Demystifying Language Assessment Literacy among High School Teachers of English as a Foreign Language in Iran: Implications for Teacher Education Reforms

Rahim Watmani[1], Hassan Asadollahfam[1*], Bahram Behin[2]

Abstract

The promotion of language assessment literacy (LAL) among teachers and assessment developers has been deadly called and emphasized by professionals and scholars in the field of language assessment. In line with this urgent call for research, the present study sets out to investigate the EFL teachers' literacy assessment with the aim of bringing modifications for teacher education reforms. To meet the objectives of the study, a total of 200 EFL teachers (N=88, with a TEFL background and N=112, with a non-TEFL background) with B.A. (N=125) and M.A. (N= 75) degrees were selected through stratified random sampling at high schools of West Azerbaijan Province, Iran. The required data were gathered using Plake, Impara, and Fager's (1993) Teacher Assessment Literacy Scale (TALS). Hence, the participants' knowledge concerning components of assessment literacy (AL) was sought for. A multivariate analysis of variances (MANOVA) was run and the results, in general, revealed that the participants were not significantly familiar with assessment literacy principles and procedures. Specifically, the results indicated that EFL teachers with a TEFL background and those with a non-TEFL background differed in terms of their assessment literacy competence, especially in terms of their perceptions of AL components. The implications of the findings for teacher education programs, teacher educators and EFL teachers are discussed and some suggestions for further research are offered.

Keywords: Assessment literacy, EFL teachers, Non-TEFL background, Teacher education, TEFL background
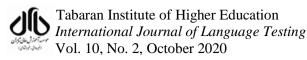
1. Introduction

One of the most crucial roles of EFL teachers in the language classroom is assessing the EFL learners' performance. The seminal work of Black and Wiliam (1998) on assessment for learning has increased the focus on assessment in language classrooms. It shed light on this fact that assessment plays a key role in improving students' learning (Fulcher, 2012). In actual fact,

[1] Department of English Language, Bonab Branch, Islamic Azad University, Bonab, Iran.
[2] English Department, Azarbaijan Shahid Madani University, Iran.
* (corresponding author: asadollahfam@gmail.com; h_asadollahfam@yahoo.com)

it led to this conclusion that if the teachers aim to achieve the desired goals, they need to be skilled in many aspects of assessment literacy (Poehner & Inbar-Lourie, 2020). The reason for this lies in the strong link between instruction and assessment such that "the quality of instruction in any classroom turns on the quality of the assessments used there" (Stiggins, 1999, p. 20).
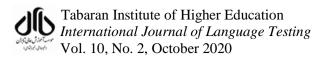
In the assessment literature, one of the first scholars who raised attention toward the significance of assessment literacy was Stiggins (1991) through his well-known statement "we are a nation of assessment illiterates." In 2006, Popham highlighted the same concern when he rightly termed the lack of appropriate training in the assessment as "professional suicide" (p. 82). He more opined that cultivating language teachers' assessment literacy can be highly advantageous for instruction. To justify his claim, he argued that the language teachers who enjoy a high level of assessment literacy, they can assess better their instruction by realizing their students' needs, monitoring learning and instructional processes, diagnosing their learning difficulty, and confirming their learning achievement (Gronlund & Linn, 1990). In actuality, empirical studies have demonstrated that one of the keystones to the success of the process of teaching (Wang, Wang & Huang, 2008), the quality of students' learning (DeLuca, Klinger, Searle, & Shulha, 2010; Mertler, 2004; Stiggins, 2007; Vogt, Tsagari, & Csépes, 2020; White, 2009), development of pragmatic ability (Cohen, 2014), and student learning motivation (Alkharusi, 2013; Dorman & Knightley, 2006) is teachers' assessment literacy.

Given the fact that language teachers' are the most important stakeholders directly involved in the enhancement of second language instruction and their undeniable impacts on EFL learners' performance, their competence in educational assessment is considered a very important professional requirement (Volante & Fazio, 2007). They are expected to put the principles of language assessment into practice to cultivate the EFL learners' performance, as well as to develop their own understanding of the nature of assessment itself and their role and practice as teacher assessors (Scarino, 2013). Despite such great importance attached to assessment in most educational settings, research literature demonstrates that many language teachers are not adequately prepared to deal with assessment issues (Kremmel & Harding, 2020; Xu & Brown, 2017) and they need some assistance in implementing various classroom assessments and in making assessment-related decisions (Janatifar & Marandi, 2018; Mertler, 1999; Mertler & Campbell, 2005); therefore, there is an urgent need for research the findings of which can contribute to and raise the awareness of language assessment literacy across teacher education programs.

2. Review of Literature

2.1. Language Assessment Literacy (LAL)

In the assessment literature, the term assessment literacy, coined by Stiggins (1995), represents the knowledge and skills a stakeholder (be they teachers or administrators) needs to practice appropriately assessment (Fulcher, 2012). However, it seems that there has been a lack of a comprehensive consensus over the basic assessment literacy competencies/components assessment (Fulcher, 2012; Taylor, 2013). Most of the definitions presented in the literature for LAL are in accordance with the *Standards for Teacher Competence in Educational Assessment*

of Students announced by the American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), and National Education Association (NEA) (Stiggins, 1991). The standards entail a number of competencies that assessment literate teachers should possess. For example, they emphasize that teachers need to select and develop assessments for the classroom, administer and score tests, use scores to aid instruction, communicate results to stakeholders, and be aware of the inappropriate and unethical uses of tests (Kremmel & Harding, 2020; Stiggins, 1991). Stiggins (1995) also states "assessment literates know the difference between sound and unsound assessment" (p. 240). He adds that assessment literates should know: (a) what they are assessing, (b) why they are doing so, (c) how best to assess the goal of interest, (d) how to generate sound samples of performance, and (e) what can go wrong and how to prevent those problems before they occur.

In a more recent attempt, Fulcher (2012) reconciled the underdevelopment perceived in assessment literacy by debating and redefining the term itself and considering the sociocultural and sociopolitical influences of assessment. Based on his analysis of previous research that examined the development of assessment literacy through language testing programs and the textbooks used in these programs for a decade (Brown & Bailey, 2008), as well as a teacher needs survey, Fulcher accurately argued, "research into assessment literacy is in its infancy" (p. 4). Therefore, he emphasized the need for three different aspects of assessment literacy, including practices, principles, and contexts and defined assessment literacy as follows:
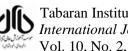
> "The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals." (2012, p. 13)

Perhaps what makes Fulcher's definition rather unique is that it is empirically derived from qualitative and quantitative data gathered on the assessment and testing needs of 278 language teachers, whereas most previous studies on teacher assessment needs only took a quantitative approach. In addition, the definition placed equal importance on the historical, social, political, and philosophical context of assessment and its practices and impact.

All in all, assessment literacy for the EFL teacher represents the knowledge, skills, and practices teachers possess and use to understand, design, administer, make decisions from, and evaluate classroom-based assessment according to the principles and concepts relevant to the fields of language testing and evaluation as well as language teaching and learning in accordance with the affordances (e.g., digital literacy) and needs (of both teachers and students) within their contexts (Brookhart, 2011; Eyal, 2012; Levi & Inbar-Lourie, 2020).

2.2. Factors that Impact Assessment Literacy

A large number of researchers have attributed the underdeveloped teacher assessment literacy to the lack of formal education on assessment (Mertler, 2004; Popham, 2006, 2009; Stiggins, 1995; Taylor, 2009) and/or in-service assessment training (Birgin & Baki, 2009; Fan, Wang & Wang, 2011; Mertler, 2009). The literature highlights some factors that impact teacher assessment literacy that could be classified into three major levels. The first level concerns the

teachers, who are regarded as the most important factor in classroom-based assessment (Vogt & Tsagari, 2014). Clearly, teacher assessment needs/competencies compose the most studied factor in assessment literacy research (Alkharusi, 2015; Fulcher, 2012; Inbar-Lourie & Levi, 2020; Vogt & Tsagari, 2014). The second level reports on the educational institutions that are responsible for designing and providing opportunities for professional development programs for pre-service (Alkharusi, 2013) and in-service teachers (Coniam, 2009; Koh, 2011). The third level concerns research on assessment literacy (Levi & Inbar-Lourie, 2020; Poehner & Inbar-Lourie, 2020). It is necessary to note that these classifications are not meant to separate factors from each other, as they are interrelated and usually affect each other.
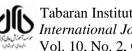
2.3. Related Studies in the Literature

In this section, some relevant studies are critically reviewed. In one of the early studies, Plake and Impara (1997) found that teachers receive little or no formal assessment training in the preparatory programs, and often they are ill-prepared to undertake assessment-related activities. The results of their study also indicate low levels of assessment competency for teachers. They further explained that the education community needs to recognize that teachers are ill-equipped to successfully undertake students' assessment as one of the most prevalent activities of their instructional program.

In another study, Alkharusi, Aldhafri, Alnabhani, and Alkalbani. (2014) assessed 167 teachers' assessment knowledge and reported weak results (an average of 12.42 of 32 items answered correctly). Using a similar approach to assess teacher assessment literacy, Mohamed, Kamis, and Ali (2017) also found that the participating teachers ($N = 187$ secondary school teachers of vocational subjects in Malaysia) showed moderate to low levels of assessment literacy. On a larger scale, Xu and Brown (2017) used an adapted version of the Teacher Assessment Literacy Questionnaire (TALQ) to examine EFL teachers' ($N = 891$) assessment knowledge in Chinese universities. The results showed an insufficient assessment of literacy levels. The authors urged language-teaching institutions to provide their teachers with ongoing professional development programs on assessment.

Moreover, Öz and Atay (2017) researched the assessment literacy of 12 teachers of the English Preparatory Program at a Turkish university using semi-structured interviews. The results of the study showed a mismatch between teachers' assessment perceptions and in classroom practices. Even though teachers viewed assessment as very important and they identified good assessment practices and principles, their practices were found to lack those principles of good assessment. So they did show some assessment literacy knowledge, but how that knowledge was reflected in practice calls for concern.

In the Iranian EFL context, Azadi (2018) has studied the conceptual factors of teacher assessment literacy among ESP instructors in the Iranian context and has found that Iranian ESP teachers taking part in the study were not well familiar with standards of teacher competencies, nor were they well familiar with the standards of assessment literacy. The findings imply that ESP teachers' literacy on educational assessment is inadequate especially on communicating assessment results to others and valid grading of the learners' assessment. Finally, Janatifar and Marandi (2018) aimed at investigating the components of LAL in the

Iranian EFL context using an adapted version of Fulcher's (2012) LAL survey with two types of constructed and closed response items. A total of 280 English language teachers from seventeen different provinces were selected in Iran. The findings indicated that LAL in the Iranian context is comprised of four factors, namely: test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects (which mainly includes social and ethical aspects of language testing/assessment), and reliability and validity. Furthermore, the results turned out that the EFL teachers in this study believe that besides the theoretical issues of assessment, they should also receive hands-on skills-based instruction in language assessment.

Kremmel and Harding (2020) who have developed a comprehensive LAL scale argue that besides the language teacher, professional raters, policymakers, employers, parents, and the learners themselves need to gain AL knowledge based on their levels of needs. In line with the same idea, Inbar-Lourie and Levi (2020) propose that LAL cannot meet its purposes unless it acts as e mediator between the teachers' AL knowledge and learning practices.

As the studies reviewed above clearly show, there is a paucity of research in the Iranian EFL context regarding the perceptions of EFL teachers with a TEFL background and the ones with a non-TEFL background toward the status of language LAL components. Moreover, there has been a lacuna in the literature concerning the role of teaching experiences and major in EFL teachers and non- EFL teachers' Assessment literacy. Hence, the present study sought to fill up the gap by addressing the following research questions:

1. To what extent are Iranian high school EFL teachers familiar with language assessment literacy components?
2. Is there any significant difference between the perceptions of the EFL teachers with a TEFL background and the ones with a non-TEFL background in terms of assessment literacy components?

3. Method

Give adequate information to allow the experiment to be reproduced. This section will include sub-sections.

3.1. Participants and Setting

The study participants were selected out of 30 high schools of West Azarbaijan Province in Iran. The sample consisted of a total of 200 male and female EFL teachers with a TEFL background (n=88) and non-TEFL background (n=112). Likewise, they had two different educational degrees, namely B.A. (N=125) and M.A. (N=75). Great care was exercised to select a representative sample of both groups of teachers using stratified random sampling according to the educational background and experience strata. Stratified random sampling, according to Mackey and Gass (2016), "provides precision in terms of the representativeness of the sample and allows preselected characteristics to be used as variables" (p. 120). It is worth noting that the participation in the study was voluntary and the teachers were assured that their names would remain confidential and they would be kept informed about the final findings.

3.2.Instrumentation

The Teacher Assessment Literacy Scale (TALS) developed by Plake and Impara (1997) was used to measure the participant's assessment literacy. In addition to the biographical information collected in terms of self-reported gender, age, major, and years of experience, TALS comprises 35 items related to the seven standards for language teacher competence in the educational assessment of students. These standards are used to guide what teachers should know and be able to do with respect to classroom assessment (Mertler, 2009). The standards consist of the following seven principles. In fact, teachers should be skilled in:

Standard 1   Choosing appropriate assessment methods
Standard 2   Developing appropriate assessment methods
Standard 3   Administering, scoring, and interpreting the results of assessments
Standard 4   Using assessment results to make decisions
Standard 5   Developing valid grading procedures
Standard 6   Communicating assessment results
Standard 7   Recognizing unethical or illegal practices

The items were developed to align with the Standards for Teacher Competence in Educational Assessment of Students (Stiggins, 1991). The items were dichotomously scored (0 = incorrect response, 1 = correct response) with a high total score reflecting a high level of assessment literacy. Campbell, Murphy, and Holt (2002) reported a KR20 reliability coefficient of .74 for the pre-service teachers' scores (See Table 1). Concerning its validity, after the scale items were translated into the participants' mother tongue (Persian) by a well-experienced translator, two university professors in applied linguistics were kindly invited to judge the items in line with the research purpose. Based on their comments, some modifications were made.

Table 1
Reliability statistics; Teachers' assessment literacy

		Cronbach's Alpha	N of Items
1	Choice	0.67	5
2	Develop	0.71	5
3	Administration	0.65	5
4	Decision	0.73	5
5	Grading	0.75	5
6	Communicating	0.72	5
7	Ethics	0.70	5
	Total	0.70	35

3.3.Procedures

   3.3.1. Data Collection Procedures

Some steps were followed to run the present study. At first, the instruments were translated into the Persian language by a well-experienced translator. Then, the instruments were given

to two university professors in applied linguistics to check if they were validated enough to be used in the current study. In fact, the aims were to provide ideas and to test the relevance of the instruments to the setting wherein the language teachers were working. At the next stage, some semi-structured interviews were conducted with several language teachers to ascertain that all the questions in the study are easy to comprehend. The next stage was the administration of the adapted scale to the selected sample. To do so, TLAS was given to the teacher participants and they were requested to complete the survey during non-instructional times at their convenience, enclose and return them to the researcher within one week of receipt. The respondents were guaranteed that their identity, as well as their responses, would be kept confidential.

3.3.2. Data Analysis Procedures

Prior to reporting the results, it is worth mentioning that the independent variables of the current study were the high school EFL teachers with and without a TEFL background, while the dependent variables were the 7 subscales of the TALQ, namely Standard 1 (Choosing appropriate assessment methods), Standard 2 (Developing appropriate assessment methods), Standard 3 (Administering, scoring, and interpreting the results of assessments), Standard 4 (Using assessment results to make decisions), Standard 5 (Developing valid grading procedures), Standard 6(Communicating assessment results), and Standard 7(Recognizing unethical or illegal practices).

To answer the research questions of the study, in addition to common descriptive statistics, the following statistical analyses were used. For the first research question, the mean of the scores for the participants' responses to each standard was calculated and reported. In order to answer the second research question, the statistical analysis of the MANOVA test was run.
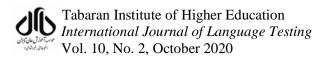
4. Results and Discussion

In this section, the results of the data analysis are presented. The first research question investigated to what extent the Iranian EFL teachers are familiar with language assessment literacy components. As reported in Table 2, the mean score on total assessment literacy was 13.16. The scores of the participants were calculated out of the total number of 35 (i.e., for each correct response the participants received 1 mark). That is to say, the participating teachers acquired 52.64 percent of the total score. This means that they were not able to gain 47.36 percent of the total score which accounts for a lot of information in the domain of assessment literacy. However, it can be claimed that they could gain 52.64 percent of the total score which means that they were moderately aware of the components of assessment literacy.

Table 2
Descriptive statistics; Participants' total assessment literacy scores

	N	Minimum	Maximum	Mean	SD
Total	200	1	27	13.16	5.187

Furthermore, the assessment literacy scale consisted of seven sub-sections, including choosing appropriate assessment methods, developing appropriate assessment methods, administering, scoring, and interpreting the results of assessments, using assessment results to make decisions, developing valid grading procedures, communicating assessment results, and recognizing unethical or illegal practices, each measured through five items. As presented in Table 3, the respondents had the highest means on choice (M = 2.60; 52.00 % total score), decision making (M = 2.43; 48.60 % total score) and development (M = 1.76; 35.20 % total score). They had the lowest means on ethics (M = 1.56; 31.20 % total score), administration (M = 1.55; 31.00 % total score) and grading (M = 1.48; 29.60 % total score).

Table 3
Descriptive statistics of components of total assessment literacy (EFL teachers with and without a TEFL background)

	N	Minimum	Maximum	Mean	SD	% out of 5	Rank
Choice of methods	200	0	5	2.60	1.252	52.00	1
Decision Making	200	0	5	2.43	1.369	48.60	2
Test Development	200	0	4	1.76	1.007	35.20	3
Communicating Results	200	0	4	1.75	1.036	35.00	4
Ethics	200	0	3	1.56	.720	31.20	5
Administration	200	0	4	1.55	.781	31.00	6
Grading	200	0	4	1.48	.716	29.60	7

In terms of ranking, *choosing appropriate assessment methods* and *using assessment results to make decisions* take the highest ranks showing that the EFL teachers irrespective of their teaching experiences and majors are moderately (Choice= % 52.00 and Decision Making = % 48.60) familiar with these assessment components. However, in other cases, the high school teachers, in general, are not knowledgeable as their performance in the test was not high (between % 29 and % 35.20).

The second research question examined if there was any statistically significant difference between the perceptions of the EFL teachers with a TEFL background and without with a TEFL background regarding assessment literacy components. To answer this question, a multivariate analysis of variances (MANOVA) was run. Before running MANOVA, it should be noted that the assumption of homogeneity of covariance matrices was retained (Box's M = 55.38, p = .003 > .001) (Table 4.). This assumption assumes that the correlations between any two variables are roughly the same across groups. As noted by Field (2018), Box's test should be reported at .001 levels.

Table 4
Box's test of equality of covariance matrices; Components of assessment literacy by groups

Box's M	55.385
F	1.902
df1	28
df2	121525.284
Sig.	.003

Considering the assumptions of ANOVA, the results of the Levene's test of homogeneity of variances indicated that the assumption of homogeneity of variances was met for decision making (F (1, 198) = 1.73, p = .190), grading (F (1, 198) = .981, p = .323), ethics (F (1, 198) = .009, p = .923); however, the assumption was violated for choice (F (1, 198) = 17.89, p = .000), development (F (1, 198) = 20.05, p = .000), administration (F (1, 198) = 5.43, p = .021) and communicating (F (1, 198) = 15.76, p = .000). One simple solution to overcome the problem of heterogeneity of variances is to reduce the significance level (alpha) to .025 or .01; as noted by Tabachnick and Fidell (2014, p 120), "violations of homogeneity of variances usually can be corrected by transformation of the dependent variable (DV) scores. Interpretation, however, is then limited to the transformed scores. Another option is to use untransformed variables with a more stringent level (for nominal α, use .025 with moderate violation and .01 with severe violation)". To be on the safe side, the researcher decided to discuss the results of MANOVA at .01 levels of significance. Table 5 presents the descriptive statistics; components of assessment literacy by groups.

Table 5
Descriptive statistics; Components of assessment literacy by groups

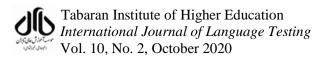| Dependent Variable	Background	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Choice of methods	Non-TEFL	2.182	.128	1.930	2.433
	TEFL	2.938	.113	2.714	3.161
Test Development	Non-TEFL	1.352	.100	1.155	1.550
	TEFL	2.089	.089	1.914	2.265
Administration	Non-TEFL	1.295	.080	1.138	1.453
	TEFL	1.750	.071	1.610	1.890
Decision Making	Non-TEFL	1.602	.124	1.358	1.846
	TEFL	3.071	.110	2.855	3.288
Grading	Non-TEFL	1.261	.074	1.116	1.407
	TEFL	1.652	.065	1.523	1.780
Communicating Results	Non-TEFL	1.364	.104	1.158	1.570
	TEFL	2.054	.093	1.871	2.236
Ethics	Non-TEFL	1.330	.074	1.184	1.475
	TEFL	1.741	.065	1.612	1.870

Also, based on the results (F (7, 192) = 11.71, p = .000 < .01, Partial eta squared =.299 representing a large effect size), it can be said that there were statistically significant differences between the two groups' means on the seven components of assessment literacy. Likewise, comparison of the means on the components of assessment literacy for the EFL teachers with and without a TEFL background showed that the teachers with a TEFL background had higher means on all components of assessment literacy than the ones without a TEFL background.

It can also be claimed that the EFL teachers with a TEFL background (M = 2.93) had a significantly higher mean on choice (F = 19.64, p = .000, partial eta squared = .090 representing a moderate effect size) than the ones without a TEFL background (M = 2.182). Moreover, it can be said that the EFL teachers with a TEFL background (M = 2.08) had a significantly higher mean on development (F = 30.25, p .000, partial eta squared = .133 representing an almost large effect size) than the ones without a TEFL background (M = 1.35). Likewise, the EFL teachers with a TEFL background (M = 1.75) had a significantly higher mean on administration (F = 18.11, p .000, partial eta squared = .084 representing a moderate effect size) than the ones without a TEFL background (M = 1.29). Further, the EFL teachers with a TEFL background (M = 3.07) had a significantly higher mean on decision making (F = 79.02, p .000, partial eta squared = .285 representing a large effect size) than the ones without a TEFL background (M = 1.60). Plus, the EFL teachers with a TEFL background (M = 1.65) had a significantly higher mean on grading (F = 15.75, p .000, partial eta squared = .074 representing a moderate effect size) than the ones without a TEFL background (M = 1.26). In addition, the EFL teachers with a TEFL background (M = 2.05) had a significantly higher mean on communicating (F = 24.44, p = .000, partial eta squared = .110 representing a moderate effect size) than the ones without a TEFL background (M = 1.36). Also, the EFL teachers with a TEFL background (M = 1.74) had a significantly higher mean on ethics (F = 17.40, p .000, partial eta squared = .081 representing a moderate effect size) than the ones without a TEFL background (M = 1.33).

5. Conclusion

As stated above, the first research question investigated to what extent the Iranian EFL teachers are familiar with language assessment literacy components. The results firstly revealed that the EFL teachers were moderately familiar with two of the assessment components; choosing appropriate assessment methods and using assessment results to make decisions. On the other hand, the findings indicated that they were weak in other components of AL. One of the significant issues in this regard was that the high school teachers with a TEFL background were mostly aware of standard 1 (choosing appropriate assessment methods) and standard 4 (using assessment results to make decisions). However, they were less aware of other standards such as standard 2 (developing appropriate assessment methods), standard 6 (communicating assessment results), standard 5 (valid grading procedure), standard 7 (recognizing unethical or illegal practices), and standard 3 (administering assessment processes).

Also, it was revealed that the EFL teachers with a non-TEFL background were, to some extent, aware of the components of the teacher assessment literacy components. They had the
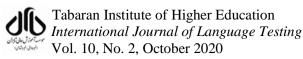
highest mean on standard 4 (decision making) and standard 1 (choosing appropriate assessment methods). However, they were less aware of standard 2 (developing assessment methods), standard 6 (communicating assessment results), standard 3 (administration of assessments), standard 7 (ethical issues), and standard 5 (valid grading procedure). In addition, the results indicated that there were significant differences between the EFL teachers with a TEFL background and the ones with a non-TEFL background in terms of their means on the seven components of assessment literacy. As the results showed, irrespective of the fact that both groups of teachers had higher means on standard 1 (choosing appropriate assessment methods) and standard 4 (using assessment results to make decisions), they differed in their degree of awareness toward other components and standards of assessment literacy (AL). The EFL teachers with a TEFL background had higher means on standard 2 (developing appropriate assessment methods), while the EFL teachers with a non-TEFL background had higher means on standard 6 (communicating assessment results).

The second research question, as mentioned above, explored if there was any statistically significant difference between the perceptions of EFL teachers with a TEFL background and the ones with a non-TEFL background regarding assessment literacy components. In general, the findings indicated that there existed a statistically significant difference between the two groups in terms of their perceptions germane to assessment literacy components. In fact, the results showed that the EFL teachers with a TEFL background enjoyed higher AL knowledge of the AL different components compared to the ones with a non-TEFL background. In addition, it was revealed that Iranian high school teachers (both with a TEFL and a non-TEFL background) had the highest means on standard 1 (choosing the appropriate assessment method) and standard 4 (decision making based on assessment results), while they had the lowest means on standard 3 (administrative issues) and standard 5 (valid grading of scores).

The findings of both research questions are in line with the previous studies, such as Alemi, Miri, and Mozafarnezhad (2019), Azadi (2018), Barootchi and Keshavarz (2002), Elshawa, Heng, Abdullah, and Rashid (2016), Kremmel & Harding (2020), Melter (2004), Mohamed et al. (2017), and Plake, Impara, and Fager (1993). For example, the findings of the present study are in line with Mertler's (2004) study focusing on pre-service versus in-service teachers' assessment literacy. It was found that both in-service and pre-service teachers performed highest on standard 3, and standard 1, respectively, while they had the lowest performance on standard 5 (developing valid grading procedures). Further, the findings partially lend support to the findings of Plake (1993) and Plake, Impara, and Fager (1993). They used TALQ and focused on the assessment literacy of in-service teachers. In general, they found that the highest mean performance was on standard 1 (choosing appropriate assessment methods) and the lowest mean was on standard 6 (communicating assessment results).

The findings also confirmed Volante and Fazio's (2007) study exploring teacher candidates' assessment literacy and examining the assessment literacy of primary/junior teacher candidates in all four years of their concurrent program. The results revealed that levels of self-efficacy remained relatively low for teacher candidates across each of the four years of this program. Most candidates suggested summative purposes for assessment and only a
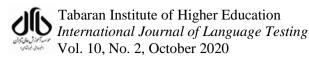
minority expressed formative purposes. They favored observational techniques and personal communication. Likewise, the results are in line with Janatifar and Marandi's (2018) reporting that the Iranian EFL teachers are not well familiar with the basic components of the LAL. Hence, they asked that besides the theoretical issues of assessment, they are badly in need of receiving hands-on skills-based instruction in language assessment in in-service training courses.

One possible reason for the findings might be related to the courses that the participants have passed on assessment and testing. It seems that these courses have not been so qualified to prepare the participants to administer the basic principles of assessment literacy. In fact, it may be argued that although assessment literacy is of paramount importance for EFL teachers to get access to second language education enterprise, unfortunately, the Iranian EFL teachers are usually accustomed to working with testing issues rather than assessment concepts (Barootchi & Keshavarz, 2002). The findings may be discussed from this perspective that EFL teachers can run their classes in line with the basic tenets of LAL unless they are trained in terms of LAL theories and practices. As such, pre-service and in-service training programs on assessment will gain importance in this regard as through such programs teachers get motivated to learn new assessment methods and implement them in their career duties. In addition, the significant difference between the EFL teachers with a TEFL background and the ones with a non-TEFL background may be attributed to the nature of the major and educational courses they have passed over their under graduation. That is, the courses that the TEFL graduates have to take up during their study may more or less affect their perceptions about LA. Hence, it is reasonable to expect the former participants outweigh the latter group.

Previous research has shown that traditional teacher preparation courses in classroom assessment are not well matched with what teachers need to know for classroom practice (Xu & Brown, 2017). The traditional programs have historically focused on large-scale (standardized) testing, although this trend is changing. That's why it can be argued that one or two courses in assessment and measurement may not truly be sufficient to cover everything teachers need to know on LAL (Mertler, 2009). In fact, this argument lends credence to Mertler and Campbell's (2005) study on measuring teachers' knowledge and application of classroom assessment concepts. They confirmed that in a lot of cases teachers were not well familiar with assessment literacy principles. This is made even worse when considering the fact that many teacher preparation programs in Iran do not even require a course in assessment (Azadi, 2018). In addition, the findings can be ascribed to the unwillingness of the Iranian teachers to participate in workshops on assessment are run here and there across the country. As Rezaee, Khoshsima, Behtash, and Sarani (2018) reported, most of the Iranian EFL teachers are highly demotivated to improve their professional competence and, accordingly, their job performance due to the demoralized organizational climate in schools.

In line with the findings, some implications are suggested. As the findings clearly demonstrated, since the Iranian EFL teachers seem to lack language assessment literacy in important domains, an urgent evaluation of the current provisions of teacher preparation programs in terms of language assessment literacy is needed. The implication of the findings is, in fact, considering the urgent needs for shaping these programs to expand their theoretical

as well as practical assessment courses to encompass the various assessment domains. In addition, as the findings of this study revealed that there is not a huge difference between the EFL teachers with a TEFL background and the ones with a non-TEFL background, who had never had any pre-service and in-service training on assessment, it is recommended that state schools and private language institutions evaluate the types of pre-service and in-service training they currently offer and how to better modify them to meet EFL teachers' needs. Likewise, the findings recommend that Teachers (Farhangian) Universities in Iran support targeted professional development programs that are based on teachers' assessment needs. Finally, the results suggest that testing and assessment materials developers include practical guides in assessment textbooks such that the EFL teachers easily know how to put the theoretical concepts into practice. Otherwise, the theoretical concepts would languish in academic journals and books without any pedagogical benefits for EFL practitioners.

In accordance with the current study's findings, some suggestions for further research are presented here. Future research can explore how EFL teachers had pre-service courses on language assessment are to use student-involved assessment such as peer-assessment. In addition, qualitative future research can unearth the Iranian university professors' perceptions regarding theoretical and practical assessment literacy components. Moreover, a study with an experimental approach needs to investigate EFL teachers' assessment literacy before and after a practical training on assessment principles and procedures. In the end, those certain aspects of assessment which have been neglected in teacher assessment courses can be highlighted by observing and focusing on the ways in which teachers are practicing assessment in their classroom. These aspects can be made the focus of future language assessment research.

6. Acknowledgments:

References

Alemi, M., Miri, M., & Mozafarnezhad, A. (2019). Investigating the effects of online concurrent group dynamic assessment on enhancing grammatical accuracy of EFL learners. *International Journal of Language Testing, 9*(2), 29-43.

Alkharusi, H. (2013). Canonical correlational models of students' perceptions of assessment tasks, motivational orientations, and learning strategies. *International Journal of Instruction, 6*(1), 21-38.

Alkharusi, H. (2015). An evaluation of the measurement of perceived classroom assessment environment. *International Journal of Instruction*, *8*(2), 45-54.

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Classroom assessment: Teacher practices, student perceptions, and academic self-efficacy beliefs. *Social Behavior and Personality: An International Journal, 42*(5), 835-855.

Azadi, A. (2018). *A study on the conceptual factors of teacher assessment literacy among ESP instructors*. Unpublished master's thesis, Islamic Azad University, Electronic Branch, Tehran, Iran.

Barootchi, N., & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, *44*(3), 279-288.

Birgin, O., & Baki, A. (2009). An investigation of primary school teachers' proficiency perceptions about measurement and assessment methods: The case of Turkey. *Procedia-Social and Behavioral Sciences*, *1*(1), 681-685.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31.

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

Brown, J. D., and Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing, 25*(3) 349-384.

Cohen, A. D. (2014). Towards increased classroom assessment of pragmatic ability. *International Journal of Language Testing, 4*(1), 5-25.

Coniam, D. (2009). Investigating the quality of teacher-produced tests for EFL students and the effects of training in test development principles and practices on improving test quality. *System*, *37*(2), 226-242.

DeLuca, C., Klinger, D. A., Searle, M., & Shulha, L. M. (2010). Developing a curriculum for assessment education. *Assessment Matters*, *2*(3), 20-42.

Dorman, J. P., & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educational Studies*, *32*(1), 47-58.

Elshawa, N., Heng, C. S., Abdullah, A. N., & Rashid, S. (2016). Teachers' assessment literacy and washback effect of assessment. *International Journal of Applied Linguistics and English Literature, 5*(4), 135-141.

Eyal, L. (2012). Digital assessment literacy-the core role of the teacher in a digital environment. *Educational Technology & Society*, *15*(2), 37-49.

Fan, Y. C., Wang, T. H., & Wang, K. H. (2011). A Web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education*, *57*(2), 1727-1740.

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

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. New York: MacMillan.

Inbar-Lourie, O., & Levi, T. (2020). Assessment literacy as praxis: Mediating teacher knowledge of assessment-for-learning practices. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment* (pp. 241-259). Springer, Cham.

Janatifar. M., & Marandi, S. S. (2018). Iranian EFL teachers' language assessment literacy (LAL) under an assessing lens. *Applied Research on English Language, 7*(3), 361-382.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*(3), 255-276.

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.

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly, 17*(2), 168-182.

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, *120*(2), 285-285.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, *33*(1), 49-64.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, *12*(2), 101-113.

Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "Assessment Literacy Inventory." *Online Submission*. Retrieved from http://files.eric.ed.gov/fulltext/ED490355.pdf

Mohamed, S., Kamis, A., & Ali, N. (2017). Gauging the assessment literacy of Malaysia's home economics teachers: An empirical study. *Geografia-Malaysian Journal of Society and Space, 12*(3), 130-138.

Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal, 6(1),* 25-44.

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher, 6*(1), 21-27.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning adjustment, and achievement* (pp. 53-58). San Diego: Academic Press, Inc.

Plake, B. S., Impara, J. C., & Fager, J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10-39.

Poehner, M. E., & Inbar-Lourie, O. (2020). *Toward a reconceptualization of second language classroom assessment*. Springer International Publishing.

Popham, W. J. (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.

Rezaee, A., Khoshsima, H., Bahtash, E. Z., & Sarani, A. (2018). A mixed method study of the relationship between EFL teachers' job satisfaction and job performance in Iran. *International Journal of Instruction, 11*(4), 391-408.

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

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 1*(52), 534-539.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *3*(77), 238-245.

Stiggins, R. J. (1999). Are you assessment literate? *The High School Journal, 6*(5), 20-23.

Stiggins, R. J. (2007). Assessment through students' eyes. *Education Leadership*, *64*(8), 22-26.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics.* Harlow, Essex: Pearson Education.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, *2*(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. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374-402.

Vogt, K., Tsagari, D., & Csépes, I. (2020). Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly*, 1-24. DOI: 10.1080/15434303.2020.1776714

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749-770.

Wang, T. H., Wang, K. H., & Huang, S. C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, *51*(1), 448-462.

White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *On Cue Journal, 30*(1), 3-25.

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