score was 29, showing that none of the teachers could answer all the items correctly. The results revealed that the participant EFL teachers did not have a satisfactory level of LAK in spite of their claims in the needs assessment stage. The results reported in Table 2 helped answer the research question of the study about the extent of teachers' LAK. With low levels, as they claimed, EFL teachers need to improve their knowledge on the essential topics of language assessment.

Next, to explore the interrelationship of scores on different parts of the test and the total score, Pearson correlations were performed, which are reported in Table 3.

Table 3
Pearson Correlations between Different Parts of LAK Test

LAK Test Parts	Part A	Part B	Part C	Part D	Part E	Part F	Total
Part A	1	.15*	.30**	.20**	.23**	.25**	.55**
Part B		1	.20*	.23**	.12	.21**	.55**
Part C			1	.33**	.26**	.30**	.62**
Part D				1	.20*	.34**	.66**
Part E					1	.38**	.57**
Part F						1	.70**
Total							1

* Significant at .05 level
** Significant at .01 level

As it can be seen in Table 3, the correlation coefficients between different parts of the LAK test ranged from .12 to .38, all significant at the .05 probability level, except the correlation between Parts B and E. One possible reason for low correlations might have been due to the fact that each part of the test targets an aspect of the LAK construct that is different from the others. On the other hand, all parts of the test had high significant correlations (ranging from .55 to .70) with the total.

The next important step was checking the construct validity of the LAK test, which was done by running factor analysis (Brown, 2005). Preliminary exploratory factor analysis resulted in 14 underlying factors. The number of factors was too many for 33 items in the test and there were no clear patterns of item loadings. Therefore, confirmatory factor analysis was run with 6 factors corresponding to the theoretically presupposed factors. The results of the confirmatory factor analysis are shown in Table 4.

Table 4
Confirmatory Factor Analysis on LAK Test

LAK Test Parts	Items	Components 1	2	3	4	5	6
Part A	Item1				.58		
	Item2				.31	.41	-.37
	Item3				.55		-.44
	Item4	.41	.36				-.34
	Item5		.60				
Part B	Item6		.37				
	Item7		.51				
	Item8			.67			
	Item9			.79			
	Item10			.79			
Part C	Item11				.45		
	Item12	.35			.44		
	Item13				.66		
	Item14		.55				
Part D	Item15						.39
	Item16				.35		.52
	Item17	.35					
	Item18				.59		
	Item19		.44				.37
	Item20		.48				
	Item21	.42					
	Item22		.55				
Part E	Item23	.47	.37				
	Item24	.46					
	Item25	.30				.54	
	Item26	.33					
Part F	Item27					.58	
	Item28					.53	
	Item29	.58					
	Item30	.60					
	Item31					.45	
	Item32	.67					
	Item33						.49

Extraction Method: Principal Component Analysis
Rotation Method: Varimax

As observed in Table 4, items in Part A (test types/ functions) loaded differently on several factors, though most of the items loaded on factors 2, 4, and 6. Two of the items loaded on one factor whereas three of them loaded on two or three factors. This shows that in spite of their seeming similarity in terms of checking knowledge of test types, the items tap some other components of the LAK construct. This signals contradictions between the theoretical

assumptions of the model and the outcome of real data. On the other hand, there was less variation among the items comprising Part B (stages of language test design). Two of the items loaded on one factor, and three of them loaded strongly on another factor. The items in Part C (test characteristics) also showed a pattern similar to those in Part B; three loaded on one factor and one loaded on another factor. Part D (key terms/ concepts), like Part A, showed some variation in terms of the loading of the items comprising it. This was not surprising and even seemed natural as the items in this part checked a variety of terms and concepts, which were not very similar to each other even in theory. The best pattern of loadings was seen for items in Part E (alternative assessment techniques) as all loaded on only one factor. Even though two items in this section showed some loadings on two factors, they all loaded strongly on one main factor as well. This confirms the theoretical assumption that varieties of new assessment techniques have the similar underlying construct. Finally, the items in Part F (assessing language components and skills) loaded on three factors. The items checking assessing vocabulary, listening, and speaking loaded on one common factor; the items checking assessing grammar, pronunciation, and reading loaded on another common factor. However, the item focusing on assessing writing loaded on a separate factor. The loadings of vocabulary, listening, and speaking items on the same factor may be because assessing listening and speaking are related to each other as both are oral skills. It also indicates that vocabulary plays a significant role in the assessment of oral skills. The loadings of the items in grammar, pronunciation, and reading on the same factor may be due to the fact that assessing knowledge of grammar and pronunciation is crucial in the assessment of the reading skill. Further, the separate loading of the item on assessing writing may be due to the fact that the content of this item focused on rater reliability in assessing writing rather than how to assess writing, which is not a major issue in testing other components and skills except assessing speaking.

The last stage was to examine the possible relationships between teachers' LAK scores and certain variables such as gender, field of study, educational level, teaching context, and teaching experience that were assumed to influence their performance. Since teachers with higher academic degrees would perform better on a knowledge test, no comparison was made among teachers with different educational levels. In addition, because out of 164 participant teachers, 115 had a BA degree, this group was selected as the norm and the relationships between teachers' LAK scores and other factors were examined only for this main group.

The first comparison was made to see the difference between the scores of female vs. male BA teachers on the LAK test. The result of the independent samples t-test showed no significant difference. Next, to investigate the differences in the performance of the BA teachers with different university majors (i.e., TEFL, English literature, English translation, and other fields), a one-way ANOVA was used which showed no significant differences due to teachers' university majors, either.

Then, the difference in the performance of the BA teachers teaching at different contexts (public schools or private institutes) was examined. Table 5 shows the descriptive statistics and the independent samples t-test results on the LAK scores of these teachers. Contrary to the previous analyses, this analysis showed a significant difference between the performance of the two groups of teachers on the LAK test ($t = -2.18$).

Table 5
Descriptive Statistics of BA Teachers at Different Teaching Contexts on LAK Test

Teaching Context	N	Mean	SD	t	df	Sig. (2-tailed)
Public schools	25	17.60	6.10			
Private institutes	89	14.89	5.28	-2.18	112	.03*

As it can be seen in Table 5, EFL teachers working at public schools were more familiar with language assessment topics and had higher levels of LAK. One of the main reasons for this significant difference may be that EFL teachers at public schools in Iran are directly involved in developing tests and assessing their students' achievement and learning, whereas at private institutes in Iran, EFL teachers are not allowed to do so as testing/assessment is basically done through tests developed by supervisors or the testing staff. Consequently, teachers at private institutes do not feel the need to improve their LAK levels.

Finally, to check the relationship between teaching experience and LAK scores of BA teachers, a Pearson correlation was used which showed a significant but low value (r = .31). To recheck the relationship between teaching experience and LAK scores, the BA teachers were divided into two groups of experienced and novice teachers. The teachers whose experience was above the mean experience were considered experienced and those with experience below the mean were considered novice. No significant correlation coefficients were observed between teaching experience and LAK scores of either experienced or novice teachers reconfirming the low relationship between teaching experience and LAK.

The descriptive statistics of experienced vs. novice teachers' scores on the LAK test and the results of an independent samples t-test are presented in Table 6. Similar to the teaching context, this analysis also showed a significant difference between the performance of the two groups of teachers on the LAK test (t = -2.73).

Table 6.
Descriptive Statistics of BA Teachers having Different Teaching Experiences on LAK Test

Teaching Experience	N	Mean	SD	t	df	Sig. (2-tailed)
Experienced	32	17.63	5.45			
Novice	76	14.53	5.35	-2.73	106	.00*

The information in Tables 6 shows that teaching experience makes a difference on EFL teachers' levels of LAK, i.e., those having more experience have higher levels of LAK. It may be because during the teaching career, experienced teachers become more involved in doing assessment and they are more trusted by their supervisors or other authorities to develop tests or assess their students

Conclusion

The absence of an appropriate test to measure EFL teachers' LAK prompted the present study. Accordingly, a data-driven scenario-based LAK test was developed based on the information obtained from the needs assessment stage and the information in available language assessment textbooks to check EFL teachers' LAK levels. Contrary to the findings of the needs assessment stage where the majority of the EFL teachers claimed that they had the basic knowledge on issues related to language assessment, their poor performance on the LAK test showed that the majority of the participants had low levels of LAK. Further, the EFL teachers' low levels of LAK confirmed their need to improve their assessment knowledge on the essential topics of language assessment. The participants' performance on the LAK test also showed that there were factors such as gender and university major which did not have strong relationships to scores on LAK test. On the other hand, the context of teaching and teaching experience showed meaningful relationships with the LAK scores of BA teachers. The findings call for a reform in treating LAK in pre-service and in-service teacher education programs for EFL teachers. Although it might be challenging, there seems to be an urgent need to provide an ongoing instruction on various topics of language assessment to teachers (Malone, 2008, 2013).

It should be mentioned that the LAK test developed and validated in this study can be considered as a useful instrument for measuring the level of EFL teachers' LAK. A well-developed LAK test can help teachers to improve their assessment practices because the goal of an appropriate LAK test is to empower teachers to become better assessors and consequently better teachers (Stabler-Havener, 2018). Since this test was developed in multiple stages following a data-driven approach, it can act as a good preliminary model to develop similar data-driven tests measuring other aspects of EFL teachers' professional knowledge. Most of the available tests (e.g., NBC, Praxis tests, TKT) are developed on the basis of experts' opinions following a top-down approach with a little or no attention to the voice or needs of the involved parties, especially teachers. Performance on these tests may give useful information to policy makers and officials in designing appropriate pre-service and in-service teacher education programs for different groups of teachers. Moreover, based on teachers' performance, officials from both public and private sectors can employ EFL teachers who are professionally knowledgeable. Performance on these tests can also show the degree of professionalism of EFL teachers and may help the supervisors and coordinators to work more closely with teachers to improve the language education of the country. Teachers' performance on such tests can also help supervisors to identify the strengths and weaknesses in teachers' LAK and provide opportunities for them to improve themselves (Stabler-Havener, 2018).

Finally, even though teachers have a significant role in educational systems, and even though it is necessary to have knowledgeable and professional teachers, the role of policy makers cannot be overlooked. Research on developing and validating appropriate tests to investigate different aspects of teachers' professional knowledge is needed more than ever. The findings could inform policy makers about how to help teachers become more professional to enhance students' achievement. This is not possible only through investigating teachers' levels of professional knowledge or training them. It requires an ongoing and data-based training that would lead to the engagement of all stakeholders in the process of teachers' professional development.

References

Bachman, L.F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing, 17*(1), 1-42.

Bachman, L.F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Berry, V., Munro, S., & Sheehan, S. (2016). *Singing from the same hymn sheet? What language assessment literacy means to language teachers*. Paper presented at the international Language Assessment Literacy Symposium on Enhancing Language Assessment Literacy: Sharing, Broadening, Innovating (September 16-17). Lancaster University, Lancaster, UK.

British Council website (2018). https://www.britishcouncil.pt/en/exam/cambridge/which/tkt.

Brown, H.D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices (2nd ed.)*. Longman.

Brown, J.D. (2005). *Testing in language programs (2nd ed.)*. Prentice Hall.

Cambridge ESOL website (2018). http://www.cambridgeenglish.org/teaching-english/teaching-qualifications/tkt/.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, *8*(1), 1-44.

Downing, S.M. (2006). Twelve steps for effective test development. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development* (pp.3-25). New Jersey: Lawrence Erlbaum Associates, Publishers.

Eberharter, K., Kremmel, B., Holzknecht, F., & Konrad, E. (2016). *Fostering language assessment literacy through teacher involvement in high-stakes test development*. Paper presented at the international Language Assessment Literacy Symposium on Enhancing Language Assessment Literacy: Sharing, Broadening, Innovating (September 16-17). Lancaster University, Lancaster, UK.

Educational Testing Service (ETS) website (2018). https://www.ets.org/praxis.

Farhady, H., & Tavassoli, K. (2015). *EFL teachers' professional knowledge of assessment*. Paper presented at the 37th international LTRC Conference on From Language Testing to Language Assessment: Connecting Teaching, Learning, and Assessment (March 18-20). Toronto, Canada.

Farhady, H., & Tavassoli, K. (2017). *Developing a professional knowledge of language assessment test for EFL teachers.* Paper presented at the 39th international LTRC Conference on Language Assessment Literacy across Stakeholder Boundaries (July 19-21). Bogota, Colombia.

Farhady, H., & Tavassoli, K. (forthcoming). Assessment knowledge needs of EFL teachers. *TEL Journal.*

Farhady, H., Jafarpur, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice.* Tehran: SAMT Publications.

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

Hamp-Lyons, L. (2017). Language assessment literacy for language learning-oriented assessment. *Papers in Language Testing and Assessment, 6*(1), 88-111.

Harsch, C., Seyferth, S., & Brandt, A. (2017). *Developing assessment literacy in a dynamic collaborative project: What teachers, assessment coordinators, and assessment researchers can learn from and with each other.* Paper presented at the 39th international LTRC Conference on Language Assessment Literacy across Stakeholder Boundaries (July 19-21). Bogota, Colombia.

He, L., & Min, S. (2017). Development and validation of a computer adaptive EFL test. *Language Assessment Quarterly, 14*(2), 160-176.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402.

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing, 30*(3), 301-307.

Janatifar, M., Marandi, S.S., & Babaei, E. (2017). *Defining EFL teachers' language assessment literacy and its promotion through virtual learning teams.* Paper presented at the 39th international LTRC Conference on Language Assessment Literacy across Stakeholder Boundaries (July 19-21). Bogota, Colombia.

Jin, Y., & Jie, W. (2017). *Do workshops really work? Evaluating the effectiveness of training in language assessment literacy.* Paper presented at the 39th international LTRC Conference on Language Assessment Literacy across Stakeholder Boundaries (July 19-21). Bogota, Colombia.

Khany, R., & Azimi Amoli, F. (2016). Validating an English language teacher professional development scale in Iranian EFL context. *Issues in Language Teaching, 5*(1), 107-134.

Kim, H., & Elder, C. (2015). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing, 32*(2), 129–149.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197.

Lane, S., Raymond, M.R., Haladyna, T.M., & Downing, S.M. (2016). Test development process. In S. Lane, M.R. Raymond, & T.M. Haladyna (Eds.), *Handbook of test development* (pp.3-18). New York and London: Routledge.

Malone, M.E. (2008). Training in language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed.), Vol.7: Language testing and assessment* (pp.225-239). New York: Springer Science and Business Media.

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

National Board for Professional Teaching Standards website (2018). http://www.nbpts.org/national-board-certification/.

Pallant, J. (2005). *SPSS survival manual.* Australia: Allen & Unwin.

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

Popham, W.J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265-273.

Richards, J.C. (2001). *Curriculum development in language teaching.* Cambridge: Cambridge University Press.

Sellan, R. (2017). Developing assessment literacy in Singapore: How teachers broaden English language learning by expanding assessment constructs. *Papers in Language Testing and Assessment, 6*(1), 64-87.

Stabler-Havener, M.L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL, 18*(1), 1-22.

Stiggins, R.J. (2005). *Assessment for learning defined.* Paper presented at the ETS Assessment Training Institute's International Conference on Promoting Sound Assessment in Every Classroom (September). Portland, USA.

Stiggins, R.J. (2008). *Assessment for learning, the achievement gap, and truly effective schools.* Paper presented at the Educational Testing Service and College Board Conference on Educational Testing in America: State Assessments, Achievement Gaps, National Policy and Innovations (September 8). Washington, DC, USA.

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.

Tsagari, D., & Vogt, K. (2015). *Assessment literacy of foreign language teachers: Research, challenges and future prospects*. Paper presented at the 37th international LTRC Conference on From Language Testing to Language Assessment: Connecting Teaching, Learning, and Assessment (March 18-20). Toronto, Canada.

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*(1), 41-63.

Wallace, M.J. (1991). *Training foreign language teachers: A reflective approach*. Cambridge: Cambridge University Press.

Wright, S.P., Horn, P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*, 57-67.

Xu, Y. (2015). *Language assessment literacy in practice: A case study of a Chinese university English teacher*. Paper presented at the 37th international LTRC Conference on From Language Testing to Language Assessment: Connecting Teaching, Learning, and Assessment (March 18-20). Toronto, Canada.

Xu, Y., & Brown, G.T.L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment, 6*(1), 133-158.

Hossein Farhady is an associate professor of applied linguistics at Yeditepe University in Istanbul, Turkey. He received his MA in TESL and PhD in Applied Linguistics from University of California at Los Angeles in 1978 and 1980, respectively. His major interest is research on various aspects of language assessment. He has taught courses on language testing and research methods at the MA and PhD levels for the last four decades in several countries and presented papers in professional conferences. He has widely published on language testing, directed national and international research projects, and supervised over 50 MA theses and PhD dissertations.

Kobra Tavassoli is an assistant professor of TEFL at Islamic Azad University, Karaj, Iran. Right now, she is the head of the ELT Department at Islamic Azad University, Karaj. She is teaching language assessment, research methodology, and language teaching methodology courses at BA, MA, and PhD levels. Her areas of interest are language assessment and teacher education, and she has presented in national and international conferences on these topics.