

Investigating Iraqi EFL Student- Teachers' Assessment Literacy: Matches and Mismatches between their Perceptions and Knowledge

M.A. Entidhar Neamah Aboalshoon Janabi

Islamic Azad University South Tehran Branch

Abstract: The success of teaching, the quality of student learning, and the motivation of students to learn all depend on teachers' ability to effectively assess their students. However, studies have repeatedly shown that the levels of assessment literacy for instructors in general education and language teaching are inadequate. The purpose of this research is to gain an understanding of the current level of assessment literacy among Iraq EFL student-teachers as well as to determine the areas in which their assessment knowledge is lacking and where it excels. This investigation makes use of a descriptive quantitative methodology, which includes an assessment knowledge test produced by Farhady and Tavasouli (2018) as well as a teacher's perception assessment test. The outcomes of this research point to both strengths and shortcomings in EFL teacher assessment literacy, as well as matches and mismatches between student-teachers' self-perceived assessment literacy and the demonstrated assessment expertise of their students. Over half of all teachers of English to speakers of other languages believe they have sufficient language assessment competence. Nevertheless, the findings that were gleaned from the two tests suggested that almost two thirds of them lacked the necessary level of literacy in terms of assessment knowledge. According to these findings, it was discovered that there is a large discrepancy between students-teachers' beliefs of their assessment knowledge and the actual level of literacy that they possessed. The study may have some repercussions for EFL teacher preparation programs, teaching institutions, and other future research areas.

Keywords: Language assessment, Language assessment literacy, Formative assessment, Summative assessment.

Introduction

Recognizing the importance of LAL of language teachers and their teachers' development needs, many studies have been conducted to conceptualize LAL, examine LAL of teachers, and validate evidence in different contexts and from different perspectives. However, this research is "still in its infancy" (Fulcher, 2012, p. 117), and further research on this topic is needed to advance LAL research. To provide a comprehensive picture of current research and lay the foundation for future LAL studies, this review will first examine the conceptualizations of LAL and the empirical studies on teachers' LAL, then provide implications and future directions.

Teacher's familiarity with the main principles of assessment measures is considered an integral part of teacher assessment literacy and related directly to what happens in classroom (Popham, 2009). Language assessment literacy can also profoundly affect teachers' involvement in decision-making about their teaching and, in turn, their practices. Due to its vital role in the educational context, assessment literacy has always been the center of attention. Although some research has been carried out on grammar and pragmatic teaching, there has been a little empirical investigation into the EFL teachers' assessment literacy in Iraq. So the aim of this study was to shine new light on these debates by investigating the current levels of language assessment literacy of Iraqi EFL student-teachers (knowledge, perceptions, skills) and identify weaknesses and strengths of their assessment knowledge.

Lack of assessment literacy among EFL teachers and the quality of English language tests in EFL contexts are of concern (Cheng, 2008). A general assumption about the assessment practices of language teachers is that language teachers cannot write perfect tests. However, in the Iraqi EFL context, a large number of EFL teachers, especially high school teachers, are often involved in or responsible for developing classroom, institutional, and local assessments. Even though they have not had an independently assessment course during their undergraduate education, nor any assessment on-the-job training course, at least in the past 10 years, teachers believe their undergraduate courses have the potential to prepare them for classroom assessment (Muhammad & Bardakci, 2019). Iraqi EFL pre-service teachers (students of the colleges of Education) take a general course on the teaching methods of EFL in the English Department of the Faculty of Education due to the excessive concentration on Linguistics and literature, rather than the methods of teaching EFL in English

courses (Al-Jabbawi, 2022). The issue is that EFL teacher assessment literacy in Iraq is far from satisfactory. They are not aware of their need to improve (Muhammad & Bardakci, 2019).

Improving the language assessment of teachers can greatly help teaching. For example, teachers who assess their subjects can identify student needs, monitor teaching and learning progress, identify students' learning difficulties and monitor academic performance (Gronlund & Linn, 1990). Despite the strong emphasis on assessment in many educational institutions, the literature shows that many teachers are unprepared to deal with assessment issues and need some help in conducting classroom assessments and in making assessment-related decisions (Mertler, 1999; Mertler & Campbell, 2005); therefore, there is a need for more emphasis on language assessment literacy in teacher education programs. Language teachers who are assessment literate can enhance the quality of their instruction and respond to their students' instructional needs more effectively. In this regard, this study attempted to address the aforementioned issues in the context of EFL instruction in Iraq.

LAL is a significant component of language teachers' expertise, but is also a challenging task for most language teachers. To date, there have been relatively few studies examining the research on teachers' LAL during these decades. To fill this void, this study is important for a number of reasons. First, a large body of research deals with assessment literacy, but research specifically on the assessment literacy of EFL teachers is limited. Therefore, this study is a good complement to the EFL literature and language testing research. Second, this study reviews the conceptualizations of LAL and related empirical studies published from 1991 to 2021. It first analyzes various conceptualizations of LAL. It will significantly boost the validity of the study's findings and enrich the literature on language instruction and assessment. Finally, the future research directions of university teachers are discussed.

1. To what extent are Iraqi EFL student- teachers' assessment literate?
2. To what extent do Iraqi EFL student- teachers view themselves as assessment literate?
3. To what extent do Iraqi EFL students-teachers' assessment literacy perceptions match with their assessment knowledge?

Literature Review

Language Assessment Literacy (LAL): Concept and Competencies

Despite the increasing importance of LAL, a major concern which still remains is determining the primary focus of literacy in language assessment. This would include a variety of skills related to test design, interpretation and use of test results, test evaluation, and the role and function of assessment in education and society (Inbar-Lourie, 2008). LAL may be defined as the knowledge of the principles and concepts that guide and support practice, including the knowledge, skills and competencies, familiarity with the testing process, ethics, and codes necessary to design, develop, maintain, or evaluate large-scale standardization and/or classroom testing (Fulcher, 2012). It can be viewed as the "ability to place knowledge, skills, processes, principles and concepts in a wider historical, social, political and philosophical framework, to understand why practices occur and to assess the role and impact of tests on society, organizations and individuals" (Fulcher, 2012, p. 125). Melone (2013) focused on instructional issues and defined LAL as "language teachers' knowledge with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language" (p. 329). On the other hand, Scarino (2013) emphasized the main role of teachers in assessment; therefore, LAL has also been defined as "student performance assessment, teacher knowledge, understanding and assessment experience" (p. 310). Finally, LAL may be regarded as competency repertoire used for understanding, evaluating and constructing language tests and analyzing test data (Phill & Harding, 2013).

The Significance of Language Assessment Literacy (LAL) in Teacher Development Programs

Language teachers are recognizing the relevance of LAL for two reasons: delegating responsibility for classroom assessment to language teachers, and moving Classroom-Based Evaluation (CBA) from assessment of learning to assessment for learning. Scarino (2017) suggested that the worldwide movement of people has pushed language teachers to shift from a communicative to an intercultural perspective in language learning. This theoretical shift affects teaching, learning, and evaluation, therefore "language teachers need advanced assessment skills" (p. 21).

Tsagari and Vogt (2017) suggested that CBA requires EFL teachers with language assessment knowledge. They asked language teachers to upgrade their qualifications and not neglect classroom challenges. The shift from measuring students' achievement to comparing student performance to established criteria places additional responsibility on teachers, who should take LAL seriously. LAL seems most relevant to language testers (Malone, 2013; Popham, 2009), but a "holistic approach that goes beyond a generic knowledge-based definition of LAL" is needed for teachers (Tsagari & Vogt, 2017, p.43).

Malone (2013) compared how language testing professionals and teachers view course content. The same issue was tackled as a language tester or instructor. The former focuses on technical difficulties, while the latter on classroom assessment techniques in language testing/assessment courses. LAL can also be seen from the perspective of administrators while making IELTS admission decisions (O'Loughlin, 2013). Phil and Harding (2013) found that decision makers' lack of understanding of language testing/assessment difficulties, techniques, and objectives might lead to major misconceptions. When working with LAL, it's crucial to balance theory and practice. Language professors, course instructors, university administrators, and policymakers should also be considered (Taylor, 2013). "Teacher assessment literacy has just lately entered the language assessment community's agenda" (Razavipour, Riazi, & Rashidi, 2011), therefore more research is needed to develop the idea and its components in our environment. Previous studies examined EFL teachers' assessment literacy. One study showed that language assessment training in Hong Kong is insufficient, and courses fail to reconcile theory and practice in the context of assessment reform (Lam, 2014). Another study examined Iranian EFL teachers' evaluation skills and feedback effects. The latest research shows that English teachers' assessment skills are poor. "Low assessment skills among EFL teachers necessitate pre- and in-service training" (Razavipour et al., 2011, p.160).

Teacher Assessment knowledge

Teachers' assessment literacy is a crucial part of teachers' quality and professional requirements. According to Mertler (2003), teachers' assessment skills can either enhance or limit learning process and student achievement. Researchers have reported that teachers' lack of confidence, knowledge and assessment experience has a negative impact on student achievement and learning (Brookhart, 2011; Plake, 1993). Although there are studies on assessment literacy, Fulcher (2012) accentuates that research on this field is not well developed yet. When the related literature is examined it can be said that assessment literacy has been studied and proposed in three stages: (1) the concept of teacher assessment, (2) teachers' knowledge and skills, and (3) the practice of teacher assessment in class. The teacher's concept and value of assessment are the most important factors in interpreting and making assessment decisions. Based on the standards and recent updates (e.g., Brookhart, 2011; Stiggins, 1991), teachers ought to be skilled in selecting the correct evaluation method in making instructional choices and developing assessment methods appropriate to options. They are likely to administer, score and interpret results of external and instructor-produced methods and use results for making options about individual students' performances. Teachers will be able to communicate the results of the assessment to the appropriate stakeholders. Finally, they should identify unconstitutional, illegal, and inappropriate assessment methods and utilization of assessment information.

Previous Studies on Teacher Assessment Literacy

A large number of studies on teacher assessment literacy have concentrated on teachers' assessment expertise as well as their attitudes on assessment literacy (Coombe et al., 2012;). This should not come as a surprise given the widespread belief that the assessment knowledge and attitudes of instructors have an effect on the strategies they employ in the classroom (Alkharusi et al., 2012; Stiggins 1995). The majority of the research that has investigated teacher assessment knowledge and language teacher assessment knowledge has consistently reported limited levels of teacher assessment/language assessment literacy (e.g., Alkharusi, 2011; Mertler, 2004; Plake, Impara, & Fager, 1993). This is the case despite the fact that there have been numerous attempts to investigate teacher assessment knowledge and language teacher assessment knowledge.

Alkharusi et al. (2012), for example, evaluated the assessment knowledge of 167 instructors by using the Teachers' Assessment Literacy Questionnaire (TALQ), and found that the teachers' knowledge was lacking (an average of 12.42 of 32 items answered correctly). Using a method similar to Alkharusi et al.'s (2012), Mohamed et al. (2017) examined the assessment literacy of teachers, and discovered that the teachers who participated in the study exhibited moderate to low levels of assessment literacy. Kasikhan and Rezaee (2019) studied ESP teachers' assessment literacy. The study included 50 male and female Iranian ESP instructors with Teaching English as a Foreign Language (TEFL) and non-TEFL backgrounds. TEFL-trained and untrained Iranian instructors have the finest tools for choosing an evaluation technique and making decisions, but the least for administrative issues and accurate scoring. On the literacy components of the assessment, there were large discrepancies between the two groups of ESP instructors. Study participants lacked understanding about teacher competency standards. Neither knew the literacy requirements. Their findings suggest ESP instructors' comprehension of educational assessment is insufficient, especially in explaining assessment results and validly evaluating learners' assessments.

Xu and Brown (2017) also employed a modified version of the TALQ to investigate the assessment expertise of EFL instructors working in Chinese institutions. Additionally, the findings revealed insufficient levels of literacy in the exam. The authors strongly recommended that universities that teach foreign languages

give their instructors access to continuing professional development programs that focus on evaluation. Firoozi, Razavipour, and Ahmadi (2019) studied Iranian EFL teachers' language assessment literacy needs and reformed assessment practices. Iran interviewed 15 EFL principals. Also, curriculum reorganization documents were closely examined. Inductive data coding demonstrated that instructors must change their evaluation of language skills to accomplish reform goals. Educators need knowledge and skills to assess language usage. Educators need training in the development of speaking and writing rubrics. Students must develop higher-order cognitive skills to assess their reading and listening comprehension. Because they aren't native English speakers, Iranian English teachers must improve. Along the same lines, Farhady and Tavassoli (2018) studied EFL teachers' knowledge of language-testing devices. 246 language instructors were given Fulcher's (2012) needs assessment questionnaire to collect data. They observed that most teachers had distinct evaluation priorities despite agreeing on the same issues. Most participants wanted further evaluation training.

Teacher evaluation perception research has shown similar difficulties. Birgin and Baki (2009) revealed that most of the 512 primary teachers polled considered themselves skilled in devising and implementing standard evaluation techniques but less so in alternative approaches. In an unpublished research study by Sheehan and Munro (2017) that employed classroom observations, interviews, and focus group interviews, EFL teachers with limited assessment experience indicated their lack of confidence in creating assessment items. Although the study's findings revealed the observed instructors employed a variety of assessment approaches well without training, they again stressed the necessity for classroom-based assessment training. In their 2010 study, DeLuca and Klinger found that instructors with formal assessment training were more confident. Mertler (2009) revealed assessment training significantly improved teacher assessment views and knowledge. Tajeddin, Khatib, and Mahdavi (2022) studied the critical language assessment literacy of EFL teachers in a separate study. According to the findings, the strongest correlations existed between the items and the following five factors: (a) the instructors' understanding of assessment purposes, scopes, and types; (b) the implications of assessment usage; (c) fairness; (d) assessment policies; and (e) national policy and ideology. It was discovered that the scale had a good level of internal consistency and construct validity, indicating that this scale has the potential to be effective for measuring language instructors' CLAL and for increasing their awareness of CLAL constructs. It was determined that the scale has a high level of internal consistency due to its construct validity and high level of internal consistency.

In terms of the assessment procedures that instructors use in the classroom, several studies have shown that these procedures do not align with the procedures that are advised by testing and assessment specialists. For instance, Alsarimi (2000) discovered that the majority of the 246 science teachers in his sample used traditional assessment methods more than alternative assessment methods. This is a practice that has a negative impact on students' self-efficacy beliefs as well as their perceptions of the congruence of the assessment with planned learning. Alsarimi (2000) found that the majority of science teachers in his sample used traditional assessment methods more than alternative assessment methods (Alkharusi et al., 2014b). According to the findings of another study carried out in the same environment, instructors have a tendency to rate students based on non achievement variables such as behavior and effort (Alkharusi et al., 2012).

Oz and Atay (2017) conducted a study on assessment literacy of English instructors working in the English Preparatory Program at a Turkish university. Their study focused on teaching English as a foreign language (EFL). The findings of the study revealed a discrepancy between the evaluation opinions of instructors and the procedures actually implemented in 41 classrooms. It was discovered that teachers' assessment procedures did not adhere to the principles of good assessment, despite the fact that instructors acknowledged that assessment is highly important and that they recognized appropriate assessment practices and concepts. Therefore, they did demonstrate some awareness of assessment literacy; however, the manner in which this information was expressed in practice raises concerns.

All of these research studies demonstrate the necessity of evaluating not just technical but also sociocultural factors. EFL teachers must lack a solid understanding of assessment issues such as reliability, validity, and test design, as well as other relevant topics; teacher trainers should take this into account. Moreover, teachers of English as a foreign language (EFL) must be familiar with their students' needs in order to include relevant material on their exams. In a similar spirit, teachers of English to speakers of other languages (EFL) should work to enhance their assessment literacy because prior research indicates that assessment literacy influences their students' learning.

Method

Thirty EFL student teachers from different cities in Iraq participated in this study. The method of sampling was convenience method to ensure response validity. They were both male and female aged between 30 and 50 years old and 11 of them were secondary or preparatory schools teachers. They have already finished their third semester of Master's study in TEFL, all of them have passed assessment course. Convenience

sampling is a non-random sampling technique defined as the selection of people who chance to be available for research" (Mackey & Gass, 2016). It is the most common method of sampling in EFL research (Dornyei, 2007).

The participants were 76% females and 24% males. The age of participants ranged between 30 and 50 years. The majority was in the 30-40 age group. 63% of them didn't have any teaching experience. While 14% of the participant have from 5 to 10 years teaching experience and 23% were in the 10-15 years group. All of them were MA students in TEFL. Table (4.3) outlines the participants' demographics.

Variable		N	percent	total
Gender	Male	7	24%	30
	Female	22	76%	
Age	30-40	24	80%	30
	40-50	6	20%	
Teaching experience	No teaching experience	19	63%	30
	5-10	4	%14	
	10-15	7	%23	
Field of study at university	TEFL	30	%100	30
Educational level	MA Student	30	%100	30

To achieve more accurate results and in line with the theoretical principles of teacher assessment literacy, the following instruments were used:

Assessment knowledge test: This test was developed by Farhady and Tavasouli (2018) to assess teachers' language knowledge. There were 33 items in the test each carrying 1 point. LAK test was developed six parts, each part focusing on one major area of language assessment, with closed-item formats of 'matching', 'ordering', and 'multiple-choice'. Part A consists of 5 matching questions (questions 1-5) aimed to measure teachers' knowledge on different types of tests. Part B consists of other 5 matching questions (questions 6-10) aimed to measure teachers knowledge on the process of language test design. Part C also includes 5 matching questions (questions 11-14) aimed to measure teachers' knowledge on statistical techniques. Part D consists of 8 multiple-choice questions (questions 15-22) aimed to measure teachers' total assessment knowledge. Part E includes 4 matching questions (questions 23-26). And finally, part F consists of 7 multiple-choice questions to measure teachers' knowledge of designing tests for assessing students' language skills. The instrument enjoyed Cronbach's alpha reliability indices of .53.

Teachers' perception assessment test: this test was used to assess teachers' perceptions about their assessment knowledge. This test consists of three parts. The first part aims to gather participants' demographic information, including their gender, age, and teaching experience. The second part includes 27 items adopted from Alkharusi's (2009) Self-Confidence Scale in Educational Measurement designed to assess teachers' perceptions of confidence in their abilities to perform certain educational assessment tasks related to the following categories developing and administering assessment methods (5 items; e.g., "writing test questions for higher cognitive levels"), developing and scoring performance assessment (8 items; e.g., "Assessing students learning through oral questions"), developing grading procedures (4 items; e.g., "Avoiding bias (personal preferences) in grading") and communicating assessment results to various audiences (5 items; "Providing written feedback to students"). In addition, 5 items on recognizing ethics of assessment (5 items; e.g., "Preventing students from cheating on tests"). And, third part consists of 29 items from Alkharusi's (2010) Teachers' Assessment Practices Questionnaire. It was designed to measure teachers' frequency of use of various assessment practices related to traditional assessment methods (5 items; e.g., "Using multiple-choice tests"), alternative assessment methods (4 items; e.g., "Using structured performance assessments"), communicating assessment with students (6 items; e.g., "Informing students about the purpose of the assessment prior to its administration"), developing assessment standards and criteria (4 items; e.g., "Defining a rating scale for performance criteria in advance"), student-involved assessment (5 items; "Providing students with opportunities to write test questions based on their understanding of the instructional objectives"), and non-achievement grading factors (5 items; e.g., "Incorporating class attendance in the calculation of students' grades"). All the items of this test were rated with Likert five-point scale (1= not competent, 2 = a little competent, 3 = somewhat competent, 4 = competent, 5 = very competent). The instrument enjoyed Cronbach's alpha reliability indices of .91. As noted by Tseng, et al. (2006), Dörnyei & Taguchi (2009), Fryer et al. (2018), and Harrison et al. (2021), a Cronbach's alpha value of .70 is the adequate reliability index for an instrument. Based on this

criterion it can be concluded that the survey questionnaire enjoyed an appropriate reliability index ($\alpha = 0.89$)

The present research method was a descriptive method. According to Hatch and Farhady (1982), descriptive research is a research design used to examine the situation involving the identification of attributes of a particular phenomenon based on an observational basis. So, this type of research is used to describe and interpret the data being studied based on fact that is supported by accurate theories. It means that descriptive research is related to the condition that occurs at that time, uses one variable or more and then investigates in fact.

To conduct the study, an email was first sent out to the participants explaining the number of tools that this thesis was planning to use. It was also stated that the individual's participation in all tasks that this research project would include would yield better results. After sending each instrument to the participants, they were asked to provide their demographic information such as age, gender and years of experience. A pop-up window clarified the purpose behind these questions.

Result & Discussions

This study was undertaken in order to achieve the following objectives. First; it investigated to what extent Iraqi EFL student- teachers were assessment literate. Second; it probed to what extent Iraqi EFL student-teachers viewed themselves as assessment literate; and finally, it explored to what extent Iraqi EFL students-teachers' assessment literacy perceptions matched with their assessment knowledge. The results are discussed below.

Iraqi EFL Student- Teachers Assessment Knowledge

Table 4.3 shows the Pearson correlations among the six parts of the AKT. There were non-significant; i.e. $p > .05$, correlations among all sections except for the correlation between Part E and Part D ($r(28) = .409$ representing a weak effect size, $p < .05$).

Table 4.3
Pearson Correlations between Components of Assessment Knowledge Test

		Part A	Part B	Part C	Part D	Part E	Part F
Part A	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	30					
Part B	Pearson Correlation	.263	1				
	Sig. (2-tailed)	.160					
	N	30	30				
Part C	Pearson Correlation	.163	-.070	1			
	Sig. (2-tailed)	.388	.712				
	N	30	30	30			
Part D	Pearson Correlation	-.154	.152	-.082	1		
	Sig. (2-tailed)	.418	.423	.666			
	N	30	30	30	30		
Part E	Pearson Correlation	-.023	.012	.177	.409*	1	
	Sig. (2-tailed)	.905	.951	.349	.025		
	N	30	30	30	30	30	
Part F	Pearson Correlation	-.091	.328	.266	.048	-.001	1
	Sig. (2-tailed)	.632	.077	.155	.803	.998	
	N	30	30	30	30	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

And finally; non-parametric Friedman's Test was run to compare the median scores on the six parts of the AKT. The non-parametric Friedman's Test was run due to the fact that the assumption of normality was not retained in Part B. As shown in Table 4.4 the ratios of skewness and kurtosis over their standard errors were within the ranges of ± 1.96 except for Part B. It should be noted that the criteria of ± 1.96 were suggested by Raykov and Marcoulides, 2008; Coaley, 2010; Field, 2018; and Abu-Bader, 2021. Table 4.4 also shows the mean ranks on six parts of the AKT which will be discussed below.

Table 4.4

Skewness and Kurtosis Indices of Normality of Assessment Knowledge Test								
N	Mean Rank	Skewness		Kurtosis		Ratios		
		Statistic	Std. Error	Statistic	Std. Error	Skewness	Kurtosis	
Part A	30	3.43	-.356	.427	-.138	.833	-0.83	-0.17
Part B	30	2.00	2.356	.427	6.266	.833	5.52	7.52
Part C	30	3.58	-.362	.427	-1.141	.833	-0.85	-1.37
Part D	30	3.10	.579	.427	.620	.833	1.36	0.74
Part E	30	4.75	-.062	.427	-1.291	.833	-0.15	-1.55
Part F	30	4.13	.144	.427	-.530	.833	0.34	-0.64

Table 4.5 shows the results of the Friedman Test. The results ($\chi^2(5) = 39.19, p < .05$, w coefficient of concordance = .261 representing a weak effect size) indicated that there were significant differences between the mean ranks scores on six parts of the AKT. The mean ranks scores will be compared two by two in Table 4.6.

Table 4.5

Friedman Tests for Assessment Knowledge Test

Total N	30
Test Statistic	39.152
Degree Of Freedom	5
Asymptotic Sig.(2-sided test)	.000

And finally; Table 4.6 shows the post-hoc comparison tests. The results indicated that;

The Iraqi EFL student-teachers had a significantly higher mean rank on Part A (MR = 3.43) than Part B (MR = 2.00) ($Z = 2.96, p < .05$).

The Iraqi EFL student-teachers had a significantly higher mean rank on Part C (MR = 3.58) than Part B (MR = 2.00) ($Z = -3.27, p < .05$).

The Iraqi EFL student-teachers had a significantly higher mean rank on Part F (MR = 4.13) than Part B (MR = 2.00) ($Z = -4.41, p < .05$).

The Iraqi EFL student-teachers had a significantly higher mean rank on Part E (MR = 4.75) than Part B (MR = 2.00) ($Z = -5.69, p < .05$); and finally,

The Iraqi EFL student-teachers had a significantly higher mean rank on Part E (MR = 4.75) than Part D (MR = 3.10) ($Z = -3.41, p < .05$). Figure 4.1 shows the mean ranks on six parts of the AKT.

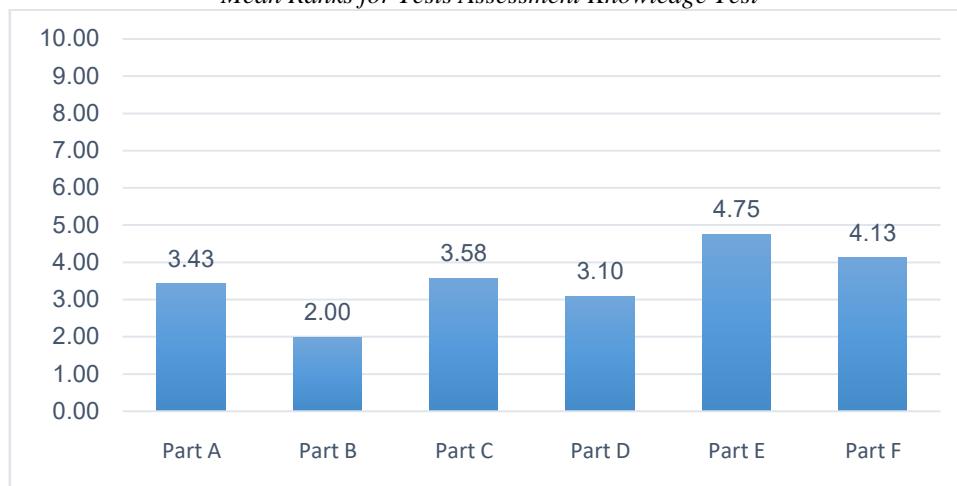
Table 4.6

Pairwise Comparisons Tests Assessment Knowledge Test

Sample 1-Sample 2	Test Statistic	Std. Error	Z	Sig.	Adj. Sig. ^a
Part B-Part D	-1.100	.483	-2.277	.023	.342
Part B-Part A	1.433	.483	2.967	.003	.045
Part B-Part C	-1.583	.483	-3.278	.001	.016
Part B-Part F	-2.133	.483	-4.416	.000	.000
Part B-Part E	-2.750	.483	-5.693	.000	.000
Part D-Part A	.333	.483	.690	.490	1.000
Part D-Part C	.483	.483	1.001	.317	1.000
Part D-Part F	-1.033	.483	-2.139	.032	.486
Part D-Part E	-1.650	.483	-3.416	.001	.010
Part A-Part C	-.150	.483	-.311	.756	1.000
Part A-Part F	-.700	.483	-1.449	.147	1.000
Part A-Part E	-1.317	.483	-2.726	.006	.096
Part C-Part F	-.550	.483	-1.139	.255	1.000
Part C-Part E	-1.167	.483	-2.415	.016	.236
Part F-Part E	.617	.483	1.277	.202	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Figure 4.1
 Mean Ranks for Tests Assessment Knowledge Test



Iraqi EFL Student- Teachers’ Perception of their Assessment Literacy

Table 4.9 shows the Pearson correlations among the five sections of the questionnaire. There were significant; i.e. $p < .05$, correlations among all sections; except for the correlation between ethics in assessment and grading. The highest and lowest correlations were between performance assessment and constructing and administering assessment ($r(28) = .744$ representing a large effect size, $p < .05$), and ethics and grading ($r(28) = .256$ representing a weak effect size, $p > .05$).

Table 4.9
 Pearson Correlations between Components of Assessment Questionnaire

		Constructing	Performance	Grading	Communicating	Ethics
Constructing	Pearson Correlation	1				
	Sig. (2-tailed)					
	N	30				
Performance	Pearson Correlation	.744**	1			
	Sig. (2-tailed)	.000				
	N	30	30			
Grading	Pearson Correlation	.583**	.587**	1		
	Sig. (2-tailed)	.001	.001			
	N	30	30	30		
Communicating	Pearson Correlation	.499**	.543**	.530**	1	
	Sig. (2-tailed)	.005	.002	.003		
	N	30	30	30	30	
Ethics	Pearson Correlation	.395*	.409*	.256	.583**	1
	Sig. (2-tailed)	.031	.025	.171	.001	
	N	30	30	30	30	30

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

And finally; non-parametric Friedman’s Test was run to compare the median scores on the five sections of the questionnaire. The non-parametric Friedman’s Test was run due to the fact that the assumption of normality was not retained. As shown in Table 4.10 the ratios of skewness and kurtosis over their standard errors were higher than ± 1.96 for all sections; except constructing and administering assessment. It should be noted that the criteria of ± 1.96 were suggested by Raykov and Marcoulides, 2008; Coaley, 2010; Field, 2018; and Abu-Bader, 2021. Table 4.10 also shows the mean ranks on five sections of the questionnaire which will be discussed below.

Table 4.10

Skewness and Kurtosis Indices of Normality for Assessment Questionnaire								
	N	Mean Rank	Skewness		Kurtosis		Ratios	
			Statistic	Std. Error	Statistic	Std. Error	Skewness	Kurtosis
Constructing	30	2.15	-.407	.427	.005	.833	-0.95	0.01
Performance	30	2.12	-1.186	.427	3.056	.833	-2.78	3.67
Grading	30	3.30	-1.089	.427	1.108	.833	-2.55	1.33
Communicating	30	3.73	-1.248	.427	.876	.833	-2.92	1.05
Ethics	30	3.70	-.986	.427	-.157	.833	-2.31	-0.19

Table 4.11 shows the results of the Friedman Test. The results ($\chi^2(4) = 32.25, p < .05$, w coefficient of concordance = .256 representing a weak effect size) indicated that there were significant differences between the mean ranks scores on five sections of the questionnaire. The mean ranks scores will be compared two by two in Table 4.12.

Table 4.11

Friedman Tests for Assessment Questionnaire

Total N	30
Test Statistic	32.253
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

And finally; Table 4.12 shows the post-hoc comparison tests. The results indicated that; The Iraqi EFL student-teachers had a significantly higher mean rank on grading (MR = 3.30) than performance (MR = 2.12) ($Z = -2.89, p < .05$).

Table 4.12

Pairwise Comparisons Tests Assessment Questionnaire

Sample 1-Sample 2	Test Statistic	Std. Error	Z	Sig.	Adj. Sig. ^a
Performance-Constructing	.033	.408	.082	.935	1.000
Performance-Grading	-1.183	.408	-2.899	.004	.037
Performance-Ethics	-1.583	.408	-3.878	.000	.001
Performance-Communicating	-1.617	.408	-3.960	.000	.001
Constructing-Grading	-1.150	.408	-2.817	.005	.048
Constructing-Ethics	-1.550	.408	-3.797	.000	.001
Constructing-Communicating	-1.583	.408	-3.878	.000	.001
Grading-Ethics	-.400	.408	-.980	.327	1.000
Grading-Communicating	-.433	.408	-1.061	.288	1.000
Ethics-Communicating	.033	.408	.082	.935	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

The Iraqi EFL student-teachers had a significantly higher mean rank on ethics (MR = 3.70) than performance (MR = 2.12) ($Z = -3.87, p < .05$).

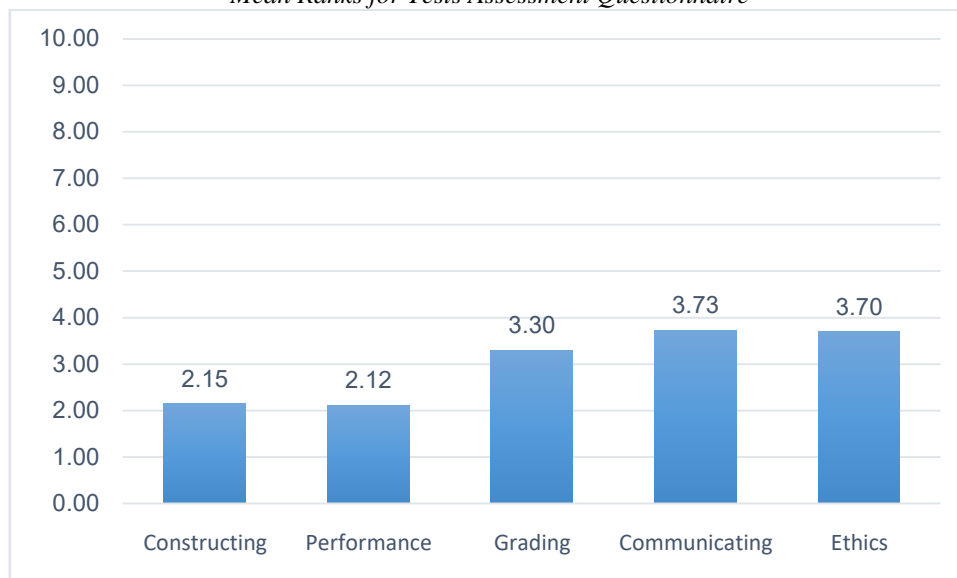
The Iraqi EFL student-teachers had a significantly higher mean rank on communicating results (MR = 3.73) than performance (MR = 2.12) ($Z = -3.96, p < .05$).

The Iraqi EFL student-teachers had a significantly higher mean rank on grading (MR = 3.30) than construction (MR = 2.15) ($Z = -2.81, p < .05$).

The Iraqi EFL student-teachers had a significantly higher mean rank on ethics (MR = 3.70) than construction (MR = 2.15) ($Z = -3.79, p < .05$).

The Iraqi EFL student-teachers had a significantly higher mean rank on communicating results (MR = 3.73) than construction (MR = 2.15) ($Z = -3.87, p < .05$).

Figure 4.2
 Mean Ranks for Tests Assessment Questionnaire



Matched and Mismatches between Iraqi EFL Student-Teachers Assessment literacy perceptions and Their Assessment Knowledge

Pearson correlation was run to probe to what extent Iraqi EFL students-teachers' assessment literacy perceptions matched with their assessment knowledge. The results (Table 4.13) showed that the total scores on AKT and questionnaire had a non-significant correlation ($r(28) = .073$, representing a weak effect size, $p > .05$). The results also showed that Part A of the test was the only variable which had significant correlations with constructing and administering assessment ($r(28) = .525$, representing a large effect size, $p < .05$), performance ($r(28) = .393$, representing a moderate effect size, $p < .05$), ethics ($r(28) = .391$, representing a moderate effect size, $p < .05$), and total questionnaire ($r(28) = .370$, representing a moderate effect size, $p < .05$).

Table 4.13
 Correlations

	Constructing	Performance	Grading	Communicating	Ethics	Total
Part A Pearson Correlation	.525**	.393*	.043	.130	.391*	.370*
Part A Sig. (2-tailed)	.003	.032	.821	.493	.033	.044
Part A N	30	30	30	30	30	30
Part B Pearson Correlation	.015	-.181	-.259	-.059	.158	-.099
Part B Sig. (2-tailed)	.935	.338	.167	.757	.405	.604
Part B N	30	30	30	30	30	30
Part C Pearson Correlation	.182	.309	.149	.268	.187	.286
Part C Sig. (2-tailed)	.335	.096	.431	.152	.322	.125
Part C N	30	30	30	30	30	30
Part D Pearson Correlation	-.337	-.320	-.192	-.202	-.063	-.289
Part D Sig. (2-tailed)	.068	.084	.310	.285	.740	.121
Part D N	30	30	30	30	30	30
Part E Pearson Correlation	.014	.045	.237	-.130	-.003	.043
Part E Sig. (2-tailed)	.941	.815	.206	.493	.989	.823
Part E N	30	30	30	30	30	30
Part F Pearson Correlation	-.022	.097	-.040	.143	.243	.107
Part F Sig. (2-tailed)	.909	.609	.834	.450	.196	.573
Part F N	30	30	30	30	30	30
Total Pearson Correlation	.041	.041	-.030	.003	.258	.073
Total Sig. (2-tailed)	.829	.828	.876	.988	.169	.701
Total N	30	30	30	30	30	30

- ** . Correlation is significant at the 0.01 level (2-tailed).
- * . Correlation is significant at the 0.05 level (2-tailed).

This finding is in line with the findings of previous related studies. Since assessment knowledge as part of teachers' professional knowledge is essential for successful teaching (Farhady & Tavassoli, 2018; Tavassoli & Farhady, 2018; Popham, 2011), it seems vital for supervisors, teacher educators, and policy-makers to provide ample opportunities for teachers to improve their language assessment knowledge (Popham, 2009). According to Olmezer-Oztürk and Aydin (2019), the primary reason for the low level of EFL teachers' language assessment knowledge is most likely the insufficiency of teacher education courses. Since teachers do not receive sufficient training on assessment, their knowledge is restricted as a result (Olmezer-Oztürk & Aydin, 2019). The findings of this study were consistent with those of Jannati (2015), who found that teachers had an understanding of the fundamental principles underlying language assessment. Despite the fact that teachers had this knowledge of assessment, however, they were unable to apply it in their classrooms. As an illustration, they were unable to apply their understanding of things like dependability, validity, fairness, and sincerity into reality. According to the findings of the study's semi-structured interviews, however, just providing educators with information on language testing does not appear to be an adequate method of achieving the desired objectives. Formative assessment and classroom settings, in particular, need for teachers to comprehend and acquire the skills necessary to effectively use their prior language content knowledge in their classroom instruction (Popham, 2009). According to Olmezer-Oztürk and Aydin (2019), this may be feasible through the implementation of continuous training courses on assessment in which testing professionals instruct instructors on how to educate and work on language assessment knowledge through the use of hands-on activities with the students. Because accurate evaluation is a potent tool for education, these classes have to be structured to enhance the experience of instructors in evaluating linguistic abilities (Popham, 2011). It is possible that attending training sessions alone will not result in increased language assessment knowledge levels. According to Olmezer-Oztürk and Aydin (2019), it is preferable and more successful for educators to participate in comprehensive training programs that are delivered over an extended period of time and are conducted by specialists in the field of language testing and assessment. In order to aid their students in becoming better language learners, instructors of English as a foreign language (EFL) need to become proficient in the creation of relevant assessments and the application of such examinations in the classroom. Due to the fact that experienced educators do not have a strong understanding of evaluation, they are unable to assist novice educators (Popham, 2011). Therefore, participating in professional development opportunities is the most viable choice that is still open to educators. If instructors are given the resources to complete assessments and are made aware of the significance of evaluation in the overall instructional process, then the evaluation process may become something that is pleasurable for both the teachers and the students. More significantly, it appears essential for many groups of stakeholders, such as professionals, educators, administrators, supervisors, policy-makers, students, and parents, to become involved in the process and collaborate in order to change the current situation (Kremmel & Harding, 2020; Yan et al., 2017).

This study, in line with Jannati (2015), Olmezer-Oztürk and Aydin (2019), and Vogt et al. (2020), was one of the few studies in the literature in which the teachers' voice about their language assessment knowledge was investigated by collecting information about various aspects of EFL teachers' language assessment knowledge. This study was unique in that it was one of the few studies in which the teachers' voice about their language assessment knowledge was investigated. The most important thing that came out of this study was the demonstration that EFL student- teachers with low levels of language assessment knowledge were unable to implement proper assessment techniques in their classrooms. Equally as important, it demonstrated that there is a significant divide between teaching methods and assessment practices in the setting of the classroom, which stands in contradiction to the new ideas of language assessment. It would appear that teachers are not very familiar with recent developments in the field of assessment, such as learning-oriented assessment, which is an approach in which assessment, instruction, and learning are all integrated, and teachers are supposed to be at the center of assessment practices in the classroom (Purpura & Turner, 2014). EFL teachers should have sufficient knowledge of language assessment (Olmezer-Oztürk & Aydin, 2019). Assessment is an unavoidable component of the teaching-learning process, and teachers are at the center of all assessment-related activities, including the writing of items, the construction of tests, and the provision of scores.

The data indicated that there was a general deficiency in assessment literacy of participants. Therefore, there was a contrast and mismatch between teachers' perceptions of their assessment knowledge and their actual level of knowledge. In other words, they were not as competent in assessment knowledge as they thought.

This study made a contribution to the field of EFL assessment as well as educational assessment by responding to calls for the development of a new measure to assess teacher assessment literacy.

The findings of this study indicated that teacher assessment literacy is limited, and that it still has its deficits. These findings supported the findings of several other studies that have been conducted within the last two decades. Although it has been shown in this and other studies that pre-service training has a significant effect on assessment literacy, future research may need to investigate the objectives, tasks, contents, and outcomes (essentially the effectiveness) of assessment courses provided by teacher preparation institutions and professional development programs.

The results of this study have the potential to be reflective, in particular, of the current status of assessment literacy among EFL teachers in Iraq. The ramifications of this study, on the other hand, may be applied to a variety of various stakeholders who are involved in education in general and EFL across the world since they concern teacher training programs and teaching institutions. The results of the research showed that pre-service assessment courses had a considerable influence on EFL teachers' assessment ability and practice. As a result, the findings of this research suggest that EFL preparation programs should continue to offer courses on evaluation to language teachers in order to improve the assessment literacy of those teachers.

It was necessary to conduct a reassessment of assessment literacy since much has changed in the field of language assessment theories, teacher credentials, and assessment training since Plake and Impara (1992). Despite this, it appears that teachers still have a significant lack of assessment literacy in several crucial domains. As a result, this research recommends conducting an exhaustive analysis of the current provisions of teacher preparation programs with regard to assessment in general and language assessment in particular. Additionally, it suggests that consideration be given to the urgent needs for reshaping these programs to expand their theoretical as well as practical assessment courses to encompass the various assessment domains highlighted by this research.

To fully understand classroom assessment literacy, future research should address methodology and sample population concerns. Methodologically, future research should track teachers' classroom assessment literacy growth. A longer-term longitudinal study of teachers' classroom assessment literacy may reveal its growth. Future research should also evaluate the link between classroom assessment literacy and practice by monitoring teachers as they prepare and conduct assessments in their classrooms. To understand teachers' classroom assessment literacy, such studies could examine their quizzes, tests, exams, assignments/projects, and written feedback to students. Future study should employ random sampling across schools and colleges, especially with bigger sample sizes, to generalize findings. In addition to instructors, curriculum developers, lawmakers, and administrators are all significant educational decision makers regarding assessment outcomes, therefore additional research should include them. Recently, it has been proposed that teachers, policymakers, and administrators must be assessment-literate to properly perform their duties and responsibilities (Popham, 2014). Quality assurance, reporting and disseminating assessment results, and using assessment data for evaluation/accountability are examples.

Instructors, students, curriculum creators, policymakers, and administrators must become classroom assessment-literate to enhance classroom-based assessment and utilization of assessment findings. According to Popham (2014), a reasonable chunk of assessment literacy is helpful for practically everyone" who implements assessments or uses assessment findings for educational decision-making.

References

- [1]. Abu-Bader, S. H. (2021) *Using statistical methods in social science research: With a complete SPSS guide*. Oxford University Press, USA.
- [2]. Al-Jabbawi, M. (2022). Designing a Training Program for Iraqi EFL In-service Teachers. *Journal of the University of Babylon for Humanities*, 30(2).
- [3]. Alkharusi, H. A. (2009). Correlates of Teacher Education Students' Academic Performance in an Educational Measurement Course. *International Journal of Learning*, 16(2).
- [4]. Alkharusi, H. (2010). A multilevel linear model of teachers' assessment practices and students' perceptions of the classroom assessment environment. *Procedia-Social and Behavioral Sciences*, 5, 5-11.
- [5]. 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.
- [6]. 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.
- [7]. Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Educational assessment profile of teachers in the Sultanate of Oman. *International Education Studies*, 7(5), 116.
- [8]. Alsarimi, A. M. (2000). Classroom assessment and grading practices in the Sultanate of Oman. (Unpublished Doctoral dissertation). University of Pittsburgh, Pennsylvania.

- [9]. 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.
- [10]. Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30 (1), 3–12.
- [11]. Brown, G. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318. <http://dx.doi.org/10.1080/0969594042000304609>
- [12]. Campbell, C., Murphy, J. A. & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- [13]. Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7), (pp. 349-364). New York: Springer.
- [14]. Coaley, K. (2010). An Introduction to Psychological Assessment and Psychometrics.
- [15]. Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. *The Cambridge guide to second language assessment*, 1, 20-29.
- [16]. 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.
- [17]. DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- [18]. Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford, England: Oxford University Press.
- [19]. 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. <https://doi.org/10.30466/ijltr.2018.120602>
- [20]. Field, A. (2018). *Discovering Statistics Using IBM SPSS, Statistics for Statistics*. (5th ed.). London: SAGE Publications.
- [21]. Firoozi, T., Razavipour, K., and Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(1), 1-14
- [22]. Fryer, L. K., Larson-Hall, J., & Stewart, J. (2018). Quantitative methodology. In *The Palgrave handbook of applied linguistics research methodology* (pp. 55-77). Palgrave Macmillan, London.
- [23]. Fulcher, G. (2010). *Practical language testing*. London, England: Hodder Education.
- [24]. Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- [25]. Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge.
- [26]. Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Taylor & Francis.
- [27]. Gronlund, N.E & Linn, R.L. (1990). *Measurement and evaluation in teaching* (6th edn). New York: Macmillan Publishing Company.
- [28]. Harrison, R., Jones, B., Gardner, P., & Lawton, R. (2021). Quality assessment with diverse studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed-or multi-method studies. *BMC Health Services Research*, 21(1), 1-20.
- [29]. Hatch, E., & Farhady, H. (1982). Research design and statistics for applied linguistics.
- [30]. Hatipoğlu, C. (2015). “English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers.” *ELT Research Journal*, 4 (2), 111-128.
- [31]. Inbar-Lourie, O. (2008b). Language assessment culture. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopaedia of Language and Education* (2nd ed., Vol. 7). (pp. 285-299). New York: Springer
- [32]. Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26–37.
- [33]. Kaya, T. (2018). *Exploring language assessment literacy of EFL instructors in language preparatory programs* (Unpublished master's thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>.
- [34]. Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>
- [35]. Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
-

- [36]. Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344.
- [37]. Mede, E., Atay, D. (2017). “English language teachers’ assessment literacy: The Turkish context.” *DilDergisi*, 168 (1), 1-5.
- [38]. Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120, 285-296.
- [39]. Mertler, C. A. (2000). Teacher-centered fallacies of classroom assessment reliability and validity. *Mid-Western Educational Researcher*, 13(4), 29-35.
- [40]. Mertler, C. A. (2003). Pre-service versus in-service teachers' Assessment literacy: Does classroom experience make a Difference?. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus.
- [41]. Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113.
- [42]. Mertler, C. A., & Campbell, C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- [43]. Messick, S. (1989). Validity. In R. L. Linn (Ed), *Educational measurement* (3rd ed.), (pp. 13-103). New York: MacMillan.
- [44]. 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
- [45]. Muhammad F., & Bardakci, M. (2019). Iraqi EFL Teachers’ Assessment Literacy: Perceptions and Practices, *Arab World English Journal (AWEJ)*, Volume 10. Number 2. June 2019 Pp. 431-442 DOI: <https://dx.doi.org/10.24093/awej/vol10no2.33>
- [46]. Ölmezer-Öztürk, E., & Aydin, B. (2019). Voices of EFL teachers as assessors: Their opinions and needs regarding language assessment. *Eğitimde Nitel Araştırmalar Dergisi - Journal of Qualitative Research in Education*, 7(1), 373–390. <https://doi.org/10.14689/issn.2148-2624.1.7c1s.17m>
- [47]. Ölmezer-Öztürk, E., & Aydin, B. (2018). Toward measuring language teachers’ assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia*, 8, 1-15.
- [48]. O’Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380.
- [49]. Oz, S., & Atay, D. (2017). Turkish EFL instructors’ in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- [50]. Perry, M. (2013). *Teacher and Principal Assessment Literacy*. The University of Montana
- [51]. Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402.
- [52]. Plake, B. S. (1993) Teacher assessment literacy: teachers’ competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(2), 21–7.
- [53]. Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12.
- [54]. Plake, B. S., & Impara, J. C. (1992). Teacher competencies questionnaire description. *Lincoln, NE: University of Nebraska*, 312.
- [55]. Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63(6), 84–85.
- [56]. Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>.
- [57]. Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator’s confession. *The Teacher Educator*, 46(4), 265–273. <https://doi.org/10.1080/08878730.2011.605048>.
- [58]. Purpura, J. E., & Turner, C. E. (2014, October). *A learning-oriented assessment approach to understanding the complexities of classroom-based language assessment*. Paper presented at the Roundtable on learning-oriented Assessment in Language Classrooms and Large-Scale Contexts. Teachers College: Columbia University, New York.
- [59]. Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge.
- [60]. Razavipour, K., Riazi, A. & Rashidi, N. (2011), On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers, *English Language Teaching* 4(1), DOI: [10.5539/elt.v4n1p156](https://doi.org/10.5539/elt.v4n1p156)
- [61]. Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing* 30(3), 309-327.

- [62]. Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72 (7), 534–539.
- [63]. Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- [64]. Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of the EFL teachers: Scale construction and validation. *Language Testing*, X, 1-30. <https://doi.org/10.1177/02655322211057040>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36
- [65]. Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412.
- [66]. Tavassoli, K., & Farhady, H. (2018). Assessment knowledge needs of EFL teachers. *Teaching English Language*, 12(2), 45-65.
- [67]. Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6, 41–63.
- [68]. Tseng, W. T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78-102.
- [69]. 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.
- [70]. 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.
- [71]. Yan, X., Fan, J., & Zhang, C. (2017). Understanding language assessment literacy profiles of different stakeholder groups in China: The importance of contextual and experiential factors. In *language testing research colloquium, Bogotá, Colombia* (Vol. 257).
- [72]. Alkharusi, H. (2010). A multilevel linear model of teachers' assessment practices and students' perceptions of the classroom assessment environment. *Procedia-Social and Behavioral Sciences*, 5, 5-11.
- [73]. Alkharusi, H. A. (2009). Correlates of teacher education students' academic performance in an educational measurement course. *International Journal of Learning*, 16(2) 1-16.
- [74]. Sheehan, S., & Munro, S., (2017). Assessment: attitudes, practices, needs. In: BAAL Language Learning and Teaching SIG 2017, 67 July 2017, University of Central Lancashire, Preston. (Unpublished)