



IMAM KHOMEINI  
INTERNATIONAL UNIVERSITY



Print ISSN: 2676-5357  
Online ISSN: 2676-5965

## The Impact of an In-Service Professional Development Course on Teachers' Language Assessment Knowledge

Kaveh Jalilzadeh<sup>1</sup>, Adel Dastgoshadeh<sup>2</sup>, Rahele Khosravi<sup>3</sup>

<sup>1</sup>Department of Foreign Languages, Istanbul University, Cerrahpaşa, Turkey  
(Corresponding author)  
*Kaveh.j@iuc.edu.tr*

<sup>2</sup>Department of English, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran  
*Adastgoshadeh@gmail.com*

<sup>3</sup>MA in TEFL, Kurdistan University, Sanandaj, Iran  
*rahelekhosravi52@gmail.com*

---

Article info    Abstract

---

Article type:    **Abstract**  
Research  
article

Received:  
08/05/ 2023  
Accepted:  
26/08/2023

With recent advancements, teachers are required to be able to implement effective assessments in the classroom context. This is applicable through in-service professional development programs that assist teachers' improvement and provide them with opportunities to enhance their knowledge and to use what they know practically, especially their knowledge regarding assessment and its concepts. Therefore, the present study, which was a quasi-experimental study, was designed to explore the effect of an in-service professional course on teachers' language assessment knowledge (LAK). To this end, a number of 76 EFL teachers, working at private institutes in Sanandaj, Iran, were selected through convenience sampling on the basis of availability to fill in the Language Assessment Test developed by Farhady and Tavassoli (2018) in order to estimate their current level of LAK. Then, a three-day workshop was held for the participants on the components and skills of language assessment. The instruction of each session took 6 hours. After 6 weeks, teachers' ALK was tested again after the treatment. The results of the paired T-test revealed that the in-service professional development course had a significant effect on teachers' LAK. It was also shown that teachers' knowledge of different concepts and key terms in assessment was enhanced. Further, various steps of test design were introduced to them. Also, based on the results it was claimed that teachers would be able to design and familiarize their learners with standard tests. Additionally, they would be capable of using alternative assessment techniques.

**Keywords:** *Assessment, Language Assessment Knowledge, Professional Development Course*

---

**Cite this article:** Jalilzadeh, K., Dastgoshadeh, A., & Khosravi, R. (2024). The impact of an in-service professional development course on teachers' language assessment knowledge. *Journal of Modern Research in English Language Studies*, 11(2), 131-153.

DOI: 10.30479/jmrels.2023.18114.2152

©2024 by the authors. Published by Imam Khomeini International University. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0>



## 1. Introduction

In the domain of language testing, a transition from a psychometric approach, which emphasizes standardized testing and reductionist measures, to an Edu metric approach that leans towards a more holistic perspective is observed. According to Wolf et al.(1991), the testing culture primarily focuses on using tests and exams to determine grades and achievements, whereas the assessment culture prioritizes using assessments to enhance instruction and facilitate student learning. Sheppard (2013) further highlights the global emphasis on new and creative types of assessment methods, including criterion and/or performance-based assessments, which has raised concerns among teachers about the underlying intentions and motivations behind these approaches.

Language testing and assessment serve a fundamental function in second language acquisition, as evidenced by research showing that teachers typically dedicate at least one-third of their teaching time to activities that are related to assessment (Bachman, 2014). The quality of assessments used in classrooms is closely linked to the provision of effective instruction and student learning. Therefore, as Earl (2013) pointed out, a crucial responsibility of educators is to combine testing with teaching and student learning, aligning them with the objectives of preparing students for lifelong learning skills and meeting the demands of the twenty-first century.

As a solution to address this issue, a specific term called Language Assessment Literacy (LAL) has been introduced in the field of language assessment and learning, particularly focusing on classroom settings (Inbar-Lourie, 2008a; Taylor, 2009). Stiggins (1999) states that “assessment literacy” is a term widely employed to describe the set of skills and knowledge that stakeholders require to navigate the evolving landscape of assessment. However, there is a lack of consensus on the specific components that constitute ‘assessment literacy,’ despite various recommended approaches aimed at promoting its development (e.g., Walters, 2010).

LAL encompasses the acquiring of knowledge, skills, and principles related to the way the construction of tests are done, interpretation, use, and evaluation, along with a critical understanding of the role of assessment in the broader context of education (O’Loughlin, 2013). While it is widely recognized that teachers need to assess students’ progress, many educators lack sufficient knowledge of the fundamentals of assessment (Popham, 2009). The competency dimensions of LAL can be classified into those related to the ‘what’ of language testing and assessment (the construct) and those related to the ‘how’ of language testing and assessment (the method) (Shohamy, 2001).

The crucial point to understand is that the aspects of ‘what’ and ‘how’ in language testing and assessment cannot be isolated from the underlying ‘why’ of testing and assessment. In other words, comprehending the ‘what’ and executing the ‘how’ requires an understanding of the reasons and rationale behind these actions, represented by the ‘why’. Each of these dimensions is influenced by language-related factors and the broader educational and assessment cultures. Davidson (2007) emphasizes that any discussion on language assessment literacy should consider the current developments in assessment, particularly the increasing support for ‘assessment for learning’ (AfL) approaches in various regions worldwide.

Recently, several authors have introduced the term ‘assessment literacy’ to define the knowledge that language teachers and instructors should possess concerning assessment topics (Stoynoff & Chapelle, 2005). It is evident that teacher training should not only concentrate on equipping pre-service teachers with effective teaching methods but also ensure that they understand various types of assessment, especially if they are engaged in creating, scoring, and enhancing classroom-based assessments.

We formulated the following two research questions to address the gap in the literature:

1. Does an in-service professional development course have any significant effect on the improvement of language assessment knowledge of Iranian EFL teachers?
2. Does an in-service professional development course have any significant effect on the improvement of Iranian EFL teachers’ knowledge of components of language assessment?

## **2. Literature Review**

### **2.1. Assessment Literacy**

The concept of assessment literacy was initially introduced in the context of general education and later adopted in language education, specifically concerning assessment and measurement practices in educational contexts. However, assessment is perceived to have flourished over time due to researchers’ growing interest in exploring this concept further. Fulcher (2012), Taylor (2009), and Walters (2010) claim that despite this exploration, there is no single agreed-upon definition for assessment literacy.

The term ‘assessment literacy’ was first proposed by Stiggins (1991a, 1995) in general education. He used it to characterize the idea that those classroom teachers

could pinpoint the differences between useful and unuseful assessment practices. Stiggins (1999) explains that ‘assessment literacy’ has become a common term for referring to the set of skills and knowledge that stakeholders need to guide the new world of assessment. Despite the increasing variety of recommended approaches to foster assessment literacy’s development (e.g., Walters, 2010), a consensus on its precise components remains elusive.

Inbar-Lourie (2008a) referred to assessment literacy as the capability of teachers to perceive the importance of social dimensions of assessment and the association between language knowledge and assessment task types. Furthermore, Language Assessment Literacy (LAL) is defined as acquiring knowledge, skills, and principles in test construction, interpretation, use, and evaluation, along with developing a critical perspective on assessment’s role in the broader educational context (O’Loughlin, 2013). While it is widely acknowledged that teachers need to assess students’ progress, many lack sufficient knowledge of assessment fundamentals (Popham, 2009). LAL competency involves dimensions related to the ‘what’ of language testing and assessment (the construct) and the ‘how’ of language testing and assessment (the method) (Shohamy, 2001).

Inbar-Lourie (2008b) emphasized that language assessment knowledge is not a singular concept but rather a combination of assessment literacy skills and language-specific competencies, creating a unique entity known as “language assessment literacy” (p. 389). Similarly, Malone (2013) referred to assessment literacy as language educators’ degree of acquaintance with terminologies in language testing and their ability to apply this knowledge to classroom practices, particularly in relation to language assessment.

Popham (2009) discusses the significance of assessment literacy for teachers’ professional growth, considering it an essential objective for their development. On the other hand, McMillan (2014) points out that traditional or objective assessment primarily assesses lower levels of learners’ thinking ability, while new alternatives in language assessment focus on evaluating higher-order thinking abilities. This alternative assessment is often referred to as authentic assessment, defined by Bachman and Palmer (1996) as tasks that relate to students’ everyday lives. Newmann and Archbald (1992) expand on the concept of authenticity by explaining that assessment quality and usefulness depend on the extent to which outcomes reflect meaningful and appropriate achievements.

The primary aspect of teachers’ classroom assessment literacy is their knowledge base concerning assessment. According to Bandura et al. (1999) and Fishbein and Ajzen (2011), the understanding and expertise individual teachers possess in this area are recognized as crucial factors that benefit the effectiveness of assessment practices. Farhady and Tavassoli (2018) emphasized

the significance of Language Assessment Knowledge (LAK) for EFL teachers, as it aids in developing appropriate assessments, making informed decisions about students' development and performance, and ultimately enhancing their professional achievements. Moreover, Boud (2006), Joughin (2009), and Earl (2013) in their studies have indicated that there is a strong correlation between the quality of assessment and instruction and learners' achievements.

Consequently, it is of main importance for language educators to understand the term assessment comprehensively and use it effectively in their teaching contexts. Experts in the field have asserted that the depth of teachers' assessment knowledge base directly influences their ability to implement assessments that improve instruction and promote student learning (Stiggins, 1991a, 1995; Popham, 2006, 2009). Price et al. (2012) add, in order for language educators to achieve the intricacies of the process of assessment in the classroom setting, possessing a comprehensive knowledge of language assessment is crucially needed. It also helps the educators in understanding the whole assessment procedure as well as making the best decisions about what skills and abilities to be measured and most importantly assists them in choosing and applying tasks for assessment of the learners' performance.

Also, Popham (2006, 2009) and Stiggins (1991a, 1995) believe that teachers' knowledge of assessment enables them to interpret the collected data from the process of assessment meticulously and effectively address any potential biases that might arise in the assessments they create or choose.

This issue has persisted for more than five decades that there is a scarcity of studies regarding assessment knowledge among teachers across various schools in numerous countries worldwide (Davidheiser, 2013; Gotch & French, 2013; Mayo, 1967; Plake, 1993). The following parts will be devoted to several studies that have utilized self-reported assessment and specific instruments that were utilized to measure the level of school educators' assessment literacy.

It is mentioned in Ajzen (1991, 2005) that individuals who lack the essential information and abilities are unable to perform their intended tasks successfully. In this context, if a teacher lacks the required knowledge and skills to carry out assessments effectively, it is highly probable that they will struggle to conduct the intended assessments or produce assessments of low quality (Bandura, 1989).

Harlen (2005) and Alkharusi et al. (2012) and other numerous studies have been conducted to explore the relationship between language educators' practical assessment and their knowledge of assessment. One such study by Tavassoli and Farhady (2018) focused on crucial skills that language educators must obtain

in order to possess comprehensible Language Assessment Knowledge (LAK). The objective of this research was to investigate the LAK requirements of EFL teachers. A total of 246 EFL teachers participated in the study and completed Fulcher's needs assessment questionnaire (2012), which aimed to gauge their perceptions regarding the significance of major aspects of language assessment and assess their own level of knowledge. The findings revealed that specific subjects related to language assessment are crucial components that should be included in an assessment course.

Esafandari and Nouri (2016) conducted a mixed-method study focusing on the Assessment Literacy (AL) of Iranian University Instructors. They emphasized the significance of teachers being assessment literate to improve their ability to assess students effectively. The study involved two groups of instructors, namely English Language Instructors and Content Instructors, and aimed to examine the disparities in their assessment literacy. Three main components for assessment knowledge were introduced in the results as follows: the theoretical aspects of assessing, analyzing, and constructing tests, as well as, the essential information about the statistical process.

Moreover, Mertler (2000) surveyed teachers from various school levels and subject areas to assess their assessment practices. The findings revealed that teachers with limited knowledge and understanding of validity and reliability concepts used statistical analysis less than those with a higher level of assessment knowledge. In the same line of inquiry, a study was conducted by Black et al. (2010) on 12 school language educators in Oxford Shire, UK, in order to estimate their level of understanding of some features of assessment such as validity in various contexts, for instance, summative assessments and the impact that it might have on the steps that they carry out practical assessment. Therefore, it was concluded that studies educators with insufficient knowledge of assessment are unable to apply assessment practically well and their practice will be affected negatively.

To address this gap in literature it seems necessary for school and institute authorities to hold regular educational courses to compensate for the breakdown in knowledge and understanding of teachers especially their knowledge regarding the concept of assessment and its components (Ölmezer-Öztürk & Aydın, 2019). Additionally, Mohammadi and Babaii (2022) in their study marked the importance of such professional development programs in which they measured the participants' knowledge of dynamic assessment before and after the educational program which focused on improving teachers' information about dynamic assessment. The findings showed that the participants found those training sessions both practical and informative and their knowledge about dynamic assessment increased significantly after the course.

## 2.2. Components of Assessment Literacy

Assessment literacy encompasses the essential knowledge and skills that teachers need to fulfill two main aspects:

First, identifying, selecting, or creating assessments that are well-suited for various purposes, including accountability, evaluating instructional programs, monitoring student growth and progress, promoting students, and diagnosing individual learning needs or gaps. This involves the ability to design assessments that align with specific objectives and are appropriate for the intended use.

Second, analyzing, evaluating, and effectively utilizing both quantitative and qualitative evidence derived from different types of assessments, including external summative and interim assessments, classroom summative assessments, and formative assessments embedded in instruction. These assessment results are used to make informed decisions aimed at enhancing educational programs and tailoring instructional approaches to promote student learning and achievement. A strong understanding of test quality considerations and comparability issues is essential for making appropriate decisions based on assessment data.

The goal of assessment literacy is to empower teachers to make well-informed decisions about assessment practices, ultimately leading to improved student learning outcomes. It involves the ability to choose appropriate assessments for specific purposes, interpret and utilize assessment results effectively, and apply this knowledge to enhance instructional strategies and educational programs. By being assessment-literate, teachers can ensure that assessment becomes a valuable tool in fostering student growth and success in the classroom. (Kahl et al., 2013).

## 3. Method

### 3.1. Participants

A pool of 90 EFL teachers who had degrees in TEFL were chosen as the participants of this study through convenience sampling, the most widespread non-random type of sampling in EFL studies (Dornyei, 2007), which is defined as “the selection of individuals who happen to be available for study” (Mackey & Gass, 2016, p. 122). The participants were teachers who were teaching the English language in private institutes in Sanandaj, Iran. The participants’ age range varied from 20 to 32 including both genders 50 females and 40 males. Finally, a pool of 76 EFL teachers participated in the present study and 14 participants were excluded from the study due to mortality. The educators who were the subjects of this study had different teaching experiences (5 to 15 years of experience) at various ages (6 to 28) and levels (basic to advanced) of students. There was an assumption that various demographic information of the participants did not have

any specific impact on their LAK (Ölmezer-Öztürk & Aydin, 2019). Consequently, the researchers did not intend to control any of the demographic features of the participants such as gender, age, etc.

### **3.2. Materials and Instruments**

In order to collect more accurate results, the participants of the current study were asked to complete the Language Assessment Knowledge test, which was developed by Farhady and Tavassoli (2018).

#### ***3.2.1. Language Assessment Knowledge Test***

In this research, the Language Assessment Knowledge (LAK) test developed by Farhady and Tavassoli (2018) was utilized to measure the educators' knowledge of assessment. The test consisted of 33 items covering six main topics: Part A, focused on test types/functions (matching items); Part B, covered the stages of language test design (ordering items); Part C, dealt with test characteristics (matching items); Part D, assessed key terms/concepts (multiple-choice items); Part E, explored alternative assessment techniques (matching items); and Part F, involved assessing language components and skills (multiple-choice items). The test's reliability was determined using Cronbach's Alpha.

### **3.3. Procedure**

Before starting the in-service program, the Language Assessment Knowledge test was given to the participants as the pretest of the study in order to estimate their assessment knowledge. In this phase, teachers took the Language Assessment Knowledge test designed by Farhady and Tavassoli (2018). This test helped the researchers to have a thorough perspective regarding the participants' knowledge of language assessment due to the fact that the components that were measured by the given test are considered as important parts of the workshop. Then, in the treatment phase, the researchers held a 3-day workshop. In so doing, the researchers, first of all, tried to change the belief of the teachers regarding testing culture and directed them to assessment culture, and based on their experience in teaching at private institutes, they designed a straightforward syllabus for 3 days during which they introduced new concepts in assessment and provided the teachers with informative sources and then, they gave the teachers practical samples and examples regarding alternative assessment techniques. The instruction of each session took 6 hours. During the first session of the workshop, researchers elaborated on test characteristics and key terms and concepts in language testing. On the second day, the participants got familiar with different test types as well as their functions and they also learned about different stages of a language test design. Furthermore, they gained experience about what types of tasks to use in their tests. In the last session of the workshop, they were taught alternative



assessment techniques. Additionally, statistical analysis was utilized to test the reliability and validity of the test as well as the characteristics of the items which were designed by them as well as the ways of assessing language components and skills for example how to grade speaking and writing and how to use the results gathered from their assessment in order to make proper decision about their students' achievements. In the end, as the post-test of the study, the teachers were requested to complete the Language Assessment Knowledge test again after 5 weeks.

### 3.4. Data Analysis

In this phase of the study, with the purpose of examining the teachers' knowledge of assessment prior to and following the treatment, the data collected from a language assessment knowledge test administered to obtain accurate information about the EFL teachers' assessment knowledge were analyzed.

## 4. Results and Discussion

### 4.1. Results

The present study was undertaken in order to estimate the effect of professional development on the development of language assessment knowledge of Iranian EFL teachers. Paired-samples t-test and repeated measures ANOVA were utilized to analyze the data. Table 1 displays the skewness and kurtosis statistics for the pretest and posttest of LAK and their components. It is believed that in order to achieve the normality of data, the skewness and kurtosis of the data should be lower than +/- 2 (Bachman, 2005; Bae & Bachman, 2010).

**Table 1**  
*Descriptive Statistics; Testing Normality of Data*

|                     | N  | Skewness  |            | Kurtosis  |            |
|---------------------|----|-----------|------------|-----------|------------|
|                     |    | Statistic | Std. Error | Statistic | Std. Error |
| Pre-Type/Function   | 76 | .127      | .276       | -.119     | .545       |
| Pre-Design          | 76 | -.128     | .276       | -.321     | .545       |
| Pre-Characteristics | 76 | -.203     | .276       | -.211     | .545       |
| Pre-Concepts        | 76 | .383      | .276       | .136      | .545       |
| Pre-Alternative     | 76 | -.186     | .276       | -.467     | .545       |
| Pre-Skills          | 76 | -.396     | .276       | .117      | .545       |
| Post-Type/Function  | 76 | -.301     | .276       | -.587     | .545       |

|                      |    |       |      |       |      |
|----------------------|----|-------|------|-------|------|
| Post-Design          | 76 | -.652 | .276 | -.305 | .545 |
| Post-Characteristics | 76 | .157  | .276 | -.955 | .545 |
| Post-Concepts        | 76 | -.279 | .276 | -.359 | .545 |
| Post-Alternative     | 76 | -.446 | .276 | -.232 | .545 |
| Post-Skills          | 76 | -.230 | .276 | -.619 | .545 |
| Pretest              | 76 | -.448 | .276 | .121  | .545 |
| Posttest             | 76 | -.530 | .276 | .155  | .545 |

## 4.2. Exploring the First Research Question

*Does an in-service professional development course have any significant effect on the improvement of language assessment knowledge of Iranian EFL teachers?*

With the aim of comparing the Iranian EFL teachers' means on the pretest and posttest of LAK a paired-samples t-test was used to which was to probe the first research question. Based on the results displayed in Table 2 it can be claimed that the EFL educators' mean on the posttest of LAK was higher than their mean on the pretest (M = 22.67, SD = 5.65) (M = 13.61, SD = 4.41).

**Table 2**

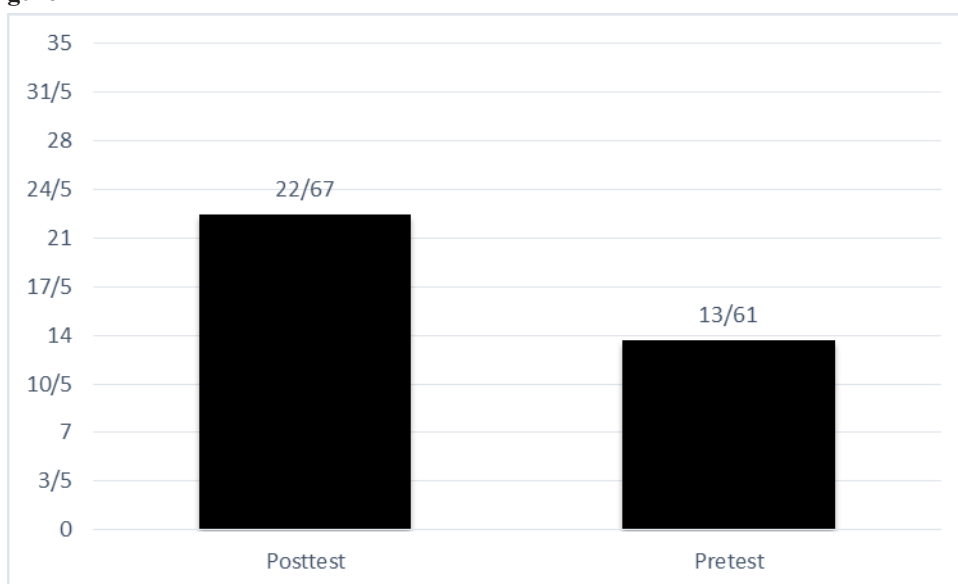
*Descriptive Statistics; Pretest and Posttest of Language Assessment Knowledge*

|          | Mean  | N  | Std. Deviation | Std. Error Mean |
|----------|-------|----|----------------|-----------------|
| Posttest | 22.67 | 76 | 5.651          | .648            |
| Pretest  | 13.61 | 76 | 4.412          | .506            |

Paired-samples t-test results show a large effect size ( $t(75) = 22.78$ ,  $p = .000$ ,  $r = .935$ ) (Table 3) that, in turn, indicates that EFL teachers had a significantly higher mean on the post-test of LAK than the pretest. As a result, the hypothesis stating "professional development did not have any significant effect on the improvement of language assessment knowledge of Iranian EFL teachers" was rejected.

**Table 3***Paired-Samples t-test; Pretest and Posttest of Language Assessment Knowledge*

| Paired Differences |                |                 |   |       |        |    |                 |
|--------------------|----------------|-----------------|---|-------|--------|----|-----------------|
| Mean               | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference |       | T      | f  | Sig. (2-tailed) |
|                    |                |                 | Lower                                     | Upper |        |    |                 |
| 9.066              | 3.469          | .398            | 8.273                                     | 9.859 | 22.781 | 75 | .000            |

**Figure 1***Means on the Pretest and Posttest of Language Assessment Knowledge*

### 4.3. Exploring the Second Research Question

*Does an in-service professional development course have any significant effect on the improvement of Iranian EFL teachers' knowledge on components of language assessment?*

In this part of the study in order to make the comparison between the Iranian EFL teachers' improvement of means from pretests to posttests of components of LAK, a repeated measures ANOVA as well as simple effect analysis were used. Since the present study did not include any independent variables, there was no need to probe the assumption of homogeneity of variances and the assumption of homogeneity of covariance matrices. Table 4 displays the results of Mauchly's sphericity tests. The sphericity assumptions requires that the differences between any two means should have roughly equal variances. As

noted by Field (2018), there should be at least three tests to probe the sphericity assumption. That was why the probability was not computed for the effect of Time because the components of LAK were measured at two-time intervals of pretest and posttest. The other findings collected from Mauchly's tests were important that showed the violation of the assumption of sphericity. However, there is no need to worry about the violation of these assumptions because the repeated measures procedures computed four F-values (Table 6). As noted by Field (2018, p 847), "If sphericity is violated then you can apply a correction proportionate to the extent of the violation (for example, the Greenhouse–Geisser correction) so you may as well ignore Mauchly's test and always apply the correction". Since the assumption of sphericity was not met, the results of the Greenhouse–Geisser correction (Table 6) were reported.

**Table 4**

*Mauchly's Test of Sphericity*

| Within Subjects Effect | Mauchly's W | Approx. Chi-Square | df | Sig. | Epsilon <sup>b</sup> |             |             |
|------------------------|-------------|--------------------|----|------|----------------------|-------------|-------------|
|                        |             |                    |    |      | Greenhouse-Geisser   | Huynh-Feldt | Lower-bound |
| Time                   | 1.000       | .000               | 0  | .000 | 1.000                | 1.000       | 1.000       |
| Tests                  | .466        | 55.778             | 14 | .000 | .757                 | .802        | .200        |
| Time * Tests           | .598        | 37.569             | 14 | .001 | .827                 | .882        | .200        |

Table 5 displays the Iranian EFL teachers' means on pretests and posttests of components of LAK. The results indicated that the participants, following the professional development course, showed improvement in their means on post-tests.

**Table 5**  
*Descriptive Statistics; Pretests and Posttests of Components of Language Assessment Knowledge*

| Time     | Tests                  | Mean  | Std. Error | 95% Confidence Interval |             |
|----------|------------------------|-------|------------|-------------------------|-------------|
|          |                        |       |            | Lower Bound             | Upper Bound |
| Pretest  | Pre - Type / Function  | 2.000 | .099       | 1.803                   | 2.197       |
|          | Pre-Design             | 2.066 | .098       | 1.871                   | 2.261       |
|          | Pre - Characteristics  | 1.789 | .086       | 1.617                   | 1.962       |
|          | Pre-Concepts           | 3.079 | .155       | 2.770                   | 3.388       |
|          | Pre-Alternative        | 1.684 | .096       | 1.493                   | 1.875       |
|          | Pre-Skills             | 2.987 | .143       | 2.701                   | 3.272       |
| posttest | Post - Type / Function | 3.395 | .139       | 3.118                   | 3.672       |
|          | Post-Design            | 3.750 | .129       | 3.494                   | 4.006       |
|          | Post - Characteristics | 2.697 | .101       | 2.496                   | 2.898       |
|          | Post-Concepts          | 5.329 | .202       | 4.926                   | 5.732       |
|          | Post-Alternative       | 2.803 | .109       | 2.585                   | 3.020       |
|          | Post-Skills            | 4.697 | .185       | 4.328                   | 5.067       |

According to Table 6, a significant difference was observed between the means of the pretest and posttest disregarding their components ( $F_{(1, 75)} = 518.98$ ,  $p = .000$ , partial  $\eta^2 = .874$  indicating a large effect size). Significant differences were observed between the overall means on six components of LAK disregarding time interval; i.e., pretests and posttests ( $F_{(5, 71)} = 92.52$ ,  $p = .000$ , partial  $\eta^2 = .867$  representing a large effect size). Statistically significant relationships were spotted between time and components of LAK ( $F_{(5, 71)} = 8.83$ ,  $p = .000$ , partial  $\eta^2 = .384$  representing a large effect size).

**Table 6***Multivariate Tests; Pretests and Posttests of Components of Language Assessment Knowledge*

|              | Effect             | Value | F       | Hypothesis<br>df | Error<br>df | Sig. | Partial<br>Eta Squared |
|--------------|--------------------|-------|---------|------------------|-------------|------|------------------------|
| Time         | Pillai's Trace     | .874  | 518.987 | 1                | 75          | .000 | .874                   |
|              | Wilks' Lambda      | .126  | 518.987 | 1                | 75          | .000 | .874                   |
|              | Hotelling's Trace  | 6.920 | 518.987 | 1                | 75          | .000 | .874                   |
|              | Roy's Largest Root | 6.920 | 518.987 | 1                | 75          | .000 | .874                   |
| Tests        | Pillai's Trace     | .867  | 92.528  | 5                | 71          | .000 | .867                   |
|              | Wilks' Lambda      | .133  | 92.528  | 5                | 71          | .000 | .867                   |
|              | Hotelling's Trace  | 6.516 | 92.528  | 5                | 71          | .000 | .867                   |
|              | Roy's Largest Root | 6.516 | 92.528  | 5                | 71          | .000 | .867                   |
| Time * Tests | Pillai's Trace     | .384  | 8.836   | 5                | 71          | .000 | .384                   |
|              | Wilks' Lambda      | .616  | 8.836   | 5                | 71          | .000 | .384                   |
|              | Hotelling's Trace  | .622  | 8.836   | 5                | 71          | .000 | .384                   |
|              | Roy's Largest Root | .622  | 8.836   | 5                | 71          | .000 | .384                   |

Table 7 displays the findings of simple effects analysis which enabled the researchers to compare the improvement of means from pretest to posttest for each of the components of LAK. Based on these results and the descriptive statistics displayed in Table 5 it can be concluded that EFL teachers showed significant improvement in their means from;

- Pretest (M = 2) to posttest (M = 3.39) of test types and functions (Mean Difference = 1.39,  $p = .000$ ).
- Pretest (M = 2.06) to posttest (M = 3.75) of stage of test design (Mean Difference = 1.68,  $p = .000$ ).
- Pretest (M = 1.78) to posttest (M = 2.69) of test characteristics (Mean Difference = .908,  $p = .000$ ).
- Pretest (M = 3.07) to posttest (M = 5.32) of key terms and concepts (Mean Difference = 2.25,  $p = .000$ ).
- Pretest (M = 1.68) to posttest (M = 2.80) of alternative test

techniques (Mean Difference = 1.11,  $p = .000$ ).

- Pretest ( $M = 2.98$ ) to posttest ( $M = 4.69$ ) of assessment of language components and skills (Mean Difference = 1.71,  $p = .000$ ).

**Table 7**

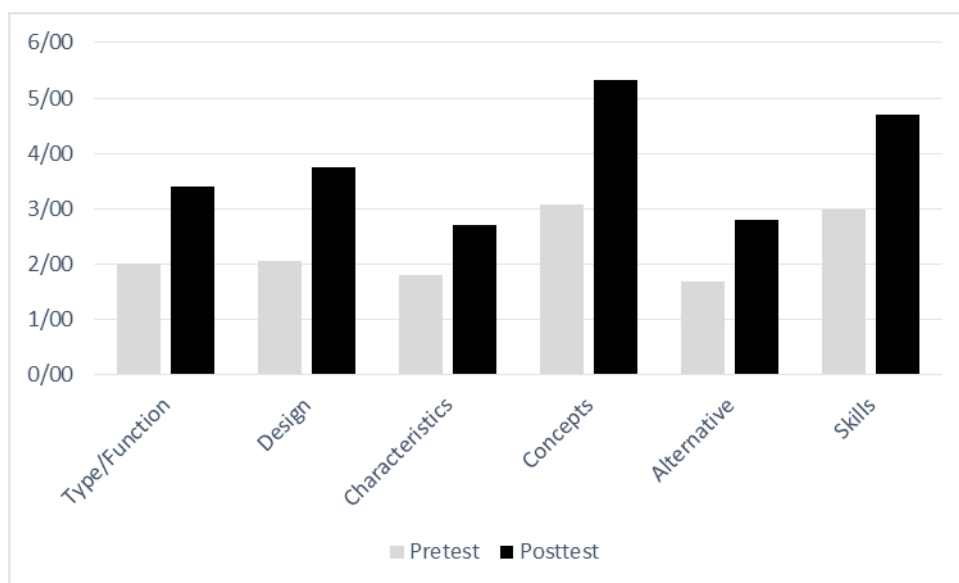
*Pairwise Comparisons; Pretests and Posttests of Components of Language Assessment Knowledge*

|                 | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval for Difference |             |
|-----------------|-----------------------|------------|------|--|-------------|
|                 |                       |            |      | Lower Bound                            | Upper Bound |
| Test Types      | 1.395*                | .131       | .000 | 1.133                                  | 1.656       |
| Test Design     | 1.684*                | .126       | .000 | 1.433                                  | 1.935       |
| Characteristics | .908*                 | .103       | .000 | .703                                   | 1.113       |
| Concepts        | 2.250*                | .190       | .000 | 1.871                                  | 2.629       |
| Alternative     | 1.118*                | .108       | .000 | .904                                   | 1.333       |
| Skills          | 1.711*                | .175       | .000 | 1.363                                  | 2.058       |

\*. The mean difference is significant at the .05 level.

**Figure 2**

*Means on Components of Pretests and Posttests of Language Assessment Knowledge*



#### **4.4. Discussion**

This study was a report on the effect of the professional development of EFL teachers on their Language Assessment Knowledge (LAK) as well as their knowledge of the assessment components. The findings of data analysis showed that professional development had a significant effect on the betterment of language assessment knowledge of Iranian EFL teachers. Also, it is indicated by Farhady and Tavassoli (2018) and Popham (2011) that the higher the level of LAK of the teachers the more student's achievements will result in the classes as assessment knowledge plays an important role in teachers' professional knowledge.

Secondly, the participants, following the professional development course, showed improvement in their means on post-tests of components of LAK. The tests showed that the EFL teachers had more information about terms and concepts of assessment but had little information about test designs, functions, and characteristics as well as alternative test techniques. However, the post-test shows that they had significant progress in their knowledge of assessment components which was also in parallel with a study conducted by Jannati (2015) in which the instructors were aware of the basic topics in language assessment. Even though the teachers had enough knowledge in assessment, they weren't able to use this knowledge in practice. Therefore, the instructors needed a training course in order to learn how to use their knowledge in their classroom and teaching practice (Ölmezer-Öztürk & Aydin, 2019; Popham, 2009).

These findings are in line with some other previous studies such as Melter (2003), Esfandiari and Nouri (2016), Heng, Abdullah and Rashid (2016), Elshawa et al. (2016), Farhady and Tavassoli (2018). For example, Farhady and Tavassoli's (2018) study focused on EFL teachers' language assessment knowledge (LAK) improvement and their LAK needs. They found that a large number of EFL teachers believed that they needed to augment their knowledge of assessment to improve their professional development. Moreover, Mertler (2003) in his study focused on the difference between the knowledge of assessment of pre-service and in-service educators. The results of his study revealed that on one hand, both teachers in the in-service and pre-service programs performed highest on some components of assessment literacy. On the other hand, they had the lowest performance on one component of assessment knowledge which was developing valid grading procedures.

Therefore, it is important for a successful teaching program to provide educators with various courses to develop their LAK (Popham, 2009) as it is claimed by Ölmezer-Öztürk and Aydin, 2019, the major cause of inadequate knowledge of EFL educators regarding LAK results from the inadequacy of



teacher education and professional development courses. These professional development programs should be presented to augment the teachers' skills in measuring language abilities. However, mere attendance in these training courses may not result in higher LAK. But, according to Ölmezer-Öztürk and Aydin (2019), ongoing training sessions regarding knowledge of assessment and testing in which teachers can utilize practices of assessment are favorable and effective. It is necessary for EFL teachers to learn how to design proper tests and utilize them in class with the aim of helping their learners achieve their goals in learning. As stated by Kremmel and Harding (2020) and Yan et al. (2017), if teachers have a good command of assessment and are aware of its importance in the teaching process, then, they will find assessment as an enjoyable experience and put more effort in the process of assessing their learners.

### **5. Conclusion and Implications**

The results of the current study were in line with previous research conducted by Ölmezer-Öztürk and Aydin (2019), Jannati (2015), and Vogt et al. (2020). This study stands out as one of the few quantitative investigations in the literature that explored Iranian teachers' Language Assessment Knowledge (LAK) and their understanding of assessment components. The study's primary contribution was demonstrating a significant gap between what a teacher can teach practically and what assessment practices he can apply in the context of the classroom which deviates from the latest theories of language assessment. It became evident that EFL teachers with limited expertise in assessment and low LAK might struggle to appropriately assess their classes. Additionally, teachers were found to lack sufficient information about assessment components, such as test types, functions, characteristics, and test design. Having that assessment knowledge is crucial to the process of teaching-learning and proves that educators are at the center of all practices related to assessment, Ölmezer-Öztürk and Aydin (2019) mention that it is imperative for EFL teachers to possess adequate knowledge in the assessment process.

The findings of this study carry important implications for both teacher educators and policymakers. It is recommended that teacher education programs undergo necessary changes to prioritize practical professional development with a specific focus on language assessment for teachers (Afshar & Ghasemi, 2020; Jalilzadeh et al., 2023). Policymakers, being responsible for crucial decisions related to education programs and assessments, should be knowledgeable about various assessment types and must ensure that teachers and researchers are given adequate opportunities and resources to conduct further research on Language Assessment Knowledge (LAK) (Taylor, 2013).

This study had some limitations related to the data collection techniques used and the small sample size. To obtain more valuable insights into teachers' assessment knowledge, it is recommended to incorporate classroom observations to observe how language teachers assess their learners and provide feedback in the actual classroom setting, particularly considering the diverse levels of Language Assessment Knowledge (LAK) among teachers. Additionally, employing a larger sample size would offer a more comprehensive understanding of EFL teachers' realities and how assessment techniques are employed in various contexts by different teachers.

Another limitation concerns the lack of differentiation among the participants based on their demographic characteristics, such as age, academic degree, and teaching experience. Future studies should address this aspect to better comprehend how these factors may influence teachers' assessment practices. Lastly, this study was confined to EFL teachers in the context of Iran. To gain a more extensive perspective, it is suggested to conduct similar research in various EFL and ESL contexts, considering teachers with diverse demographic backgrounds.

### **Acknowledgments**

We would like to thank the participants who agreed to take part in this study.

## References

- Afshar, H. S., & Ghasemi, S. (2020). Investigating the barriers to teachers' professional development in an EFL context. *Journal of Modern Research in English Language Studies*, 7(2), 101-122.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Ajzen, I. (2005). *Attitudes, personality and behaviour*. McGraw-Hill Education.
- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, 1(2), 217-232.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, 2(1), 1-34.
- Bachman, L.F. (2013). Ongoing challenges in language assessment. In A. J. Kunnan (Ed.) *The Companion to Language Assessment*, (pp.1-18). Wiley Blackwell.
- Bae, J., & Bachman, L. F. (2010). An investigation of four writing traits and two tasks across two languages. *Language Testing*, 27(2), 213-234.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9), 1175-1184.
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control. *Journal of Cognitive Psychotherapy*, 13(2), 158-166.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215-232.
- Boud, D. (2006). Creating the space for reflection at work. In *Productive reflection at work* (pp. 158-169). Routledge.
- Davidheiser, S. (2013). *Identifying Areas for High School Teacher Development: A Study of Assessment Literacy in the Central Bucks School District*. Drexel University, Philadelphia, Pennsylvania.
- Davidson, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly*, 4(1), 37-68.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative,*

- qualitative, and mixed methodologies*. Oxford University Press.
- Earl, L. M. (2013). Assessment for learning; Assessment as learning: Changing practices means changing beliefs. *Assessment, 80*, 63-71.
- Elshawa, N. R., Heng, C. S., Abdullah, A. N., & Rashid, S. M. (2016). Teachers' assessment literacy and washback effect of assessment. *International Journal of Applied Linguistics and English Literature, 5*(4), 135-141.
- Esfandiari, R., & Nouri, R. (2016). A mixed-methods, cross-sectional study of assessment literacy of Iranian university instructors: Implications for teachers' professional development. *Iranian Journal of Applied Linguistics, 19*(2), 115- 154. <http://dx.doi.org/10.29252/ijal.19.2.115>.
- Farhady, H., & Tavassoli, K. (2018). Developing a language assessment knowledge test for EFL teachers: A data-driven approach. *Iranian Journal of Language Teaching Research, 6*(3), 79-94.
- Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113-132.
- Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Taylor & Francis.
- Gotch, C. M., & French, B. F. (2013). Elementary teachers' knowledge and self-efficacy for measurement concepts. *The Teacher Educator, 48*(1), 46-57.
- Harlen, W. (2005). Teaching, learning and assessing science 5-12. *Teaching, Learning and Assessing Science 5-12*, 1-264.
- Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402.
- Inbar-Lourie, O. (2008). Language assessment culture. *Encyclopedia of Language and Education, 7*(2), 285-299.
- Jalilzadeh, K., Alavi, S. M., & Siyyari, M. (2023). EFL/ESL Teachers' writing assessment literacy: State of the arts. *Iranian Journal of Applied Language Studies, 15*(1), 101-120.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education, 6*(2), 26-37.

- Joughin, G. (2009b). Assessment, learning and judgement in higher education: A critical review. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 13-27). Springer Science & Business Media.
- Kahl, S. R., Hofman, P., & Bryant, S. (2013). *Assessment literacy standards and performance measures for teacher candidates and practicing teachers*. Paper prepared for the Council for the Accreditation of Educator Preparation (CAEP). Measured Progress.
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Mayo, S. T. (1967). *Pre-service preparation of teachers in educational measurement*. Final report. Chicago, IL: Loyola University (ERIC Document Reproduction Service No.: ED021784).
- McMillan, J. (Ed.). (2013). *Sage handbook of research on classroom assessment*. Sage.
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson.
- Mertler, C. A. (2000). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation*, 7(1), 25.
- Mohammadi, S. and Babaii, E., 2020. The Dynamic Assessment of Reading Comprehension: An Exploration of EFL Teacher's Perception and Practice. *Journal of Modern Research in English Language Studies*, 9(2), 123-148.
- Newmann, F. M., & Archbald, D. A. (1992). The nature of authentic academic achievement. In B. Harold, F. M. Newmann, E., Adams, D. A. Archbald, T. Burgess, J. Raven, & T. A. Romberg, *Toward a new science of educational testing and assessment* (pp. 71-83). State University of New York press.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380.
- Ölmezer-Öztürk, E., & Aydin, B. (2019). Investigating language assessment knowledge of EFL teachers. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(3), 602-620.

- Plake, B. S., & Impara, J. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- Popham, J. W. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63(6), 84-85.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental?. *Theory into Practice*, 48(1), 4-11.
- Price, M., Rust, C., O'Donovan, B., Handley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. The Oxford Centre for Staff and Learning Development.
- Sheppard, J. D. (2013). Perceptions of Teachers and Administrators Regarding the Teacher Evaluation Process, *Ph.D. Dissertation*, Electronic Theses & Dissertations, 852, available at: <https://digitalcommons.georgiasouthern.edu/etd/852>.
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23-27.
- Stiggins, R. J. (1999). Teams. Learning teams can help educators develop their collective assessment literacy. *Journal of Staff Development*, 20, 17-21.
- Stoynoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing: A resource for teachers and program administrators*. TESOL.
- Tavassoli, K., & Farhady, H. (2018). Assessment knowledge needs of EFL teachers. *Teaching English Language*, 12(2), 45-65.
- 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.

- Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. *Language Assessment Quarterly*, 17(4), 386-409.
- Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language- test specification reverse engineering. *Language Assessment Quarterly*, 7(4), 317-342.
- Wolf, D., Bixby, J., Glenn III, J., & Gardner, H. (1991). Chapter 2: To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17(1), 31-74.
- Yan, X., Fan, J., & Zhang, C. (2017, July). *Understanding language assessment literacy profiles of different stakeholder groups in China: The importance of contextual and experiential factors*. Paper presented at the Language Testing Research Colloquium, Bogotá, Colombia.