How Literate are Prospective English Teachers in Language Assessment? A Case Just Before Entering the Teaching Profession

Mehmet ABİ *a 1 🆔, Şeyda Selen ÇİMEN a 2 🆔, Şevki KÖMÜR a 3 🆔

a Muğla Sıtkı Koçman University, Muğla, Türkiye

Abstract

Assessment is defined as 'an essential component of all learning and teaching endeavors'. Training prospective teachers in assessment is of primary importance in teacher education. This study employs a mixed methods design to uncover (a) how literate prospective English teachers in language assessment based on scenarios that include some instances of assessment applications, (b) how they perceive their knowledge of language assessment, and (c) how they apply their knowledge when developing language tests. Participants are the fourth-year prospective English teachers who took the course Testing and Evaluation in ELT. The data for the study were collected through open-ended questionnaires and scenarios designed to reflect how successfully the participants transferred their knowledge into practice. The results reveal that assessment courses fail to achieve their objectives in the practical sense. Additionally, the inability to obtain high levels of language assessment literacy might result from cognitive errors which individuals have developed throughout their education.

Introduction

---

1 Corresponding author.
E-mail address: mehmetabi@mu.edu.tr
2 *E-mail address*: s.selen.ozkan@gmail.com
3 *E-mail address*: coal@mu.edu.tr

It is commonly agreed upon that assessment is vital to language teaching and learning practices (Ashraf & Zolfaghari, 2018). Whether its aim is the process, progress, or the product of learning, various assessment procedures are employed to provide data for all stakeholders regarding the success of those practices. To obtain reliable results from any test, one must be able to control the numerous variables that influence the test's quality and outcome. When all parties that play a significant role in assessment procedures are considered, classroom teachers have the most substantial role since they are typically the practitioners who prepare, administer, evaluate, and share the results. Therefore, it would be correct to assert that becoming competent as a practitioner requires amassing a substantial amount of knowledge that can be used to make sound, systematic decisions (Koh, 2011). Therefore, enhancing the assessment literacy of prospective and/or in-service teachers should be a priority.

The practical application of assessment knowledge is not devoid of problems or criticism. Thus, some researchers have attempted to identify the source where prospective teachers are expected to acquire the necessary knowledge. It can be stated that studies conducted on teacher preparation courses with regard to classroom assessment practices do not match the actual needs for classroom practice (Xu & Brown, 2016). What is taught in university assessment courses can be insufficient to prepare pre-service teachers for actual field assessment practices adequately. Researchers such as Jawhar and Subahi (2020), Graham (2005), and Beziat and Coleman (2015) specify that teachers have been unable to improve their assessment knowledge and skills sufficiently at the tertiary level; therefore, as Latif (2021) suggests, more attention should be paid to identify the assessment literacy of prospective language teachers in university contexts. As Latif (2021) states, the disparity and deficiencies between university education for language teachers and professional school teachers might be a serious concern.

There is a need to establish a link between theoretical considerations at universities and practical applications in classrooms (Coombe et al.,2020). To narrow the gap and increase the success rate of university assessment courses, Yan and Fan (2020) remark that trainers should consider contextual factors and develop realistic expectations about Language Assessment Literacy (LAL) development. In that way, it could be possible to claim that theories and skills might best be reflected in the assessment practices in classroom environments. Therefore, this study aims to explore the level of language assessment

literacies of preservice language teachers and to what extent they can reflect their language assessment literacies on assessment scenarios and test developments.

Throughout literature, it could be observed that many language education researchers have conducted studies on assessment literacy. Some of these researchers have aimed to define LAL. Among them, Mohebbi and Coombe (2020) state that "[it] includes skill in the procedures for evaluating language abilities, the ability to construct sound and suitable assessments, and, just as importantly, the understanding of appropriate mechanisms of feedback that assist learners to improve their target language and reach their short- and long-term goals" (p.1). In another study, Giraldo (2018) denotes that language assessment literacy consists of skills (instructional, design, measurement, and technological), knowledge (awareness of applied linguistics, theory and concepts, and context), and principles (awareness of and actions towards critical issues). From another perspective, Quilter and Gallini (2002) theorize assessment literacy as a fundamental factor affected by teachers' attitudes towards assessment. In contrast, Heitink, Kleij, Veldkamp, Schildkamp, and Kippers (2016), Wu, Zhang, and Liu (2021), and Davison (2019) consider it as an integral part of the assessment for learning practices. Stiggins (1995) points out that it covers knowing the purpose of the assessment, how it is administered, and developing the ability to identify the potential problems that could emerge throughout the assessment procedure and how to overcome those problems. That is, when teachers become assessment literate, they are equipped with the necessary knowledge and tools (Djoub, 2017) and become competent in applying the appropriate methodology in terms of preparation, administration, and dissemination of the results (Zulaiha, Mulyono & Ambarsari, 2020).

Some researchers also try to find answers to questions such as how to become literate in assessment. In this respect, Xu and Brown (2016) have developed a framework called 'Teacher Assessment Literacy in Practice (TALiP)' as a conceptual framework of assessment literacy. Likewise, Taylor (2013) enumerates that language assessment literacy comprises eight core components: knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practice, personal beliefs and attitudes, and scores and decision-making. High competencies in these dimensions can indicate a high level of assessment literacy. By the same token, Lee and Butler (2020), who investigated several LAL models, conclude that to become LAL, one should develop adequate knowledge of theories related to LAL and the role of different contexts; in addition, individuals need

upgrade their practical skills regarding the development and interpretation of assessment; and finally, social consequences of assessment must be understood.

There are researchers such as Vogt, Tsagari, and Spanoudis (2020), Yan and Fan (2020), and Levi and Inbar-Lourie (2020) who focus on the role of contextual factors while trying to explain LAL. Yan and Fan (2020) and Yan, Zhang, and Fan (2018) argue that language assessment literacy has a social dimension and cooperative construct in which several stakeholders play a role. Therefore, Yan and Fan (2020) add that LAL is not merely accumulating knowledge and skills but an interactive process in which contextual and experiential factors act together. Similarly, Levi and Inbar-Lourie (2020) point out that LAL embraces generic, language-specific, and contextual factors. With reference to their research, Yan and Fan (2020) have developed 'an apprenticeship-based, experience-mediated model' of language assessment literacy development (p. 238), in which they examine the impact of assessment context and experience by highlighting the importance of collaborative and reflective assessment practices. Another LAL model has been developed by Davies (2008), which comprises components of knowledge, skills, and principles. In the same vein, Fulcher's (2012) LAL model consists of layers of context, principles, and practice, which provide evidence of the role of context. In this respect, it could be said that language assessment literacy is context-bound. As Coombe, Vafadar, and Mohebbi (2020) expound, various components of teacher assessment literacy are interrelated. Lee and Butler (2020) also acknowledge the interconnected construct of language assessment literacy and the role of primary stakeholders such as teachers. However, they criticize overlooking learners' role in language assessment procedures and argue "… the importance of incorporating learners' perspectives on language assessment in the conceptualization of LAL" (p.1099). Vogt, Tsagari, Csépes, Green, and Sifakis (2020) are among the few researchers who claim that learners are one of the main stakeholders in assessment procedures. Therefore, they should constitute a part of LAL.

There are a number of studies on the benefit of becoming literate in assessment in language classrooms. The findings from the research indicate that becoming assessment literate will contribute to the quality of the assessment procedures (Coombs, DeLuca, LaPointe-McEwan, & Chalas 2018). Gürsoy (2017), for example, signifies that when teachers cannot improve their assessment literacy levels, this will result in malpractices in language classrooms, significantly impacting the assessment quality. Gürsoy (2017) emphasizes that it is not sufficient for a teacher to know assessment; they have to develop

their competencies in assessment literacy and should know how to apply that knowledge for practical purposes if quality is the desired outcome. From a similar perspective, Latif (2021) points out that teachers can administer more comprehensive assessment practices when they have adequate assessment literacy. In addition, Boylu (2021) sees assessment literacy as one of the most fundamental factors affecting the evaluation of the success of foreign language teaching practices. That is, with the successful integration of assessment in teaching, an assessment-literate teacher can significantly contribute to the achievement levels of the students (McMillan, 2003).

As it is understood, assessment-literate language teachers in real classrooms who have improved their assessment literacies can contribute to achieving language teaching and learning goals in classrooms. However, when real classrooms are put in the center, there are various points of view by researchers. Researchers such as Alderson (2005), Stiggings (1991), Metler (2003), Ölmezer-Öztürk and Aydın (2019), Popham (2006), Rahman (2018), Yeşilçınar and Kartal (2020) argue that the situation in language classrooms is not at the desired level. In other words, as Sevimel-Şahin and Subaşı (2021) and Shim (2009) indicate, most teachers have problems applying their assessment knowledge to practice. Vogt and Tsagari (2014), Latif (2017), and Al-Bahlani (2019) have similar opinions and indicate the inadequacy of knowledge and understanding of assessment literacy of language teachers.

On the other hand, Watmani, Asadollahm, and Behin (2020) handle the problem relatively gently and indicate that language teachers are not unaware of all assessment procedures. Accordingly, they argue that teachers can choose the appropriate assessment methods and have developed an acceptable level of competence in using assessment results for decision-making. However, Watmai et al. (2020) add that language teachers need to be more knowledgeable in terms of other components of assessment literacy. That is, they know the basics but need help to ameliorate their literacy levels in a way that integrates all dimensions of assessment literacy. A parallel notion is held by Öz and Atay (2017) and Yeşilçınar and Kartal (2020), who say that although in-service teachers feel they have enough knowledge on assessments, they have difficulty putting them into practice in their classrooms. Their finding leads us to make a deduction: knowing something and putting it into practice are two sides of a coin, and when one has a problem, the other is inevitably stuck into that problem.

On the other hand, Vogt and Tsagari (2014) consider another dimension: familiarity with the assessment types by language teachers. They argue that language teachers tend to utilize the assessment types they are more familiar with from their previous experiences in practice than the others, such as informal or alternative assessment types. Another finding of Vogt and Tsagari's (2014) study is that language teachers who have taken part in their research complain about the quality of their education at universities. For them, their unsatisfactory assessment practices in classrooms result from inadequate training.

Some researchers approach the issue from a distinct perspective, too. For example, Hatipoğlu (2015), Hatipoğlu and Erçetin (2016), and Latif (2021), who have carried out a study on pre-service language teachers' assessment knowledge focus on the source of the problem by criticizing tertiary-level practices and add to the argument by quoting even at university level pre-service teachers obtain a limited amount of knowledge in terms their assessment literacy, which highlights the inadequacy of university courses on assessment and put it another way, the failure of the education in terms of attaining their objectives. Some researchers adopt a relatively positive standpoint. To exemplify, in their study, Zulaiha et al. (2020) argue that EFL teachers not only have a good amount of assessment knowledge. However, they are also good at putting the planning, implementation, monitoring, recording, and dissemination stages of assessment procedures into practice. However, even Zulaiha et al. (2020) question the quality of classroom assessment practices.

Considering the literature mentioned above review, this study investigates pre-service English language teachers' perceptions of their knowledge in testing and assessment practices. Another aim of the study is to determine how well pre-service teachers are ready to apply their testing and assessment knowledge in connection with given scenarios. Finally, it scrutinizes the consistency between pre-service language teachers' perceptions and their ability to apply their knowledge in different assessment settings. In this respect, the following research questions are formed:

1. How do pre-service English language teachers perceive their knowledge in testing and assessment practices?
2. How successful are the pre-service language teachers in applying their testing and assessment knowledge to different assessment (scenarios) contexts?
3. How competent are prospective teachers in developing achievement tests?

Method

This study employs a mixed-method study methodology to address the "how" questions it poses to understand the phenomenon of assessment literacy and knowledge of pre-service English teachers. The qualitative and quantitative data were collected and analyzed independently. After that, the findings were brought together, and interpretations were generated (McCrudden & Rapp, 2023). In this respect, as Creswell and Clark (2018) suggest, the current study employs a convergent concurrent mixed-method design.

Participants

This study was carried out in a state university with the participation of 57 pre-service teachers attending the English Language Teaching Department and taking the Testing and Evaluation course *in ELT*. Students were invited to take part in the study voluntarily. In the scope of this course, students are offered theoretical and practical aspects of language assessment and testing. During the course, they are introduced to the fundamental concepts for the first seven weeks. Then, they are requested to prepare a test using the knowledge they have attained during their studies, which include test specifications, the test, an answer key and a rationale in which they explain how they utilized their knowledge. All the participants were informed about the study and the procedure before consented. The participants were four-year preservice English teachers in the last semester of their university education. In addition to this course, they were doing their teaching practice in schools. Up to the study period, they had already taken some compulsory and elective courses such as Language Acquisition, Methods and Approaches in ELT, Teaching English Language Skills, and Materials Development. Thus, it would not be wrong to state that all participants were expected to have become competent enough not only in terms of language testing and assessment but also in language teaching.

Data Collection Tools and Procedure

In this study, the data were collected in three ways. Initially, reflection paper was used to collect data to diagnose participants' perceptions and knowledge regarding language testing and assessment. Then, a scenario-based assessment, adopted from Brown and Abeywickrama (2010) and comprising six scenarios, was administered to evaluate how successfully participants transferred their knowledge into practice. To this end, participants were asked to examine the scenarios regarding several fundamental assessment components

such as validity, reliability, authenticity, washback, and so on. Finally, participants were asked to prepare a language test with the given specifications.

In the study's first phase, participants were given five open-ended questions to discover their knowledge in testing and assessment practices in English language teaching and learning environments. In the first question, participants were asked to identify and interpret the contemporary testing tools and methods. The aim was to discover whether they still tended to accept the conventional ways of testing and assessment tools as dominant tools or not. In addition, they were asked to explain the tools and methods briefly to figure out their level of knowledge. The second question aimed to reveal the students' competence regarding their practical knowledge. With this aim, they were asked to indicate their views of the required qualifications and skills for ideal language testing. Then, the following question incorporated the role of technology in testing. Questions four and five were formed to seek out participants' self-awareness in language testing and their perceptions of areas with development potential concerning language testing in which they think they should improve themselves.

In the second phase, participants were given six scenarios adapted from Brown and Abeywickrama (2010, pp. 48-50) and designed to assess prospective teachers' knowledge of the key concepts and their applications in ELT settings. Each scenario had some powerful and weak sides in terms of essential principles of evaluation in language tests. Participants who had taken *Language Assessment and Testing* classes in the previous semester were supposed to accurately identify these six testing scenarios' strengths and weaknesses. In this vein, participants were asked to evaluate (low, medium, or high) the six scenarios according to the basic testing principles: reliability, validity, practicality, wash-back, and authenticity. Then, participants' papers were collected and graded by experts in terms of the accuracy of their answers. For each correct answer, participants were given one point. That is, for each scenario, a participant could get a grade between 0 and 5, and the maximum overall score they could get was 30. In this way, each participant's overall score was obtained at the end of the evaluation process, indicating their proficiency level in language assessment competencies.

In the third phase of the study, the participants were tasked with developing an achievement test to assess language skills and related subskills to be used in the schools during their practicums. In this way, it was aimed to give them a real sense of developing a

test for the target group. The tests were evaluated according to the test specifications given to the pre-service teachers. Ideally, participants were expected to prepare tests that were appropriately designed. That is, they were to develop valid and reliable tests that can be utilized in language classrooms efficiently. They were all expected to utilize authentic materials and texts and pay attention to the evaluation. Finally, pre-service teachers were assumed to attach the test specifications to their tests.

Data Analysis

Having completed the data collection part, participants' responses to open-ended questions were analyzed by carrying out (content) analysis. First, the students' responses were analyzed by two ELT experts, who prepared a report. Then, another expert checked the consistency of the data analysis. After that, the authors tried to reach a conclusion based on individual answers and scores and the participants' overall performance. They intended to explore a pattern that would have helped them to interpret the results and reach a conclusion. Since the analysis did not compromise any statistical or detailed content analyses, there was no need to make use of any data analysis software. So, all the calculations were made manually. According to the results obtained, the participants were grouped into four groups, and further analyses were performed by taking these four groups to determine whether they represented any unique characteristics.

Results

First, Questions 1, 2, and 5 aimed at discovering their knowledge of contemporary testing tools (question 1), of other skills and qualifications necessary for ideal testing (question 2), and of the suggested areas for self-improvement (question 5) were analyzed. Findings revealed that 56 participants replied to question 1. It was seen that participants mentioned portfolios (n=50) as the most popular contemporary testing tool. It was followed by peer/self-assessment (n=28), performance tasks (n=20), projects (n=24), and grid testing (n=6) as the standard contemporary testing tools.

Analysis of participants' replies to the second question (What are the other required skills and competencies besides language skills to carry out ideal testing and evaluation?) showed that 39 out of 57 participants gave appropriate answers to this question. Accordingly, almost half of the participants knew that they needed to be able to prepare tests that accurately measured what they were supposed to measure. Acceptable tests meant

consistency over time and between different participants (validity and reliability, 39%). Moreover, they pointed out that a teacher needed to know their students before testing (knowing students, 13%) and had to abide by ethical codes (ethics, 17%). Another issue that came out was that a teacher needed to be able to prepare practical tests (practicality, 11%) and good answer keys (answer key, 7%). Besides, the other required skills were having enough content knowledge (content knowledge, 6%) and working together with their colleagues while preparing tests (collaboration, 7%).

Then, participants' self-evaluation of themselves regarding their self-awareness and level of knowledge in testing were examined in Question 4. When the data were investigated, it was seen that some students thought their self-efficacy levels were relatively high. They saw themselves as competent enough to prepare and administer proper tests. On the other hand, some other participants indicated that although they had studied testing previously, they did not believe that their competencies were not high enough to enable them to prepare and administer proper tests. Based on this finding, it could be stated that there are two major groups of participants regarding pre-service perceptions of their language testing and assessment knowledge. That is, it was possible to group participants into two distinct groups: competent and incompetent.

Participants' answers to Question 5 were examined in the final part of this phase. In this question, participants were asked to indicate areas where they thought they should improve their testing knowledge. 40 out of 57 participants replied to this question. Findings showed that almost half of the participants needed further training in testing and evaluation (51%). That is, although they had studied language testing for a semester and improved their knowledge, they did not feel secure enough to prepare tests for actual use in schools. Nine participants reflected they needed to improve their knowledge of testing four skills (19%). Since they were aware of the fact that productive skills were hard to manage in terms of testing. Four participants pointed out that they needed to improve their abilities to prepare tests suitable for their students' levels (9%). Another issue that had been argued by three students as a potential area for self-improvement was 'grading' (7%). A small number of participants considered themselves as incompetent in this area. Another three participants stated that they also needed to learn more about the alternative ways/instruments of assessment (6%). Finally, other areas suggested by participants were finding distractors (2%), giving feedback (2%), preparing answer keys (2%), and fairness (2%).

In respect to Research Question 1, not all participants have at least an average level of knowledge. In other words, participants know the basics of the language testing assessment at varying degrees but not at acceptable levels. In addition, they can classify discrete types of contemporary testing tools, name the required skills and qualifications, and identify their weak areas in language testing and assessment. However, only a limited number of participants can do it effectively. Moreover, according to findings obtained from Question 4, which focused on pre-service teachers' perceptions about their own language testing and assessment competencies, it was found that not all participants had perceived themselves similarly.

After completing the analysis of the open-ended questions, data from the scenarios were examined in the next step. In the first scenario, participants were supposed to evaluate a standardized multiple-choice proficiency test. The second scenario was a timed impromptu test of written English (TWE Test). Then, in the third scenario, they were asked to evaluate one-on-one oral interviews to assess overall production ability, whereas they were expected to rate a five-minute prepared oral presentation as the fourth one. In the fifth scenario, participants evaluated the scenario, which included multiple drafts of a three-page essay, which included a peer-reviewed one; finally, in the last one, they were supposed to grade a portfolio of materials collected throughout a course.

Table 1. Key Concepts: Application Activity

Scenario	Content
Scenario 1	Standardized multiple-choice proficiency test, no oral or written production. Students receive a report that includes a total score and subscores for listening, grammar, proofreading, and reading comprehension.
Scenario 2	Timed impromptu test of written English (TWE Test). Students receive a report listing one holistic score ranging between 0 and 6. No additional comment is provided.
Scenario 3	One-on-one oral interview to assess overall production ability. Students receive one holistic score ranging between 0 and 5. No additional comment is provided.
Scenario 4	Student gives a five-minute prepared oral presentation in class. Teacher evaluates by filling in a rating sheet indicating Student's success in delivery, rapport, pronunciation, grammar, and content. The teacher uses a presentation rubric which describes each performance factor at three levels of proficiency: Very Good, Good, and Needs Improvement. Students receive a copy of the rubric as part of the presentation.
Scenario 5	Student creates multiple drafts of a three-page essay. Early drafts are peer-reviewed. Student turns in a near-final version to the teacher. Teacher comments on grammatical/rhetorical errors only and returns it to student (no grade).
Scenario 6	Student assembles a portfolio of materials over a semester-long course. Teacher conferences with student on the portfolio at the end of the semester, assigning an overall grade.

Brown and Abeywickrama (2010, pp. 48-50)

This scenario-based evaluation was expected to identify whether participants' perceptions about their testing and evaluation competencies were consistent with their testing knowledge. However, before starting, participants were assigned to two previously identified groups. Thus, participants who had stated that they felt themselves competent formed the "Competent" group (n = 26), whereas the ones who felt incompetent (n =31) were in the "Incompetent" group. Table 2 shows the total mean scores of the groups.

Table 1. Mean scores for competent and incompetent groups

Group	N	Mean
Competent	26	12,81
Incompetent	31	12,55
Total	57	12,67

When the mean scores for the two groups were investigated, it was seen that there was no significant difference between the participants who felt competent enough and the ones who felt incompetent in terms of their overall scores from scenario-based assessment. Further analysis was conducted in terms of the mean scores of the groups for each scenario. Table 2 shows the mean scores for each scenario. When the findings were studied, it was seen that participants who felt more competent scored relatively higher than the ones in the incompetent group in scenarios 1, 2, and 3, whereas the incompetent group scored higher in scenario 4. Both groups had the same mean score for scenarios 5 and 6. However, apart from scenario 6, both groups scored lower than the average score (2.5) for each scenario, which was the cut-off point that was used to determine successful and unsuccessful participants. It was anticipated that the participants in the competent group would get more than 2.5 in each scenario, which was not the case. That is, all the participants scored lower than expected and eventually failed. With these findings, it would not be wrong to state that the data were inconclusive.

Table 3. Mean scores for each scenario

Group	N	Scenario 1 Mean	Scenario 2 Mean	Scenario 3 Mean	Scenario 4 Mean	Scenario 5 Mean	Scenario 6 Mean
Competent	26	2,04	1,31	2,54	2,22	1,65	3
Incompetent	31	1,94	1,29	2,19	2,48	1,65	3
Total	57	1,98	1,3	2,35	2,39	1,65	3

However, when the data were considered on an individual basis, an interesting finding was observed, which could be worth conducting further analyses. In this respect, it was found that not all participants in a group scored in the same way, which might have been the reason

for inconclusive findings. This finding provided a rationale to assume that more sophisticated grouping was needed. Because scores of individual participants indicated the existence of a more complicated pattern, it was seen that there were some mismatches within each group.

In the first group, the participants stated they felt competent enough to prepare a sound test. However, the problem was that although some participants felt competent, they scored lower than the average (15 points), which means they failed and did not know about the basics of testing. In more precise terms, those participants were unaware of their incompetency, which was interesting. On the other hand, in the same group, there were some other students whose perceptions matched their scores. That is, there were participants who felt competent enough to create a test and who were able to transfer it into practice since they scored higher than the average. That is, they were aware of their knowledge.

It was possible to see a similar pattern within the incompetent group. There were some students who felt incompetent in testing but scored higher than the average, which meant that although they had the basic knowledge regarding language testing, their self-perceptions were low in terms of confidence. This indicated the existence of a mismatch. Because those participants were supposed to be able to prepare a proper test. Moreover, it was also possible to find participants whose perceptions and scores matched in the incompetent group. That is, their competency and practice were consistent. Thus, based on this finding, participants were further regrouped into four categories in line with whether their perceptions and scores matched or not. Table 4 shows the scores for these four groups for each scenario.

Table 2. Individual-based scores for four groups

Group	Participants	Scenario 1 Score	Scenario 2 Score	Scenario 3 Score	Scenario 4 Score	Scenario 5 Score	Scenario 6 Score	Total Score
Competence Match	P34	4	4	3	2	3	4	20
	P14	3	1	3	5	3	4	19
	P7	3	2	3	2	2	5	17
	P13	2	2	4	3	1	5	17
	P28	2	0	4	3	4	4	17
	P32	2	2	3	3	2	4	16
	P53	3	2	2	3	3	3	16
Competence Mismatch	P17	1	1	2	2	0	3	9
	P26	1	1	3	2	0	2	9
	P27	0	0	3	2	3	1	9
	P56	1	1	2	2	1	1	8
	P49	2	2	3	0	0	0	7
	P51	1	2	2	1	0	0	6
	P14	0	0	3	2	0	0	5
Incompetence Mismatch	P5	3	2	3	4	0	4	16
	P35	4	2	1	3	2	4	16
	P45	3	3	2	3	2	3	16
	P54	3	2	2	3	3	3	16
	P12	2	1	3	3	1	5	15
	P47	3	2	4	2	1	3	15
	P11	1	2	2	3	3	3	14
Incompetent Match	P1	2	0	0	2	2	4	10
	P9	3	2	2	3	0	0	10
	P16	1	1	4	2	1	1	10
	P37	2	0	2	0	2	3	9
	P43	1	2	2	2	0	1	8
	P36	1	1	2	0	1	2	7
	P55	0	0	1	0	2	3	6

Having completed the analyses of the data obtained in the first two phases and grouped participants into four distinct categories, participants' sample tests were investigated. Only the tests by participants in these four categories were examined to see whether a pattern aligned with the groups. In the data collection phase, participants had been asked to create a test with specifications so their practical knowledge could be evaluated. First, a rubric was created, and the tests were graded. However, the data were not conclusive since participants were unable to provide original papers. They included either copy and paste style, superficial specifications, or unrelated ones that did not correspond to the questions prepared by the tester, that is, the participants. Some works were directly downloaded from the Internet and did not have relevant specifications. This made the interpretation of the papers impossible because they were not the real products of the participants. Thus, it would not be wrong to assume that participants performed poorly in this respect and provided unacceptable tests. This finding might indicate a lack of LAL since participants with higher levels of LAL are supposed to be skillful practitioners.

Discussion and Conclusion

This study explores teacher candidates' competency levels relating to their capabilities and self-perceptions of LAL. Additionally, it examines pre-service teachers' achievement levels in terms of realizing their LAL knowledge in various testing contexts. For these purposes, data are collected employing reflection papers in which diagnostic procedures to identify participants' LAL self-perceptions and knowledge are assessed, a test that encapsulates six assessment scenarios to evaluate how successfully pre-service teachers transfer their knowledge into different contexts, and finally, a task that incorporates test preparation with specifications to observe their abilities to apply their knowledge.

Responses to the open-ended questions are transcribed, and content analyses have been generated. This section aims to determine whether participants have an acceptable level of technical knowledge regarding test preparation and language assessment in general. Findings reveal that not all participants are able to perform as anticipated. Therefore, it could be mentioned that the participants have not fully attained the objectives of the testing course. In the literature, various researchers came up with similar findings. Graham (2005) and Jawhar and Subahi (2020) indicate that pre-service teachers are inefficient at improving their assessment knowledge during their university studies.

The investigation of open-ended questions also displays that many participants are aware of basic testing tools and can differentiate fundamental issues such as self-assessment or formative assessment tools. Accordingly, the level of success rate in terms of this question is not at the desired level. Less than half of the students can name the required skills and competencies for a successful test. This finding supports Sayyadi (2022), Gan and Lam (2020), and Vogt and Tsagari (2104), who reflect that at the tertiary level, teacher candidates fail to put their theoretical knowledge into practice due to various factors, one of which might be too much focus on theory and ignorance of practical aspects of assessment procedures. When the details are examined, it could be observed that only a small percentage of the participants are knowledgeable enough. These findings are partially supported by Quileste and Moreno (2020), who indicate that pre-service teachers have an acceptable level of knowledge in testing but require further improvement in certain areas, such as objective writing.

Findings demonstrate enough evidence to assume that teacher candidates are knowledgeable regarding fundamental aspects of language assessment procedures, such as

reliability and validity. Nevertheless, the finding fails to provide conclusive data to argue whether participants can identify the subdimensions of reliability and validity of a language test. Another issue that has been identified is that teacher candidates are unable to show acceptable levels of achievement in identifying the other aspects related to good language test preparation, such as practicality, ethical issues, content knowledge, and collaboration. In this respect, their highly low scores (the percentage drops as low as 7%) make their performances questionable. This finding is in line with what Kömür (2108), who examines the assessment awareness of pre-service teachers, states. Kömür (2108) argues that although pre-service teachers have a good amount of knowledge regarding skills and competencies in language assessment, but they need further training and practice opportunities. Therefore, based on this finding, it would not be wrong to state that the training given at the tertiary level on language testing is not sufficient to create such an effect on LAL at desired levels, and the testing curriculum needs enriching with the addition of innovative approaches and materials. This is in line with Triastuti (2020). He identifies that when it comes to practical applications of knowledge in teaching environments, pre-service teachers are unable to reflect what they know at acceptable levels. In the same vein, Sevimel-Şahin and Subaşı (2019), in their systematic review of the literature on LAL, state that in the pre-service teacher training area, the literature reveals the fact that at the tertiary level, too much emphasis on theory regarding LAL makes it difficult to bring out satisfactory outcomes in practice. So, as Berger (2023) suggests, there is a need to develop open and dynamic program designs. However, the findings of Giraldo and Quintero (2023) contradict these findings. Their exploratory action research reveals that it is possible to expand pre-service teachers' language assessment perceptions significantly within the tertiary level courses. In addition, El-Freihat (2021) asserts that years of experience play a great role in the practical application of knowledge. Yet, he also points out that the assessment courses in universities should be revised to meet the needs of pre-service teachers in this sense.

When the data related to participants' self-perception of their testing competencies have been inspected, it is seen that the findings provide enough evidence to support the previous findings of the current study. Accordingly, there are two main groups of participants in terms of their perceptions. On the one hand, there is a group of participants who think that they have obtained the necessary knowledge during their testing course and are competent enough to create suitable language tests. On the other hand, there is another group of participants who indicate that they have failed to improve their language testing competencies

throughout the testing course and, therefore, perceive themselves as incompetent in test preparation and administration. Similarly, Yan and Fan (2020) say that there are discrepancies at individual and group levels in terms of LAL profiles. In addition to these findings, the investigation of the data set also revealed that more than half of the participants feel the need for further training to improve their testing competencies after university education when they start teaching, which parallels the other studies in the literature. Fitriyah, Masitoh, and Widiati (2022) state that although teachers, both experienced and novice ones, indicate they have a good amount of knowledge about language assessment and testing, they still demand further in-service training. With reference to these findings, it could be argued that it is normal to see people with varying degrees of competencies, and individuals may or may not feel competent enough under normal conditions.

In the next stage, the results of the scenario-based test are evaluated to examine the consistency between what participants know and how they apply that knowledge. The assumption is that participants who feel competent enough would outscore in the test the ones who feel incompetent and must be able to get a score higher than the average. The highest score that one participant could get on the test is 5, which means that participants in the competent group should get at least 2.5. However, when the overall scores are examined between groups, it is seen that no significant difference exists. That is, all the participants are at the same level no matter what level of competence they have in terms of their LAL, and regardless of their perceptions, they all fail. This finding brings out the question of how participants can perceive their LAL levels as high but unable to transfer their knowledge into practice, which is supported by Gurmesa, Birbirs, Hussein, and Tsegaye (2022), who affirm that EFL teachers' assessment knowledge is not reflected in practical applications in classroom environments, which means the efficiency of testing courses offered at the tertiary level is questionable and may be unable to ensure the attainment of all its objectives by the pre-service teachers. It could be argued that from an orthodox point of view, pre-service teachers take certain courses throughout their education at university, and then they are expected to reflect their knowledge in practice when they start teaching at schools. It is inevitable to observe varying degrees of achievement levels in practice, but if the majority of pre-service teachers fail to prepare and administer language tests at acceptable levels, it means there is a paralogism.

If an individual cannot apply the knowledge he has acquired after successfully completing a course, then it might be possible to argue that the problem may stem from

practices at the tertiary level. In this respect, the success criteria and assessment and evaluation processes of the courses might become questionable. This finding is supported by Morrison and Sepulveda-Escobar (2022). They argue that to obtain reliable, valid, and authentic assessment products in foreign language classrooms, the efficiency of teacher education programs should be improved. Similarly, Muianga (2023) says that there is a need to redesign professional programs associating with assessment literacy to become more effective based on his study on teachers' conceptions of assessment. Beyond that, the testing course curriculum and syllabuses might need to be updated in a way that meets the changing conditions and requirements to ensure that pre-service teachers develop their LALs at acceptable levels by bringing the practical hands-on applications forward. Coombe, Vafadar, and Mohebbi (2020) contribute to this by expressing that training on assessment should evolve into long-term and sustainable applications in a way to include the whole professional life of the teacher. Moreover, these kinds of programs should be built upon the findings of recent research and consider the requirements of context time and contemporary demands by stakeholders.

The above findings reveal that there are inconsistencies between groups in terms of participants' competency levels and self-perceptions, which is supported by Elhussien and Khalil (2023), who argue that pre-service teachers are unable to make accurate decisions in terms of applying their knowledge and also they are unaware of their assessment abilities. In addition, it is seen that groups do not perform as anticipated. In this respect, findings from individual scrutiny reveal a distinct pattern in both groups. Reflection of the big picture there are inconsistencies within each group. In both competent and incompetent groups, there are participants whose perceptions match their performances. That is, participants in each group can be allocated into two sub-categories: match and mismatch, which divides all participants into quartiles. Participants in matching groups perform as expected. However, there is a significant problem in mismatch groups. The ones in the competent mismatch tend to underestimate their performances. Similarly, participants in the incompetent mismatch group overestimated their performances.

These findings provide a rationale for assuming there is a cognitive error. This finding is in line with Freund and Kasten (2012) and Hofer, Mraulak, Grinschgl, and Neubauer (2022), who states that people's performances and their self-estimation of abilities and intelligence have a mediocre correlation with each other. Indeed, this is what Kruger and Dunning (1999) propose in their study that people who get low marks on a given test tend to

overestimate their performances, whereas people with higher marks underestimate their performances. As Sanchez and Dunning (2018) state, a consistent pattern is observed between the actual and estimated performances, which is the case in this study. This connection is called the Dunning-Kruger Effect (DFE). Basically, people are not aware of to what extent they do not know a specific subject area, and this, in turn, results in the misapplication of appropriate strategies and negatively affects learning and performance. Coutinho, Thomas, Alsuwaidi, and Couchman (2021) argue that due to DKE, people cannot overcome their shortfalls in knowledge or skills, which could be interpreted as the source of greater problems that might stem from these deficits at later stages. Therefore, it could be quoted that at the individual level, the problems may result from individuals' cognitive biases regarding their performance perceptions. That is, the ineffectiveness of course design or the practices carried out by the lecturers might not be the only factors instigating participants' inadequacies. So, any future modification of the testing courses at the tertiary level should incorporate practices which aim at reducing cognitive error and DKE so that pre-service teachers can be equipped with better testing and assessment knowledge, which is a view also supported by Lee and Butler (2020), and Taylor (2013) who argue that among many stakeholders regarding LAL, learners (pre-service teachers) have a crucial role. Therefore, their perspectives need to be taken into consideration for better conceptualization of LAL.

In conclusion, based on the above-mentioned findings, it is concluded that pre-service language teachers fail to reflect what they have learned in assessment courses into practice, which makes the efficiency of assessment courses at the tertiary level questionable, which means assessment courses fail to achieve their objectives from practical considerations. Pre-service teachers know the basics of assessment procedures superficially but are unable to perform at an acceptable level regarding the other aspects of test preparation, such as creating collaborative tests, abiding by ethical codes, and assessing the content knowledge appropriately. These issues could be seen as the manifestation of low levels of LAL from both practical and theoretical perspectives. This study provides evidence regarding the problems mentioned above, too. Accordingly, in some measure, the source is the cognitive errors that pre-service teachers bear while self-evaluating their individual language assessment competencies. The intricated inconsistencies between pre-service teachers' self-perceptions of their LAL and their actual performances reveal the existence of the Dunnig-Kruger Effect, which may result in the selection of inaccurate strategies and/or paths in pre-service teachers' learning journey, and without overcoming this issue, it might not be

possible to create efficient programs. Considering these results, it would not be wrong to conclude that the assessment courses at tertiary levels need revising in a way that focuses on practical applications of theoretical knowledge and that is supported by interventions aiming to decrease the levels of cognitive errors of teacher candidates.

The current study is a case study, and the results are data-driven. It will evidently contribute to the field but is not free from some limitations. First, data are collected from a limited group of participants, and it is difficult to generalize the findings to other contexts. In addition, there is a risk of bias. Although participants have been informed about the ethical aspects and asked for their sincere responses, their personal opinions and perceptions may influence the results. Moreover, it is not possible to generate the same study with the same population, either. Even if it is possible, it is open to the recall bias.

For further studies, materializing an empirical study, which will be conducted with the participation of more participants from various contexts, could provide quantitative data that would promote more sound results. Additionally, researchers may focus on the role of cognitive errors in university-level courses on student achievement and motivation. It is also recommended that design studies focus on specific solutions for the entailing problems.

Declaration of Conflicting Interest
The Authors declare that there is no conflict of interest.

References

Al-Bahlani, S. M. (2019). Assessment literacy: A study of EFL teachers' assessment knowledge, perspectives, and classroom behaviors [Published doctoral dissertation, The University of Arizona]. UA Campus Repository, University Libraries

Alderson, J. C. (2005). Editorial. *Language Testing*, *22*(3), 257–260. https://doi.org/10.1191/0265532205lt315ed

Ashraf, H., & Zolfaghari, S. (2018). EFL Teachers' Assessment Literacy and Their Reflective Teaching. *International Journal of Instruction*. 11(1), 425-436. 10.12973/iji.2018.11129a.

Berger, A. (2023). A difficulty-informed approach to developing language assessment literacy for classroom purposes. *Chinese Journal of Applied Linguistics, 46*, 289 - 307.

Beziat, T. L. R. & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25-30.

Boylu, E. (2021). Turkish as a foreign language teachers' attitudes towards assessment and evaluation. *International Journal of Education and Literacy Studies*. 9, 72-85. 10.7575/aiac.ijels.v.9n.1p.72.

Brown, H. D., & Abeywickrama, P. (2010). Language Assessment: Principles and Classroom Practices. Pearson Education

Coombe, C. A., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia, 10*.

Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, *71*, 134–144.

Coutinho M. V. C., Thomas J, Alsuwaidi A. S. M., & Couchman J. J. (2021). Dunning-Kruger effect: intuitive errors predict overconfidence on the cognitive reflection test. *Front. Psychol*, 12,. https://doi.org/10.3389/fpsyg.2021.603225

Creswell, J. W., & Plano Clark, V. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications, Inc.

Davies, A. (2008). Textbook trends in teaching language testing. *Lang. Test.* 25, 327–347. https://doi.org/10.1177/0265532208090156

Davison, C. (2019). "Using assessment to enhance learning in English language education," in *Second Handbook of English Language Teaching*. ed. X. Gao (Cem, Switzerland: Springer International), 433–454.

Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds.), *Revisiting EFL assessment* (pp. 9-27). Switzerland: Springer International Publishing.

Fitriyah, I., Masitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*.

Freund, P. A., and Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin,* 138, 296–321.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132, https://doi.org10.1080/15434303.2011.642041

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia, 10*, 1-18.

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179-195. https://doi.org/10.15446/profile.v20n1.62089.

Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), https://doi.org/ 10.14483/22487085.14514

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education* 21(6),607-621. https://doi.org/10.1016/j.tate.2005.05.001

Gurmesa, D. B., Birbirs, D. T., Hussein, J. W., & Tsegaye, A. G. (2022). Ethiopian secondary school EFL teachers' classroom assessment conceptions and practices from an activity theory perspective. *East African Journal of Education Studies*.

Gürsoy, G. (2017). Ölçme ve Değerlendirme Okuryazarlığı: Kavramsal Bir Analiz . *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 6 (1), 281-316.

Hatipoğlu, Ç. (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

Hatipoğlu, Ç., & Erçetin, G. (2016). Türkiye'de yabancı dilde ölçme ve değerlendirme eğitiminin geçmişi ve bugünü [Past and present of education in foreign language testing and assessment in Turkey]. In S. Akcan, & Y. Bayyurt (Eds.) 3. Ulusal Yabancı Dil Eğitimi Kurultayı: Türkiye'de Yabancı Dil Eğitimi Üzerine Görüş ve Düşünceler [The 3rd National Foreign Language Education Conference: Views and

Opinions on Foreign Language Education in Turkey] (pp. 72-89). Boğaziçi University.

Heitink, M.C., van der Kleij, F., Veldkamp, B. P., Schildkamp, K. & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50-62. https://doi.org/10.1016/j.edurev.2015.12.002.

Hofer, G., Mraulak, V., Grinschgl, S., & Neubauer, A. (2022). Less-intelligent and unaware? Accuracy and Dunning–Kruger effects for self-estimates of different aspects of intelligence. *Journal of Intelligence*, 10 https://doi.org/10.3390/jintelligence10010010.

Jawhar, S. S., Subahi, A. M. (2020). The impact of specialty, sex, qualification, and experience on teachers' assessment literacy at Saudi higher education. *International Journal of Learning, Teaching and Educational Research*, 19(5), 200-216, https://doi.org/10.26803/ijlter.19.5.12

Koh, K.H. (2011). Improving teachers' assessment literacy through professional development, *Teaching Education*, 22(3), 255-276, https://doi.org/10.1080/10476210.2011.593164

Kömür, Ş. (2018). Pre-service English teachers' assessment awareness: level of readiness or classroom readiness. *Journal of Language Teaching and Learning.*

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Latif, M. W. (2017). Tertiary level EFL practitioners' understanding of the dynamics of assessment literacy: An exploratory study. In C. Coombe, P. Davidson, A. Gebril, D. Boraie, & S. Hidri (Eds.), *Language assessment in the Middle East and North Africa: Theory, practice and future trends*, (pp. 71–87) TESOL Arabia.

Latif, M. W. (2021). Exploring tertiary the EFL practitioners' knowledge base component of assessment literacy: implications for teacher professional development. *Language Testing in Asia,* 11(1), 1-22. https://doi.org/10.1186/s40468-021-00130-9

Lee J., & Butler Y. G. (2020). Reconceptualizing language assessment literacy: Where are language learners? *TESOL Quarterly*, 54(4), 1098–1111. https://doi.org/10.1002/tesq.576

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

McCrudden, M. T., & Rapp, D. N. (2023). Inquiry worldviews, approaches to research, and mixed methods in educational psychology. In *Handbook of Educational Psychology* (pp. 48-70). Taylor and Francis. https://doi.org/10.4324/9780429433726-5

McMillan, J. H. (2003). *Classroom Assessment - Principles & Practice for Effective Instruction*. Alyn & Bacon.

El-Freihat, S. M. (2021). EFL language assessment literacy among Jordanian EFL secondary school teachers. *Cypriot Journal of Educational Sciences*. 16(6), 3266-3277. https://doi.org/10.18844/cjes.v16i6.6321

Mohebbi, H., & Coombe, C. (2020). *Research Questions in Language Education and Applied Linguistics: A Reference Guide.* Springer.

Morrison, A., & Sepulveda-Escobar, P. (2022). Chilean English as a foreign language teacher educators' conceptions and practices of online assessment. *RELC Journal*.

Ölmezer-Öztürk, E., & Aydın, B. (2019). Voices of EFL teachers as assessors: Their opinions and needs regarding language assessment . *Eğitimde Nitel Araştırmalar Dergisi*, 7(1), 373-390 .

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

Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63, 84-85.

Quilter, S., & Gallini, J. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36, 115-131, https://doi.org/10.1080/08878730009555257.

Quileste, R., & Moreno, L. A. (2020). The pedagogical content knowledge of the school of education pre-service teachers in test construction. *JPAIR Institutional Research*. DOI: https://doi.org/10.7719/irj.v14i1.802

Rahman, M. M. (2018). Teachers' perceptions and practices of classroom assessment in secondary school science classes in Bangladesh. *International Journal of Science and Research (IJSR)*. https://doi.org/10.21275/art20183034

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, 114(1), 10–28. https://doi.org/10.1037/pspa0000102

Sayyadi, A. (2022). In-Service university-level EFL instructors' language assessment literacy and training needs. *Profile: Issues in Teachers' Professional Development*.

Sevimel-Şahin, A., & Subaşı, G. (2019). An overview of language assessment literacy research within English Language Education Context. *Journal of Theoretical Educational Science,* 12(4), *1340-1364,* https://doi.org/10.30831/akukeg.501817

Shim, K. N. (2009). An Investigation into Teachers' Perceptions of Classroom-Based Assessment of English as a Foreign Language in Korean Primary Education. Doctoral Thesis. University of Exeter.

Stiggins, R. J. (1995). Professional Development: The Key to a Total Quality Assessment Environment. *NASSP Bulletin*, *79*(573), 11-19. https://doi.org/10.1177/019263659507957303

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. https://doi.org/10.1177/0265532213480338.

Triastuti, A. (2020). assessing English pre-service teachers' knowledge base of teaching: linking knowledge and self-portrayal. *TEFLIN Journal - A publication on the teaching and learning of English*.

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58,149-162, https://doi.org/10.1016/j.tate.2016.05.010.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*, 374 - 402.

Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking Learners' Perspectives on Language Assessment Practices to Teachers' Assessment Literacy Enhancement (TALE): Insights from Four European Countries. Language Assessment Quarterly, 17(4), 410–433. https://doi.org/10.1080/15434303.2020.1776714

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, 1-24,https://doi.org/10.1080/15434303.2020.1781128.

Watmani, R., Asadollahfam, H., & Behin, B. (2020). Demystifying language assessment literacy among high school teachers of English as a foreign language in Iran: Implications for teacher education reforms. Tabaran Institute of Higher Education. *International Journal of Language Testing,* 10(2), 129-144.

Wu, X., Zhang, J.L., &Liu, Q. (2021). Using assessment for learning: Multi-case studies of three Chinese university English as a foreign language (EFL) teachers engaging students in learning and assessment. *Front. Psychol*, 12, https://doi.org/10.3389/fpsyg.2021.725132

Yan, X., Zhang, C., & Fan, J. J. (2018). Assessment knowledge is important, but…: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System, 74*, 158–168. https://doi.org/10.1016/j.system.2018.03.003

Yan, X., & Fan, J. (2020). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing, 38*, 219 - 246.

Yeşilçınar, S., & Kartal, G. (2020). EFL teachers' assessment literacy of young learners: Findings from a small-scale study . *Journal of Theoretical Educational Science* , 13 (3) , 548-563.

Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An Investigation into EFL Teachers' Assessment Literacy: Indonesian Teachers' Perceptions and Classroom Practice. *European Journal of Contemporary Education, 9*, 189-201.