

## Investigating Language Assessment Knowledge of EFL Teachers\*

### İngilizceyi Yabancı Dil Olarak Öğreten Öğretmenlerin Dilde Ölçme Değerlendirme Bilgilerinin Araştırılması

Elçin ÖLMEZER-ÖZTÜRK\*\*, Belgin AYDIN\*\*\*

• Received: 13.07.2018 • Accepted: 07.09.2018 • Published: 31.07.2019

**Kaynakça Bilgisi:** Ölmezer-Öztürk, E., & Aydın, B. (2019). İngilizceyi yabancı dil olarak öğreten öğretmenlerin dilde ölçme değerlendirme bilgilerinin araştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(3), 602-620. doi: 10.16986/HUJE.2018043465

**Citation Information:** Ölmezer-Öztürk, E., & Aydın, B. (2019). Investigating language assessment knowledge of EFL teachers. *Hacettepe University Journal of Education*, 34(3), 602-620. doi: 10.16986/HUJE.2018043465

**ABSTRACT:** The purpose of this study is to provide an overall picture regarding general and skill-based Language Assessment Knowledge (LAK) level of English teachers working in Turkish higher education context by using Language Assessment Knowledge Scale (LAKS) developed by Ölmezer-Öztürk and Aydın (2018). The scale including 60 items with four constructs was completed by 542 EFL teachers working at schools of foreign languages. The statistical findings revealed that in terms of LAK level of the teachers, the participants got, on average, 25 out of 60. The results of one sample t-test revealed that the score the teachers got from the scale was significantly lower than half of the total score. It was also found that the teachers were the most knowledgeable in assessing reading whereas they had the lowest score in assessing listening. Besides, except for being a testing office member, no significant impact of demographic features was found on LAK level of the participants. Finally, the present study offers several suggestions both for future studies and for policy makers to improve EFL teachers' language assessment literacy on assessing each skill.

**Keywords:** Language assessment knowledge, EFL teachers, language assessment literacy

**ÖZET:** Bu çalışma İngilizce öğretmenlerinin dilde ölçme değerlendirme bilgisi seviyelerini ortaya koymayı amaçlamaktadır. Veriler 60 maddeli ve 4 boyutlu dilde ölçme değerlendirme bilgisi ölçeği ile toplanmış ve bu ölçek yükseköğretimde çalışan 542 öğretmen tarafından cevaplanmıştır. Öğretmenlerin dilde ölçme değerlendirme bilgi seviyeleri ile alakalı bulgular, katılımcıların ölçeğin genelinden 60 üzerinden 25 aldığını ve bu ortalamanın toplam puanın yarısından anlamlı derecede düşük olduğunu ortaya koymuştur. Bunun yanı sıra, öğretmenlerin dilde ölçme değerlendirme bilgi düzeylerinin okuma becerisini ölçmede en yüksek, dinleme becerisini ölçmede ise en düşük olduğu belirlenmiştir. Ölçme değerlendirme ofisi çalışanı olup olmama haricinde hiçbir demografik değişkenin öğretmenlerin dilde ölçme değerlendirme bilgisi üzerinde etkisinin olmadığı da çalışmanın istatistikî bulguları arasındadır. Çalışmanın sonuç bölümünde ise öğretmenlerin dilde ölçme değerlendirme okuryazarlıklarının geliştirilmesi için hem araştırma hem planlama açısından bazı önerilerde bulunulmuştur.

**Anahtar Sözcükler:** Dilde ölçme değerlendirme bilgisi, İngilizce öğretmenleri, dilde ölçme değerlendirme okuryazarlığı

## 1. INTRODUCTION

The role of assessment in teaching and learning process is undeniable. Assessment is regarded like an engine which is responsible for initiating learning (White, 2009). Teaching and assessment cannot be considered as separate constructs, because assessment is a component of

\* This study is a condensed summary of the doctoral dissertation "Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers" written by Elçin ÖLMEZER-ÖZTÜRK.

\*\* Öğr. Gör. Dr., Anadolu Üniversitesi, Eğitim Fakültesi, Yabancı Diller Eğitimi Bölümü, İngiliz Dili Eğitimi Ana Bilim Dalı, Eskişehir - TÜRKİYE. e-posta: [elcinolmezerozturk@anadolu.edu.tr](mailto:elcinolmezerozturk@anadolu.edu.tr) (ORCID: 0000-0001-7743-6361)

\*\*\* Prof. Dr., Anadolu Üniversitesi, Eğitim Fakültesi, Yabancı Diller Eğitimi Bölümü, İngiliz Dili Eğitimi Ana Bilim Dalı, Eskişehir - TÜRKİYE. e-posta: [baydin@anadolu.edu.tr](mailto:baydin@anadolu.edu.tr) (ORCID: 0000-0002-4719-7440)

learning and teaching process, and teachers are engaged in assessment and assessment-related activities in most of their professional time. Good assessment practices are crucial because the quality of the assessments that are utilized is a prerequisite for the quality of the instruction and learning (Stiggins, 1999). These assessment practices are good informants; so, with the help of these, teachers can adapt the pace of the lesson, make a decision about whether the course content is relevant or not, shape student learning during teaching process, get an idea about whether the teaching is effective or not, and help create confidence in students for the national standardized tests. A language teacher has this assessment responsibility as a part of her/his profession (Mertler, 2003). As teaching and assessment are the concepts affecting each other, they inform and improve each other (Malone, 2013); thus, teachers have great roles in bridging between these two concepts. The role of teachers is made salient in assessment process, and many scholars in the literature (Stiggins, 1999; Popham, 2009) voiced that language teachers can become more informed decision makers once they are equipped with language assessment knowledge. With this great role in language assessment, teachers' knowledge of assessment has a big impact on the quality of education (Malone, 2013). As a result, it is necessary for teachers to utilize assessment strategies to make decisions, to decide on the most suitable instruction for learners, and to get an idea about teaching and learning progress.

Stiggins (1995, p. 240) emphasized that assessment literate teachers know “what they are assessing, why they are doing it, how best to assess the skill, knowledge of interest, how to generate good examples of student performance, what can potentially go wrong with the assessment, and how to prevent that from happening”. Thus, assessment literacy covers the knowledge related to assessment and also application of this knowledge during assessment practices. Very recently, a new term, language assessment literacy, has flourished, and it is rooted in the term assessment literacy, but it has appeared as a distinct area from assessment literacy. The definitions of language assessment literacy vary in the literature. Malone (2013, p. 329) defined language assessment literacy as “language teachers' familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language”. For Inbar-Lourie (2008, pp. 389-390), “language assessment knowledge base comprises layers of assessment literacy skills combined with language specific competencies, forming a distinct entity that can be referred to as language assessment literacy”. In another work, Inbar-Lourie (2017) stated that the term LAL stems from AL, but LAL is different from AL in the sense that LAL “attempts to set itself apart as a knowledge base that incorporates unique aspects inherent in theorizing and assessing language-related performance” (p. 259). Lam (2015) also defined it as “teachers' understandings and mastery of assessment concepts, measurement knowledge, test construction skills, principles about test impact, and assessment procedures which can influence significant educational decisions within a wider social context” (p. 172). Finally, language assessment literacy was also defined as “the level of knowledge, skills, and understanding of assessment principles and practice that is increasingly required by other test stakeholder groups, depending on their needs and context” (Taylor, 2009, p. 24). Thus, as is stated in the definitions above, language assessment literacy requires additional competencies related to language when compared to assessment literacy. However, research into language assessment literacy “is still in its infancy” (Fulcher, 2012, p. 117).

It is crucial that language teachers need to have adequate knowledge in assessment-related process (Price, Rust, O'Donovan, Handley, & Bryant, 2012). However, many in-service teachers stressed that they are not adequately equipped with assessment knowledge (Plake, 1993). Stiggins (2010, p. 233) pointed out this problem with a very assertive utterance by stating that “assessment illiteracy abounds”. This indicates that teachers are responsible for assessing learners, but whether they have the necessary knowledge to assess learners is open to discussion.

Xu and Brown (2017, p. 134) stated that assessment literacy should start with the investigation of its knowledge base; thus, the assessment knowledge lies at the heart of assessment literacy. Some of the studies focusing LAL are as follows: Lam (2015) carried out a study to find out how two language assessment courses facilitated or inhibited the language assessment literacy of pre-service teachers in five Hong Kong institutions. The analysis of the programmes showed that there was insufficient support to foster LAL, and the training for LAL was inadequate. Next, Tsagari and Vogt (2017) carried out a mixed-design study to find out the teachers' perceptions of LAL and their individual needs related to language testing and assessment. The participant teachers were from Cyprus (n=16), Greece (n=22) and Germany (n=25) with no training related to assessment. The results demonstrated that the participants' perceived LAL was not sufficient, and they did not feel themselves prepared effectively for assessment-related practices. Additionally, Volante and Fazio (2007) carried out a study with 69 pre-service teachers from each of the four years in ELT programme. The findings indicated that self-efficacy ratings of the participants were very low across each of the four years of the programme. The majority of the respondents made use of assessment for mainly traditional summative purposes. Furthermore, the pre-service teachers stated for an urgent need for a specific course based on classroom assessment, and this need was verbalized by all the participants across four years.

In Turkish context, Hatipoğlu (2015) studied with 124 pre-service teachers to investigate what pre-service teachers knew about assessment and what their expectations were from testing course. The findings demonstrated that the participants expected to evaluate, select and write exams and prepare their learners for all types of exams. It was also revealed that the pre-service teachers had limited assessment knowledge after four years in ELT department. Another study belongs to Öz and Atay (2017) who investigated the Turkish EFL teachers' perceptions towards in-class language assessment and its link with their classroom practices. The findings revealed that the teachers were familiar with the basic terms related to classroom assessment; however, when it comes to the practice, they had difficulty in reflecting their assessment knowledge into their classroom practice. Finally, Mede and Atay (2017) made use of the online language testing and assessment questionnaire adapted from Vogt and Tsagari (2014) to find out the training needs and practices of 350 Turkish EFL teachers. The findings showed that the teachers had limited assessment literacy, and they needed training in many areas of testing and assessment.

The studies related to language assessment literacy of EFL teachers are mostly concerned with the needs of English language teachers with regard to language assessment, proving the inefficiency of pre-service education and lack of professional development, and revealing the self-reports of the participants related to their assessment knowledge or practices.

Thus, there is an urgent need for exploring language assessment knowledge of language teachers as the first step towards language assessment literacy in order to detect their strengths and weaknesses in language assessment. Preparatory programmes of the universities are one of the workplaces where teachers are expected to teach English, and assess their learners in each skill. The assessors in these preparatory programmes are language teachers. The problem is that language teachers are responsible for all the assessment-related activities in most of the preparatory programmes, but how knowledgeable or competent they are in assessing their learners is the question. As a starting point, language assessment knowledge of language teachers should be determined. However, in Turkey, there is not a study on measuring the language assessment knowledge of language teachers; thus, there is paucity of research in language assessment literacy to shed light on this issue. This identification is vital because by detecting the strengths and weaknesses of language teachers, the needs of language teachers could be specified. Based on these needs, testing and assessment course in pre-service education and teacher professional development programmes related to language assessment can be designed and developed.

As the present study investigates language assessment knowledge of EFL teachers in higher education in Turkey with the help of a scale that was developed and validated by the researchers, it aimed to find out answers to the following research questions throughout the study:

1. What are the general and skill-based Language Assessment Knowledge (LAK) levels of EFL teachers in Turkish higher education setting?
2. Is there a relationship among their levels of skill-based LAK?
3. Does LAK level change according to following demographic features which are years of experience, educational background, the BA programme being graduated, workplace, having a testing course in BA, attending trainings on testing and assessment and being a testing office member?
4. Does their LAK level change according to their perceived self-competency in assessing each language skill?

## 2. METHODOLOGY

### 2.1. Research Context

Turkey is an EFL context in which English does not have an official status. It is taught as a foreign language at primary, secondary and university levels. The preparatory programmes of the universities are the research context in this study. With the English preparatory programmes they have, schools of foreign languages at universities are the institutions in which English is taught in a systematic and intensive way in Turkey (Aydın, et. al., 2017). These programmes were deliberately chosen for this study, because the language teachers in these programmes are responsible for assessment-related activities and all four skills are assessed.

### 2.2. Participants

The population of this study included Turkish EFL teachers working at preparatory programmes of the universities in Turkey. The online version of the scale was sent to all the language teachers of the universities which have English preparatory programmes. Among 122 universities (85 state and 37 private universities) in Turkey, the scale was sent to the ones with English preparatory programmes. Throughout the data collection process, the language teachers and the heads of the preparatory programmes were sent reminder e-mails every week in order to reach the maximum number of participants. The ones who responded to the scale were included as the participants in the study. 542 language teachers from 53 universities (37 state and 16 private universities) contributed to the data collection process of the study. There were 115 language teachers answering the scale from one of the state universities, which was the highest number of the participants responding to the scale, and the rest of the participants contributing to the data collection ranged from two to 30 language teachers. The demographic features and the number of the teachers are shown in the following table.

**Table 1. Demographic features and the number of the participants**

Demographic feature	Number of the Participants	Percentage
<b>Gender</b>	Male – 174	32
	Female - 368	68
<b>Years of experience</b>	1-5 years – 86	16
	6-10 years – 173	32
	11-15 years – 114	21
	16-20 years – 100	18
	More than 21 – 69	13
<b>Educational background</b>	BA – 238	44
	MA – 255	47
	PhD – 49	9
<b>The BA programme graduated</b>	ELT – 347	64

	Non-ELT - 195	36
<b>The current workplace</b>	State University – 372	68
	Private University – 170	32
<b>Being a testing office member</b>	Yes – 260	48
	No – 282	52
<b>Had a separate testing/assessment course in pre-service</b>	Yes – 260	48
	No – 282	52
<b>Attended any trainings on language testing/assessment</b>	Yes – 282	52
	No – 260	48

### 2.3. Data Collection Tool

At the initial stage of this study, a scale to measure language assessment knowledge of EFL teachers was developed by the researchers at the end of a tough validation process which included opinions from teachers, opinions from academicians as the expert opinion and a meeting with testing practitioners. In this process, the items that were perceived as too difficult, too easy, irrelevant or tricky were removed from the scale based on the comments. Besides, the items which received the same answer in the piloting process were also removed. These stages, as shown below in Table 2, considerably contributed to the content validity of the scale and the development of each item.

**Table 2. Revision process of the scale**

	<b>reading</b>	<b>listening</b>	<b>writing</b>	<b>speaking</b>	<b>in total</b>
<b>1<sup>st</sup> stage</b>	49	61	74	53	237
(Three experts with PhD in ELT checking for comprehensibility)	-3	-2	-6	-6	-17
<b>2<sup>nd</sup> stage</b>	46	59	68	47	220
(Checking with 10 teachers)	-	-	-	-	-
<b>3<sup>rd</sup> stage</b>	46	59	68	47	220
(Expert opinion)	-9	-26	-20	-12	-67
<b>4<sup>th</sup> stage</b>	37	33	48	35	153
(Training with the testing office members)	-9	-7	-14	-11	-41
<b>5<sup>th</sup> stage</b>	28	26	34	24	112
(Piloting with 50 teachers and expert opinion)	-13	-11	-19	-9	-52
<b>Final Version</b>	15	15	15	15	60

Language Assessment Knowledge Scale (LAKS), which was developed by the researchers at the end of the stages shown above, was sent in an online format to all the teachers working at preparatory programmes at universities in Turkey. It has a high reliability ( $R=.91$ ), and the findings derived from the Second Order Confirmatory Factor Analysis indicated that the model presents a good fit (Ölmezer-Öztürk & Aydın, 2018). It included 60 items with four constructs that are assessing reading, assessing listening, assessing writing and assessing speaking.

During the data collection process, there were four different formats of the scale, and in each format, the places of the skills were different. For instance, in the first format, assessing reading was the first skill in the scale whereas in the second format, it was placed in the second rank after assessing listening. What is more, only one format was sent to a university, and it was noted down by the researchers not to cause any confusion in later stages. The reason for this variability was to avoid the possible effects of any fatigue, boredom or careless responses of the participants toward the end of the scale.

## 2.4. Data Analysis Process

The quantitative data were analyzed through descriptive and inferential statistics. The following table presents the statistical methods that were used to analyze the quantitative data of this study in line with the research questions.

**Table 3. Statistical methods used in analysis**

The focus of the research question	The statistical method
R. Q. 1. The level of general and skill-based language assessment knowledge	Descriptive statistics (mean, percentage, standard deviation, etc.) One sample t-test
R. Q. 2. The relationship of their skill-based knowledge	Pearson Correlation
R. Q. 3. The impact of demographic features on the knowledge level of participants	Inferential statistics (Independent samples T-test, one-way ANOVA)
R. Q. 4. Perceived self-competency and LAK level	One-way ANOVA

## 3. FINDINGS

### 3.1. General and Skill-based Language Assessment Knowledge Level of EFL Teachers

The first research question of the study aimed to investigate general and skill-based LAK level of EFL teachers working at Turkish higher education context. The findings derived from their responses are presented in Table 4. The results showed that the participants' mean score in LAKS was 25 over 60. In other words, the number of the items answered correctly by the teachers were 25 on average. Besides, it was found that the participants had the highest mean score in assessing reading (7,055 over 15) which means that they know more about assessing reading compared to assessing other skills. According to the results in Table 4, the participants got a mean score of 4,752 over 15 in assessing listening and it was found to be the skill in which the participant teachers were the least knowledgeable.

**Table 4. General and skill-based LAK level of EFL teachers in Turkish higher education context**

ITEMS	N	True	False	Don't Know	Mean	SD
<b>ASSESSING READING (Bold ones refer to the participants with correct answers)</b>						
1. Asking learners to summarize the reading text is a way of assessing their reading skills.	542	<b>269</b>	257	16	,496	,500
2. When asking several questions about a reading text, all the questions are independent of each other.	542	<b>153</b>	343	46	,282	,450
3. Cloze test is used for assessing the main idea of the text.	542	230	<b>250</b>	62	,461	,498
4. In a reading exam, using a text learners have encountered before is not a problem.	542	278	<b>190</b>	74	,350	,477
5. One reading text is enough to be included in a reading exam.	542	108	<b>400</b>	34	,738	,440
6. The language of the questions is simpler than the text itself.	542	<b>264</b>	220	58	,487	,500
7. Errors of spelling are penalized while scoring.	542	256	<b>237</b>	49	,437	,496
8. Taking vocabulary difficulty into consideration is necessary in assessing reading skills.	542	<b>288</b>	224	30	,531	,499
9. Including not stated/doesn't say along with true/false items has advantages over true/false items.	542	<b>236</b>	221	85	,435	,496
10. The more items a reading text is followed, the more reliable it becomes.	542	<b>198</b>	200	144	,365	,481
11. Using the same words in the correct option as in the text is not a problem.	542	241	<b>243</b>	58	,448	,497
12. Simplification of reading texts is avoided.	542	243	<b>205</b>	94	,378	,485
13. Reading texts in a reading exam include various genres (essay, article, etc.).	542	<b>328</b>	188	26	,605	,489
14. In top-down approach, assessment is on overall comprehension of the reading text.	542	<b>267</b>	110	165	,492	,500

15. Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.	542	<b>296</b>	199	47	,546	,498
<b>READING-TOTAL</b>	<b>542</b>				<b>7,055</b>	<b>4,470</b>
<b>ASSESSING LISTENING</b>						
16. Using reading texts for listening purposes poses a problem.	542	<b>160</b>	292	90	,295	,456
17. Including redundancy (e.g. what I mean to say is that ...) in a listening text poses a problem.	542	243	<b>228</b>	71	,420	,494
18. Any type of listening text is used for note-taking.	542	267	<b>223</b>	52	,411	,492
19. Spelling errors are ignored in scoring the dictation.	542	<b>92</b>	400	50	,169	,375
20. Errors of grammar or spelling are penalized while scoring.	542	319	<b>169</b>	54	,311	,463
21. A listening cloze test is a way of selective listening.	542	<b>286</b>	139	117	,527	,499
22. Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative testing.	542	209	<b>63</b>	270	,116	,320
23. Scoring in note-taking is straightforward.	542	253	<b>132</b>	157	,243	,429
24. In discrete-point testing, comprehension is at the literal/local level.	542	<b>199</b>	45	298	,367	,482
25. Using dictation diagnostically in assessing listening skills does not pose a problem.	542	<b>172</b>	171	199	,317	,465
26. Giving learners a transcript of the listening text is a valid way of assessing listening skills.	542	224	<b>259</b>	59	,477	,499
27. Dictation is a kind of discrete-point testing.	542	253	<b>52</b>	237	,095	,294
28. Inference questions based on intelligence are avoided in listening tests.	542	<b>100</b>	399	43	,184	,388
29. Asking learners to listen to names or numbers is called intensive listening.	542	278	<b>126</b>	138	,232	,422
30. In selective listening, learners are expected to look for certain information.	542	<b>315</b>	187	40	,581	,493
<b>LISTENING-TOTAL</b>	<b>542</b>				<b>4,752</b>	<b>3,291</b>
<b>ASSESSING WRITING</b>						
31. Giving two options to learners and asking them to write about one ensure reliable and valid scoring.	542	312	<b>160</b>	70	,295	,456
32. Analytic scoring is used to see the strengths and weaknesses of learners.	542	<b>279</b>	177	86	,514	,500
33. The parts of a scoring scale and the scores in each part do not change for different levels of learners.	542	150	<b>335</b>	57	,618	,486
34. When there is a disagreement between the scores of the two raters, they score the written work again.	542	381	<b>134</b>	27	,247	,431
35. Learners are required to write about at least two tasks in the exam rather than one task.	542	<b>149</b>	309	84	,274	,44688
36. Giving restrictive prompts/guidelines to learners for the writing task is avoided.	542	155	<b>333</b>	54	,614	,487
37. Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.	542	420	<b>72</b>	50	,132	,339
38. Using visuals which guide learners for writing poses a problem.	542	50	<b>422</b>	70	,778	,415
39. Holistic scoring is used to see whether the learner is proficient or not at the end of the term.	542	<b>257</b>	161	124	,474	,499
40. Analytic scoring leads to greater reliability than holistic scoring in writing.	542	<b>216</b>	192	134	,398	,490
41. In controlled writing, learners have the chance to convey new information.	542	163	<b>261</b>	118	,481	,500
42. Classroom evaluation of learning in terms of writing is best served through analytic scoring rather than holistic scoring.	542	<b>214</b>	167	161	,394	,489
43. Irrelevant ideas are ignored in the assessment of initial stages of a written work in process writing.	542	173	<b>292</b>	77	,538	,498
44. Providing a reading text for writing is a way of assessing writing skills.	542	<b>250</b>	196	96	,461	,498

45. Mechanical errors (e.g. spelling and punctuation) are dealt with in the assessment of later stages of a written work.	542	<b>189</b>	298	55	,348	,477
<b>WRITING-TOTAL</b>	<b>542</b>				<b>6,573</b>	<b>2,478</b>
<b>ASSESSING SPEAKING</b>						
46. When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.	542	308	<b>191</b>	43	,352	,478
47. Giving learners one task is enough to assess speaking skills.	542	34	<b>486</b>	22	,896	,304
48. Interlocutors' showing interest by verbal and non-verbal signals poses a problem.	542	125	<b>386</b>	31	,712	,453
49. When it becomes apparent that the learner cannot reach the criterion level, the task is ended.	542	<b>157</b>	320	65	,289	,454
50. Using holistic and analytic scales at the same time poses a problem.	542	149	<b>231</b>	162	,426	,494
51. Reading aloud is a technique used to assess speaking skills.	542	<b>87</b>	380	75	,160	,367
52. In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.	542	<b>209</b>	277	56	,385	,487
53. In interactive tasks, more than two learners pose a problem.	542	<b>149</b>	316	77	,274	,446
54. The interlocutor gives the score when the learner is in the exam room.	542	72	<b>430</b>	40	,793	,405
55. In a speaking exam, production and comprehension are assessed together.	542	<b>282</b>	231	29	,520	,500
56. Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills.	542	<b>112</b>	359	71	,206	,405
57. Discussion among learners is a way of assessing speaking skills.	542	<b>312</b>	213	17	,575	,49470
58. A checklist is a means of scoring oral presentations in in-class assessment.	542	<b>288</b>	183	71	,531	,499
59. When the focus is to assess discourse, role plays are used.	542	<b>270</b>	166	106	,498	,500
60. In peer interaction, random matching is avoided.	542	<b>100</b>	342	100	,184	,388
<b>SPEAKING-TOTAL</b>	<b>542</b>				<b>6,808</b>	<b>2,784</b>
<b>LAKS-TOTAL</b>	<b>542</b>		<b>25,190</b>	<b>11,390</b>		

To reveal whether this mean score is statistically and significantly lower than the half of the total score, one sample t-test was applied. The findings are presented in Table 4 below. According to the values, it was found that the mean difference (4.81) between the participants' mean score (25.19) in the scale and the half of the maximum score (30) is statistically significant. That means their LAK level in general is significantly low.

**Table 5. One sample t-test results**

Mean diff.	df	t	p
<b>4.81</b>	541	-9.83	.000*

\*p<.05

One sample t-test was also applied for each skill to find out whether the mean score regarding each skill is significantly lower than the half of the total point for each skill. There were 15 items in each skill. The minimum and maximum scores for each skill were 0 and 15. Thus, the half of the total point was 7,5. The mean scores for each skill were 7,055 for assessing reading, 4,752 for assessing listening, 6,573 for assessing writing, and 6,808 for assessing speaking. The results shown in the Table 6 revealed that the participants' mean scores in each skill were significantly lower than the half of the total score.

**Table 6. One sample t-test results – skill based**

	Mean diff.	df	t	p
Assessing Reading	-,44	541	-2,31	.021*
Assessing Listening	-2,74	541	-19,42	.000*
Assessing Writing	-,926	541	-8,69	.000*
Assessing Speaking	-,691	541	-5,78	.000*

\*p&lt; .05

### 3.2. The Relationship among the Participants' Skill-based Assessment Knowledge

Another research question of the current study aimed to present how each skill-based knowledge correlated with the others and language assessment knowledge in general. For the analysis, Pearson correlation was employed and the findings are presented in Table 7. It is seen that all correlational values among the variables are significant. It was also found that all types of skill-based knowledge were highly and positively correlated with language assessment knowledge (LAK) in general. In addition to this, it was also revealed that all types of skill-based knowledge had high or moderate positive correlations among themselves. The highest correlational level was found between reading and listening (.816), whereas the lowest was between writing and speaking (.547) which is a moderate one. These high or moderate relationships among the skills mean that if EFL teachers' assessment knowledge in one skill increases, their assessment knowledge in others tends to increase in high or moderate levels.

**Table 7. The relationship among skill-based language assessment knowledge**

	LAK	Reading	Listening	Writing	Speaking
LAK	1	,933**	,908**	,749**	,852**
Reading		1	,816**	,573**	,737**
Listening			1	,597**	,689**
Writing				1	,547**
Speaking					1

\*\*Correlation is significant at 0.01 level; N=542

### 3.3. Effects of Demographic Features on LAK Level of the Teachers

The third research question of the study examined the language assessment knowledge of the participants in terms of the following variables: years of experience, educational background, the BA programme being graduated, workplace, testing course in BA, attending trainings on testing and being a testing office member. The findings belonging to each variable are presented in the tables below.

As for the first variable, years of teaching experience, the findings presented in Table 8 below revealed that there was no significant difference among the experience groups. Based on this, it can be said that teaching experience did not play a significant role on language teachers' LAK level.

**Table 8. Language assessment knowledge according to years of experience**

years of experience	N	M
1-5 years	86	24,97
6-10 years	173	25,03
11-15 years	114	24,86
16-20 years	100	25,62
more than 21 years	69	25,75

	Sum of Squares	df	Mean Square	F	p
<b>Between Groups</b>	60,28	4	15,071	,115	,977
<b>Within Groups</b>	70129,14	537	130,594		
<b>Total</b>	70189,42	541			

The second variable was the educational background. The results in Table 9 showed that the difference among the groups was not significant and the teachers' LAK level did not change according to their educational background.

**Table 9. Language assessment knowledge according to educational background**

Educational background	N	M
BA degree	238	25,508
MA degree	255	24,870
PhD degree	49	25,306

  

	Sum of Squares	df	Mean Square	F	p
<b>Between Groups</b>	50,805	2	25,403	,195	,823
<b>Within Groups</b>	70138,621	539	130,127		
<b>Total</b>	70189,426	541			

Additionally, as shown in Table 10, it was found that there was not a statistically significant difference between ELT and non-ELT graduates in terms of their LAK level. In other words, the programme being graduated, whether ELT or non-ELT, did not influence the language assessment knowledge of the teachers.

**Table 10. Language assessment knowledge according to the programme being graduated**

BA Graduation	N	Mean	Std. Deviation	Std. Error Mean
English Language Teaching	347	25,42	11,722	,629
Non-ELT	195	24,76	10,790	,772
<b>Mean diff.</b>	<b>df</b>	<b>t</b>		<b>p</b>
<b>,657</b>	540	,644		,52

Whether the teachers in this study worked at a state or private university was another variable investigated. Based on the results, it can be seen in Table 11 that there was not a significant difference between these two groups and workplace was found to have no effect on the teachers' language assessment knowledge.

**Table 11. Language assessment knowledge according to the workplace**

Workplace	N	Mean	Std. Deviation	Std. Error Mean
at a state university	372	25,346	11,565	,599
at a private university	170	24,847	11,022	,845
<b>Mean diff.</b>	<b>df</b>	<b>t</b>		<b>p</b>
<b>,499</b>	540	,474		,63

Another variable was the separate testing course in BA and the findings revealed that there is not a significant difference between these two groups. In other words, it can be said that the testing and assessment course given in BA programmes did not have a significant effect on the teachers' LAK level as shown in Table 12 below.

**Table 12. Language assessment knowledge according to testing course in BA**

A separate testing course in BA	N	Mean	Std. Deviation	Std. Error Mean
Yes	260	25,019	12,045	,747
No	282	25,347	10,769	,641
<b>Mean diff.</b>	<b>df</b>	<b>t</b>		<b>p</b>
<b>-,328</b>	540	-,335		,73

Attending trainings on testing and assessment was another variable. The results indicated that there was no significant difference between these two groups, which means that the training received on language assessment did not have a significant impact on the teachers' LAK level. The findings are shown below in Table 13.

**Table 13. Language assessment knowledge according to the attendance to trainings**

Attending any trainings in LTA	N	Mean	Std. Deviation	Std. Error Mean
Yes	282	25,741	11,967	,712
No	260	24,592	10,720	,664
<b>Mean diff.</b>	<b>df</b>	<b>t</b>	<b>p</b>	
<b>1,148</b>	<b>540</b>	<b>1,17</b>	<b>.24</b>	

Among all variables examined in this research question, the only significant difference was found in terms of being a testing office member or not. The findings shown in Table 14 indicated that there was a statistically significant difference between these two groups, and the LAK level of the participants having worked as a member of testing office was higher than the others. Based on this, it can be concluded that working on testing, doing institutional staff and being involved with some practical elements related with testing and assessment might have a positive impact on the LAK level of the teachers.

**Table 14. Language assessment knowledge according to being a testing office member**

Being a testing office member	N	Mean	Std. Deviation	Std. Error Mean
Yes	260	26,303	11,710	,726
No	282	24,163	11,007	,655
<b>Mean diff.</b>	<b>df</b>	<b>t</b>	<b>p</b>	
<b>2,140</b>	<b>540</b>	<b>2,19</b>	<b>.02</b>	

### 3.4. Perceived Self-competency and Actual Language Assessment Knowledge Level

Whether the teachers' LAK level changed according to their perceived self-competency was investigated based on each language skill. The findings derived from one-way ANOVA analysis are presented in the following tables. It is indicated in Table 15 below that almost 95% of the participants perceived themselves competent or very competent. On the other hand, the ones who thought that they were not very competent in assessing reading, had the highest mean score among all. According to the findings, no significant difference was found among the participants who perceived themselves as very competent, competent, and not very competent in terms of their LAK level in reading. However, it can be clearly seen that the teachers' perceived self-competency in assessing reading is far from their actual LAK level.

**Table 15. Perceived self-competency of the teachers and their LAK level in assessing reading**

Assessing Reading	N	M			
very competent	152	6,769			
competent	355	7,019			
not very competent	34	8,676			
	<b>Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>p</b>
<b>Between Groups</b>	102,202	2	51,10	2,567	.078
<b>Within Groups</b>	10709,244	538	19,90		
<b>Total</b>	10811,445	540			

In the findings of the previous research questions, the participant teachers were the least knowledgeable in assessing listening. On the contrary, the findings shown in Table 16 regarding their perceived self-competency tells the opposite since more than 80% of the teachers perceived themselves as competent or very competent. In addition to this, there was not a

significant difference among the perception groups in terms of their LAK level in assessing listening, and again, it was found that the ones who perceived themselves as not very competent had the highest mean score compared to the others.

**Table 16. Perceived self-competency of the teachers and their LAK level in assessing listening**

Assessing Listening	N	M			
very competent	112	4,821			
competent	338	4,695			
not very competent	89	4,943			
	Sum of Squares	df	Mean Square	F	p
Between Groups	4,843	2	2,42	,222	.80
Within Groups	5834,760	536	10,88		
Total	5839,603	538			

In terms of their LAK level in assessing writing, as Table 17 indicates, no significant difference was found among the perception groups. It was also revealed that more than 90% of the teachers perceived themselves as competent or very competent in assessing writing though their actual LAK level in this skill was 6.573 over 15. Finally, it is again seen that the ones who perceived themselves as not very competent had the highest mean score compared to the other groups.

**Table 17. Perceived self-competency of the teachers and their LAK level in assessing writing**

Assessing Writing	N	M			
very competent	161	6,347			
competent	333	6,657			
not very competent	45	6,866			
	Sum of Squares	df	Mean Square	F	p
Between Groups	14,381	2	7,19	1,181	.30
Within Groups	3264,695	536	6,09		
Total	3279,076	538			

For the last skill, assessing speaking, the findings were similar to the others. As Table 18 below demonstrates, there was no significant difference in terms of LAK level in assessing speaking among the participants based on their perceived self-competency in this skill. Again, almost 85% of the teachers perceived themselves as competent or very competent though they demonstrated a LAK level of 6.808 over 15, which shows a difference between their perceptions and actual level. Finally, the last important point was again similar to the other skills and the ones with “not very competent” perception had a relatively higher level of LAK in assessing speaking compared to the other groups.

**Table 18. Perceived self-competency of the teachers and their LAK level in assessing speaking**

Assessing Speaking	N	M			
very competent	129	7,0620			
competent	336	6,6786			
not very competent	74	6,918			
	Sum of Squares	df	Mean Square	F	p
Between Groups	14,851	2	7,42	,953	.38
Within Groups	4174,303	536	7,78		
Total	4189,154	538			

## 4. DISCUSSION

### 4.1. Language Assessment Knowledge Level of EFL Teachers

The teachers in the current study indicated a relatively low level of language assessment knowledge. In ELT, there are few studies aiming at revealing the language assessment knowledge of teachers. The rarity of these studies in ELT was also mentioned by Hatipoğlu (2017). The findings of the current study are in line with those of Tao (2014), Mede and Atay (2017), and Xu and Brown (2017) where they also found that language teachers had low language assessment knowledge. Popham (2009) also stated that most of the teachers do not have adequate knowledge related to language assessment, and discussed the severity of the situation by saying that for most of the teachers, test “is a four-letter word, both literally and figuratively” (p. 9).

When it comes to skill-based language assessment knowledge level of the participant teachers, the findings indicated that the participant teachers are more competent and knowledgeable in assessing reading when compared to assessing other skills. Why assessing reading has higher mean scores can be found in the utterances of Hubley (2012), and Backlund, Brown, Gurry and Jandt (1980). Hubley (2012) stated that there is agreement among scholars in the argument that reading is a crucial skill, and even maybe the most important one, and much of the input comes from reading sources surrounding us. Because of the density of input surrounding the learners in the classroom as well, learners have to read a lot. As reading skill is given importance, teaching it is highly valued, and, it is assessed by the teachers as a natural consequence. There are various ready-made materials for assessing reading; thus, it does not become a challenge for teachers to assess reading skills of their learners (Backlund, Brown, Gurry, & Jandt, 1980). Why the other three skills had lower mean scores were mentioned in the literature by touching upon the difficulties each skill possesses.

For assessing writing, Weigle (2012) expressed that assessing writing could be perceived as something easy, and people may think that teachers only give the topic and ask learners to write on that topic. Indeed, it is not as easy as people may think, because just giving the topic and asking learners to write on that topic is not a good way of assessing writing. Speaking is also regarded highly important because of the oral communication taking place a lot in our lives (Heaton, 1990). Madsen (1983) pointed out that speaking is the most difficult skill to assess because of its subjectivity and complex nature, because teachers do not know what and how to assess regarding speaking skill. In other skills, they have ready-made materials provided by the coursebooks and publishing companies; but, in speaking they are all alone.

As for listening, Flowerdew and Miller (2012) discussed that assessing listening is perceived by both learners and teachers as an issue which somehow improves by itself. Buck (2001) also mentioned this problem by stating that listening is neglected in terms of teaching and assessing, which is one of the findings of the current study. To draw attention to the ignorance, Flowerdew (1994), Nunan and Miller (1995) and Flowerdew and Miller (2005) stated that listening skill is a ‘Cinderella’ skill which majority of teachers take for granted. For Buck (2001), why listening is neglected lies on the complicated nature of listening as a skill and practicality issues related to assessing listening.

Furthermore, in Turkish context, why these three skills got lower scores could result from ill-prepared assessment programmes, lack of ongoing workshops alongside the difficulty of measuring these skills. The assessment courses in pre-service education is restricted to only one course in the last term of ELT programme. Hence, such a broad topic has to be covered in one course both theoretically and practically, which is not very possible. What is more, there is lack of ongoing workshops specifically designed for language assessment for in-service teachers. Lastly, these three skills, as mentioned above, are difficult to measure when compared to assessing reading.

## 4.2. The Relationship among the Participants' Skill-based Assessment Knowledge

The findings revealed that all the items are correlated with LAK and the skills have high or moderate positive correlations among themselves. These results indicated that all types of skill-based assessment knowledge are important elements of LAK and if teachers are trained to be more knowledgeable in assessing one skill, it is highly probable that their LAK in general will increase as well. This finding might lead us to perceive language assessment knowledge as a holistic phenomenon with its own interrelated elements. In addition to this, it was also revealed that all types of skill-based knowledge had high or moderate positive correlations among themselves. These high or moderate relationships among the skills mean that if EFL teachers' assessment knowledge in one skill increases, their assessment knowledge in others tends to increase in high or moderate levels. This finding again put forward that all types of skill-based assessment knowledge might be considered as interrelated elements.

The probable reason for this is that all the skills, though different in nature, serve for the same purpose which is LAK, and the logic behind the assessment of all skills is similar. For instance, when a teacher's knowledge in designing tasks such as multiple choice or open-ended in reading increases, that teacher can transfer this knowledge into other skills, and makes use of that knowledge in others. Another example could be the use of at least two tasks in assessing each skill. When a teacher has learnt that at least two tasks are needed to assess writing skills more reliably, then the same information could be used in other skills as well. Consequently, this increase in knowledge in one skill affects the knowledge in other skills positively, and also results in increased knowledge in LAK.

## 4.3. Effects of Demographic Features on LAK Level of the Teachers

The findings revealed that years of experience, the BA programme being graduated, educational background, workplace, testing course in BA, and attending trainings on testing and assessment do not have an effect on LAK level of the participants whereas being a testing office member has an influence on LAK level of the teachers.

To start with years of experience, in Tao (2014)'s study, it was revealed that there is not a relationship between years of experience and the actual LAK level of teachers. Thus, the results of the current study are in parallel with the findings of Tao (2014). The possible explanation might be that language assessment is not a topic that could be learned or acquired on the job, and there should be some extra driving forces for the teachers to have this knowledge. As the findings showed that the BA programme has no effect on the LAK level of the teachers, it is clear that the teachers start their jobs with insufficient knowledge in language assessment (Hatipoğlu, 2017). In other words, teaching experience, per se, does not necessarily increase the language assessment knowledge and skills of teachers. Regarding BA programme being graduated from, the results indicated that the teachers who are ELT-graduates and non-ELT graduates are not different in terms of their language assessment knowledge. The reason for the similarities of both groups could be that language assessment is not given a priority in ELT programmes and covered in one course at the fourth grade; thus, the graduates of ELT and non-ELT are not different with respect to their language assessment knowledge.

With respect to educational background, there are no significant differences among the teachers having BA degree, MA degree and PhD degree. As previously mentioned, the BA programmes were insufficient in terms of exposure in language assessment knowledge, and this finding might underline the situation that this insufficiency may not be solely the problem of pre-service education, but also might be the problem of MA and PhD programmes. Even if there is one course related to language assessment, compulsory or elective, in post-graduate level, all the topics have to be covered in one course in one academic term period, which might be short for this broad topic. These possible reasons may have led to this finding. This finding is in line

with Tao's (2014) study, showing that there is not a statistically significant difference between the teachers whose educational background is BA and the ones with MA.

Related to workplace, the findings revealed that there is not a significant difference between the teachers working at a state and private university. Private universities tend to have more training and professional development programmes. However, the presence of the professional development programmes may not guarantee the increased knowledge in all fields of language including language testing and assessment, which is also one of the findings of the current study regarding the relationship between the existence of training and professional development programmes and language assessment knowledge.

Also, it was found that there is not a significant difference between the teachers who had a separate testing and assessment course in those pre-service education and those who did not. The findings of this study could be supported by the findings of Tsagari (2008) and Tao (2014) who stated that the participants had inadequate assessment training in pre-service education. In Turkish context, this finding is in line with Köksal (2004), Mede and Atay (2017) and Hatipoğlu (2015; 2017)'s studies which stated that pre-service education is insufficient in terms of equipping pre-service teachers with necessary language assessment knowledge, and additions were needed in language testing and assessment course in pre-service education.

This finding might also lead us to the conclusion that the testing and assessment course in the program might not present a sufficient content to pre-service teachers as stated by Hatipoğlu (2015). According to this study, one course in language testing and assessment in pre-service education resulted in lack of basic training of learners in language assessment. The insufficiency of this course in pre-service education may result from several probable causes. One might be related to the competency of the teacher educators giving those courses in pre-service education. These teacher educators should be equipped with a lot of knowledge related to language assessment. Stiggins (1999), Hatipoğlu (2012) and Jeong (2013) emphasized that the teacher educators who are responsible for this language assessment course at university should have a solid background in language assessment. The second one might be the arbitrariness of the content of these courses. There is not a framework for the syllabus design for these courses, and the teacher educators giving those courses decide on the content of these courses (Hatipoğlu, 2015). The last one is even though there is a specifically designed course, the presence of this course may not be enough to cover all the information related to assessing each language skill comprehensively in just one academic term period. The learners may not have sufficient time to become familiar with all the issues related to assessing language skills, and they may also not have time to make practice such as going through ready-made exams and deciding on the appropriacy of the tasks, or designing tasks. As they are not involved in these tasks, it is more likely that the presence of that separate course in pre-service education may not be very efficient for the teachers. The importance of practice was also stated by Jin (2010).

The last variable which does not have an effect on the LAK level of the participant teachers is attending trainings on testing and assessment. This might lead testing and assessment practitioners in Turkey to re-consider the content and quality of trainings and workshops in Turkish context due to their relative effectiveness perceived by the participants since such trainings were found to be significantly influential on teachers' professional development in this domain. For instance, several scholars in the field (McNamara & Roever, 2006; Stiggins, 2010; Mede & Atay, 2017) stated that the reason why teachers are assessment illiterate is lack of professional development programmes. Moreover, it was expressed that training in language assessment does not lead to increased knowledge in language assessment and the trainings should go beyond applied psychometrics, and should have a comprehensive and to the point content. Malone (2008) also stated that training is not enough itself. Trainings should "include the necessary content for language instructors to apply what they have learned in the classroom and understand the available resources to supplement their formal training when they enter the

classroom (p. 235). Attending trainings is not sufficient for a language teacher. As stated in Malone's sentences, trainings should go hand in hand with other efforts of language teachers. This inefficiency of the trainings or programmes on the language assessment knowledge of the teachers may rely on the fact that there are not many trainings or professional development programmes on language assessment, especially there is not a conference solely focusing on language assessment with respect to assessing four skills in Turkey. Another reason could be the sustainability of these programmes. Half of the participants in Mede and Atay's (2017) study stated that they had training in language assessment, but they were short and one-shot training. Hence, sustainability of the programmes may also play a role in increasing language assessment knowledge of teachers.

Being a testing office member is the only variable that makes a difference, and has an influence on language assessment knowledge of the participants. The results showed that there is a significant difference between the testing office members and the ones who are not in testing office. This finding was another focus of the qualitative data, and for the fourth question of the open-ended protocol, the participants were asked to comment on the significant impact of being a testing office member on language assessment knowledge of teachers. The respondents expressed that when teachers are more involved in assessment-related activities, they learn more. According to them, as testing office members have to be involved in assessment-related activities in testing office, they naturally learn more.

#### **4.4. Perceived Self-competency and Actual Language Assessment Knowledge Level**

The results showed that there is not a significant difference among the participants who perceived themselves as very competent, competent, and not very competent in terms of their LAK level in assessing reading, listening, writing and speaking. Furthermore, the findings indicated that the majority of the participants perceived themselves competent or very competent. On the other hand, the ones who thought that they were not very competent in assessing each skill had the highest mean score among all. It can be concluded that the participant teachers' perceived self-competency in assessing these four skills is far from their actual LAK level. With respect to perceived self-competency, the finding of this study shows parallelism with Öz and Atay (2014) and Jannati (2015)'s studies in which the participant teachers reported that they were familiar with the concepts related to language assessment such as the features of a good test and they had enough knowledge about the concepts and terminology related to language testing and assessment.

### **5. CONCLUSION AND SUGGESTIONS**

This study revealed general and skill-based language assessment knowledge level of EFL teachers in Turkish higher education context, and it was seen that the teachers had insufficient language assessment knowledge. When the skills were analysed in detail, the highest mean score belonged to assessing reading, 7,055, and the lowest mean score belonged to assessing listening, 4,752. Besides, it was seen that all skills were highly and positively correlated with LAK in general, and with each other as well. When the effects of demographic features on LAK level of the teachers were investigated, it was seen that the only significant difference was found among the participants in terms of being a testing office member or not. Testing office members were found to have higher mean scores.

As for the limitation of this study, the context in which the study was conducted could be a limitation. As the setting is limited to the preparatory programmes of the universities in Turkey, the results reflect the language assessment knowledge level of EFL teachers in higher education setting in Turkey.

There are some implications and suggestions for further research. Based on all these findings, this study comes up with certain implications:

- As is clear from the findings, pre-service education has some limitations in terms of language testing and assessment, and pre-service teachers are not equipped with necessary knowledge in pre-service education related to language assessment. Thus, the content of the course in pre-service education might be considered to be revised. There should be more than one course related to language assessment, and more practical hands-on practice can be incorporated into these courses in pre-service education. The course should not only be based on theory but it should also have sufficient time to practice.
- Trainings and professional development programmes could be designed based on both theory and practice related to their needs, and language teachers could be supported and encouraged to attend the conferences and professional development programmes on language testing and assessment.
- A training module could be designed which is solely based on language testing and assessment regarding four skills. In these training programmes, teachers could be provided with basic, practical and to the point information related to each skill, and they can work on real exams and could be asked to make comments on ready-made exams. Thus, they can have the chance to combine theory and practice, and the training becomes more meaningful.

The results of this study have opened many doors for future studies. To start with, this study is restricted to the participants working at preparatory programmes of universities in Turkey. The same scale could be administered to the language teachers working at Ministry of Education and pre-service teachers in ELT departments. This measurement tool could also be used in other countries to indicate the language assessment level of language teachers. In addition to these, some cultural, linguistic and context-specific elements could be added to the scale, and this could be carried out with teachers.

## 6. REFERENCES

- Aydın, B., Kızıltan, N., Öztürk, G., İpek, Ö. F., Yükselir, C., & Beceren, S. (2017). Optional English preparatory programs after HEC 2016 regulation: Opinions of administrators on the current situation and problems. *Anadolu University Journal of Education Faculty*, 1 (2), 1-11.
- Backlund, P., Brown, K., Gurry, J., & Jandt, F. (1980). Evaluating speaking and listening skill assessment instruments: Which one is best for you? *Language Arts*, 57 (6), 621-627.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Flowerdew, J., & Miller, L. (2012). Assessing listening. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 225-233). Cambridge University Press, USA.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9 (2), 113-132.
- 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. (2017). History of English language teacher training and English language testing and evaluation (ELTE) education in Turkey. In Y. Bayyurt and N. S. Sifakis (Eds). *English Language Education Policies and Practices in the Mediterranean Countries and Beyond* (pp. 227-257). Frankfurt: Peter Lang.
- Heaton, J. B. (1990). *Writing English language tests*. (2nd ed.). Cambridge: Cambridge University Press.
- Hubley, N. N. (2012). Assessing reading. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 211-217). Cambridge University Press, USA.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25 (3), 385-402. *Language Testing and Assessment* (pp. 1-14). Springer International Publishing.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6 (2), 26-37.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers?. *Language Testing*, 30 (3), 345-362.

- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584.
- Köksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education (TOJDE)*, 5(1), 1- 11.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32 (2), 169-197.
- Madsen, H. S. (1983). *Techniques in testing*. New York and Oxford: Oxford University Press.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA, and Oxford, England: Blackwell.
- Mede, E., & Atay, D. (2017). English Language Teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168 (1), 1-5.
- Mertler, A. C. (2003). Secondary teachers' assessment literacy: *Does classroom experience make a difference?*. *American Secondary Education*, 33(1), 49-64.
- Ölmezer-Öztürk, E. & Aydın, B. (2018). Developing and Validating Language Assessment Knowledge Scale (LAKS) and Exploring the Assessment Knowledge of EFL Teachers. Unpublished PhD Dissertation, Anadolu University, Turkey.
- Öz, S., & Atay, D. (2017). Turkish EFL teachers' in-class language assessment literacy: perceptions and practices. *ELT Research Journal*, 6 (1), 25-44.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6 (1), 21-27.
- Popham, J. W. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48, 4-11.
- Price, M., Rust, C., O'Donovan, B., Handley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford: The Oxford Centre for Staff and Learning Development.
- Stanford, P., & Reeves, S. (2005). Assessment that drives instruction. *Teaching Exceptional Children*, 37 (4), 18-22.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77 (3), 238-245.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18 (1), 23-27.
- Stiggins, R. J. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade, G. J. Cizek (Eds.), *Handbook of formative assessment*, (pp. 233-250). New York, NY: Taylor & Francis.
- Tao, N. (2014). *Development and validation of classroom assessment literacy scales: English as a Foreign Language (EFL) teachers in a Cambodian Higher Education Setting*.
- Tsagari, D. (2008). *Assessment literacy of EFL teachers in Greece: Current trends and future prospects*. PowerPoint presentation at the 5th Annual EALTA Conference, May 9-11, Athens, Greece.
- Tsagari, D. & Vogt, K. (2017). Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. *Papers in Language Testing and Assessment*, 6 (1), 41-64.
- Weigle, S. C. (2012). Assessing writing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyanoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 236-246). New York: Cambridge University Press.
- White, E. (2009). Are you assessment literate?. *OnCue Journal*, 3 (1), 3-25.
- Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6 (1), 133-158.

## Geniş Özet

Bu çalışma, Türkiye'de yükseköğretim bağlamında çalışan İngilizce öğretmenlerinin dilde ölçme değerlendirme bilgisi seviyelerini ortaya koymayı amaçlamaktadır. Bu çalışma için, üniversitelerin hazırlık programlarında çalışan dil öğretmenleri özellikle seçilmiştir, çünkü bu programlarda çalışan öğretmenler ölçme değerlendirmeden sorumludurlar ve bu programlarda bütün beceriler ölçülür. Nicel veriye dayanan bu çalışmada, Dilde Ölçme Değerlendirme Ölçeği (LAKS) (Ölmezer-Öztürk & Aydın, 2018) ile katılımcılardan veri toplanmıştır. Bu ölçek, uzman görüşü, dil öğretmenleriyle görüşmeler ve ölçme-değerlendirme ofisinde görev alan dil öğretmenleriyle toplantı gibi kapsamlı bir geçerlilik çalışması sürecinden geçmiştir. LAKS, 60 maddeli ve 4 boyutlu bir ölçektir ve boyutlar okuma bilgisini ölçme, dinleme bilgisini ölçme, yazma bilgisini ölçme ve konuşma bilgisini ölçme olarak ayrılmıştır.

Ölçeğin geçerliliği ve güvenilirliği ile ilgili istatistiksel bulgular ölçeğin mükemmel model veri uyumuna sahip olduğunu ve güvenilirliğin yüksek olduğunu ortaya koymuştur ( $r=.91$ ).

Bu ölçek, Türkiye’de devlet ve özel olmak üzere 53 farklı üniversitede çalışan 542 öğretmen tarafından cevaplanmıştır. Öğretmenlerin bu ölçekteki maddelere verdikleri cevaplar incelendiğinde, katılımcıların ölçeğin genelinden 60 üzerinden 25 aldığı ve bu ortalamanın toplam puanın yarısından anlamlı derecede düşük olduğu sonucuna varılmıştır. Tek örneklem testinin sonuçları da göz önüne alındığında, dil öğretmenlerinin dilde ölçme değerlendirme bilgilerinin yeterli olmadığı sonucuna ulaşılabilir. Katılımcıların her bir beceriyi ölçme bilgisi incelendiğinde, en yüksek ortalamayı okuma bilgisini ölçmede, en düşük ortalamayı ise dinleme bilgisini ölçmede aldıkları görülmüştür. Yani, dil öğretmenlerin en çok bilgili oldukları alan okuma bilgisini ölçme iken en az bilgili oldukları alan dinleme bilgisini ölçme olmuştur. Katılımcıların en çok okuma bilgisini ölçmede iyi oldukları ortaya kalsa da, genel ortalamalar ve istatistikî bulgulara bakıldığında, öğretmenlerin en yüksek ortalamaya sahip oldukları beceri de bile ölçme değerlendirme bilgilerinin yetersiz olduğu görülmüştür. Tek örneklem testi, her bir beceri için yapılmıştır ve sonuçlar katılımcıların her bir beceri için ortalama puanının toplam puanın yarısından anlamlı derecede düşük olduğunu göstermiştir. Ayrıca, Pearson korelasyon analizi sonuçlarına göre, değişkenler arasındaki bütün korelasyon değerleri anlamlılık göstermektedir. Bütün becerilerin hem kendileri arasında hem de tepe kavram dilde ölçme değerlendirme bilgisi ile yüksek ve olumlu korelasyon gösterdiği çalışmanın bulguları arasındadır. Bu da şu anlama gelmektedir. Katılımcıların herhangi bir beceride ölçme değerlendirme bilgisinin artması hem diğer becerilerdeki ölçme değerlendirme bilgilerinin artmasına hem de tepe kavram olan dilde ölçme değerlendirme bilgisinin artmasını sağlayacaktır. Bu da dilde ölçme değerlendirme bilgisinin bütüncül bir kavram olduğunu ve bu kavramı oluşturan alt başlıkların da birbirleriyle ilişkili olduğu sonucunu ortaya koymaktadır. Ayrıca, en yüksek korelasyon okuma ve dinleme bilgisini ölçme arasındayken ( $.816$ ), en düşük korelasyon yazma ve konuşma bilgisini ölçme arasındadır ( $.547$ ). Bunlara ek olarak, ölçme değerlendirme ofisi çalışanı olup olmama haricinde hiçbir demografik değişkenin öğretmenlerin dilde ölçme değerlendirme bilgisi üzerinde etkisinin olmadığı da çalışmanın istatistikî bulguları arasındadır. Ölçme-değerlendirme ofis çalışanı olanların ortalamalarının bu ofiste olmayanlardan daha yüksek olduğu görülmüştür. Diğer demografik özelliklerin, mesleki tecrübe, öğrenim geçmişi, mezun olduğu lisans programı, özel veya devlet üniversitesinde çalışması, lisans eğitiminde ayrı bir ölçme değerlendirme dersi alması, ölçme değerlendirme ile ilgili konferanslara ve mesleki gelişim programlarına katılmasının, katılımcıların dilde ölçme değerlendirme bilgileri üzerinde bir etkisi görülmemiştir. Son olarak, katılımcıların kendini her beceri bazında öz algılarının ölçme değerlendirme bilgileri üzerinde bir etkisi olup olmadığına bakıldı ve sonuçlar anlamlı bir farklılığın olmadığını ortaya çıkardı. Sonuçlar, katılımcıların büyük çoğunluğunun kendilerini her beceriyi değerlendirmek için yeterli ve ya çok yeterli gördüğünü, yetersiz seçeneğini seçenlerin yok denecek kadar az olduğunu göstermiştir. Bu da öğretmenlerin kendilerini dilde ölçme değerlendirme konusunda nasıl algıladıkları ve ölçeğin sonuçlarına göre çıkan ortalamalar arasında büyük bir farklılığın olduğunu ortaya çıkarmıştır.

Öğretmenler kendilerini her beceri için yeterli görmelerine rağmen, sonuçlar öğretmenlerin dilde ölçme değerlendirmede büyük bir eksiğinin olduğunu göstermiştir. Sonuç olarak, bu çalışmanın bulguları, yüksek öğrenimde çalışmakta olan dil öğretmenlerinin ölçme değerlendirme bilgilerinin yetersiz olduğunu ortaya çıkarmıştır. Bu sonuçlar, öğretmenlerinin güçlü ve zayıf yönlerinin belirlenmesine yönelik bulgular ortaya koymaktadır. Bunlar göz önüne alınarak, öğretmenlerin her bir beceriye ilişkin ihtiyaçları belirlenip, öğretmenlere ölçme değerlendirmede mesleki gelişim programları hazırlanmalıdır ve ihtiyaçlar doğrultusunda mesleki gelişimlerine yardımcı olunmalıdır. Dilde ölçme değerlendirmeye odaklanan konferans sayısı yok denecek kadar azdır, bu konferansların çoğu ölçme değerlendirmeyi dil bazında değil de daha genel olarak odaklanmaktadır. Öncelikle, dilde ölçme değerlendirmeye odaklanan konferansların sayısı artırılmalı ve yükseköğretim kurumunda çalışma dil öğretmenlerinin bu tür konferanslara katılımları teşvik edilmelidir. Ayrıca, lisans programındaki ölçme değerlendirmenin yetersiz kaldığı da çalışmanın sonuçları arasındadır. Daha kapsamlı ve pratiğe dayalı birden fazla dilde ölçme değerlendirmeye ilişkin ders açılması, lisans öğrencilerinin ölçme değerlendirme alanında daha bilgili olarak mesleğe başlamalarına yardımcı olacaktır. Bu çalışma birçok çalışmaya zemin oluşturmaktadır. Bu ölçek kullanılarak Milli Eğitim’de çalışan İngilizce öğretmenlerinin ve İngilizce Öğretmenliği programında öğrenim gören lisans öğrencilerinin de dilde ölçme değerlendirme bilgisi seviyeleri ortaya konabilir. Ayrıca, dil öğretmenlerinin bildiklerinin ne kadarını kullanabildikleri ve pratiklerinin ne kadar farkında oldukları da araştırılabilir.