

**REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM**

MASTER'S THESIS

**BEYOND THE SURFACE: AN INVESTIGATION INTO THE
LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS
WORKING AT K12 SCHOOLS**

**ÖZLEM DOĞAN
20708008**

**THESIS SUPERVISOR
Asst. Prof. BURCU ÜNAL**

**İSTANBUL
2023**

**REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM**

MASTER'S THESIS

**BEYOND THE SURFACE: AN INVESTIGATION INTO THE
LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS
WORKING AT K12 SCHOOLS**

ÖZLEM DOĞAN

20708008

ORCID NO: 0009-0007-0468-1973

**THESIS SUPERVISOR
Asst. Prof. BURCU ÜNAL**

JULY 2023

The study “Beyond the Surface: An Investigation into the Language Assessment Literacy of EFL Teachers Working at K12 Schools” which was carried out by Özlem DOĞAN, was found successful unanimously as a result of the defense exam held on **05/07/2023** and has been approved as **MASTER’S THESIS** in English Language Teaching Program in Department of Foreign Languages Education by our jury.

Advisor

Asst. Prof. Burcu ÜNAL

Signature

.....

Jury Members

Prof. Dinçay KÖKSAL

Assoc. Prof. Suzan KAVANOZ

Signature

.....

.....

ABSTRACT

BEYOND THE SURFACE: AN INVESTIGATION INTO THE LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS WORKING AT K12 SCHOOLS

The main purpose of the current study is to investigate the general and skill-based language assessment literacy levels of EFL teachers working at K12 schools in Turkey and to analyze their needs both in pre-service and in-service years. Considering the purpose of the study, a mixed-method research design was employed. In the 1st phase of the study, the quantitative data were collected by means of the “Language Assessment Knowledge Scale”, and 272 EFL teachers working at either private or state K12 schools in Turkey participated. The responses were analyzed through descriptive statistics and the results showed that the participants’ mean score in LAKS was 31.59 out of 60. The mean scores for each skill were found out as 9,58 for reading, 6,82 for listening, 8,31 for speaking and 6,88 for writing over 15. Assessing reading was ranked as the area the participants are the most knowledgeable in whereas assessing listening was ranked as the area the teachers are the least knowledgeable in. In the 2nd phase of the study, semi-structured interviews were conducted with 10 EFL teachers to learn more about the teachers’ opinions on language assessment knowledge and their needs. Qualitative findings complemented the findings from the questionnaire by revealing details about the reasons of the participants’ scores in LAK scale. The overall results of the current study indicated that EFL teachers working at K12 schools in Turkey didn’t have sufficient knowledge in language testing and assessment and needed further assessment related trainings throughout both their pre-service and in-service years.

Keywords: Language testing and assessment, Language assessment literacy, language assessment knowledge, EFL teachers, K12 schools

ÖZET

K12 OKULLARINDA ÇALIŞAN İNGİLİZCE ÖĞRETMENLERİNİN DİL ÖLÇME DEĞERLENDİRME OKURYAZARLIĞININ İNCELENMESİ

Bu çalışmanın temel amacı, Türkiye'deki K12 okullarında görev yapan İngilizce öğretmenlerinin genel ve beceriye dayalı dil değerlendirme okuryazarlık düzeylerini araştırmak ve hem hizmet öncesi hem de hizmet yıllarında dil ölçme değerlendirme ile ilgili ihtiyaçlarını analiz etmektir. Araştırmanın amacı göz önünde bulundurularak karma yöntemli araştırma deseni kullanılmıştır. Araştırmanın 1. aşamasında nicel veriler “Dil Değerlendirme Bilgi Ölçeği” aracılığıyla toplanmış ve Türkiye'deki özel veya devlet K12 okullarında görev yapan 272 İngilizce öğretmeni katılım göstermiştir. Yanıtlar, betimsel istatistiklerle analiz edilmiş ve sonuçlar katılımcıların ölçek puan ortalamasının 60 üzerinden 31,59 olduğunu göstermiştir. Her bir beceri için ortalama puan, okuma için 9.58, dinleme için 6.82, konuşma için 8.31 ve yazma için 6.88 olarak bulunmuştur. Okuma becerisinin değerlendirilmesi öğretmenlerin en bilgili oldukları alan olarak bulunurken, dinleme becerisinin değerlendirilmesi öğretmenlerin en az bilgi sahibi oldukları alan olarak bulunmuştur. Çalışmanın 2. aşamasında öğretmenlerin dil değerlendirme bilgisi ve ihtiyaçları hakkında daha fazla bilgi edinmek için 10 EFL öğretmeni ile yarı yapılandırılmış görüşmeler yapılmıştır. Niteliksel bulgular, katılımcıların dil değerlendirme bilgi ölçeğindeki puanlarının nedenleri hakkında ayrıntıları ortaya koyarak anketten elde edilen bulguları tamamlamıştır. Mevcut çalışmanın genel sonuçları, Türkiye'deki K12 okullarında çalışan İngilizce öğretmenlerinin dil ölçme ve değerlendirme konusunda yeterli bilgiye sahip olmadığını ve hem hizmet öncesi hem de hizmet yıllarında dil ölçme değerlendirme ile ilgili daha fazla eğitime ihtiyaç duyduklarını göstermiştir.

Anahtar Kelimeler: Dilde ölçme ve değerlendirme, dil değerlendirme okuryazarlığı, dil değerlendirme bilgisi, İngilizce öğretmenleri, K12 okulları

ACKNOWLEDGEMENTS

At the end of this challenging and rewarding academic journey, I am proud and filled with immense gratitude. I would like to express my sincerest thanks to people who have an important role in the completion this thesis.

First of all, I would like to express my heartfelt gratitude to my thesis advisor Asst. Prof. Burcu ÜNAL for her invaluable guidance, mentorship and insightful feedback. I am grateful for her support and patience in every step of the way.

Additionally, I would like to thank my incredible friends who stood by me all the time. Thanks for offering an ear during the moments of doubt. Your presence has made this journey more meaningful and enjoyable. Your support and contribution along the way were one of my biggest motivators.

Finally, I wholeheartedly extend my deepest appreciation to my beloved family. Your endless support, understanding and motivation you have provided have been invaluable. I am very grateful for your presence in my life. Thank you.

ÖZLEM DOĞAN

July, 2023; İstanbul

TABLE OF CONTENTS

ABSTRACT	iii
ÖZET	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xii
1. INTRODUCTION	1
1.1. Background of the Study	1
1.2. Statement of Problem	2
1.3. Purpose of the Study and Research Questions	3
1.4. Significance of the Study	4
1.5. Definitions of Terms	5
2. LITERATURE REVIEW	6
2.1. Assessment, Evaluation and Testing	6
2.2. Types of Assessment in Language Education	6
2.2.1. Formative Assessment and Summative Assessment	7
2.2.2. Direct Assessment and Indirect Assessment	7
2.2.3. Discrete-point Assessment and Integrative Assessment	8
2.2.4. Norm-referenced Assessment and Criterion-referenced Assessment	8
2.3. Alternative Methods in Language Assessment	8
2.4. Assessment of Language Skills	10
2.4.1. Assessing Reading	10
2.4.2. Assessing Listening	11
2.4.3. Assessing Writing	12
2.4.4. Assessing Speaking	13
2.5. Assessment Literacy	15
2.6. Language Assessment Literacy	16
2.7. Studies on Language Assessment Literacy of EFL Teachers	17
2.7.1. Studies Conducted Abroad	17

2.7.2. Studies Conducted in Turkey	23
3. METHODOLOGY	28
3.1. Research Design.....	28
3.2. Setting and Participants.....	29
3.3. Data Collection Instrument	30
3.3.1. Language Assessment Knowledge Scale (LAKS).....	31
3.3.2. Semi-structured interviews	31
3.4. Data Collection Procedure	32
3.5. Data Analysis	33
4. FINDINGS	36
4.1. Quantitative Data Analysis	36
4.1.1. RQ1: EFL Teachers' General and Skill-based Language Assessment Knowledge Level	36
4.1.2. RQ2: The Relationship among EFL teachers' Skill-based Assessment Knowledge	43
4.1.3. RQ3: Certain Background Variables' Impact on LAK Level of EFL Teachers	44
4.1.4. RQ4: Perceived Self-competency and Actual Language Assessment Knowledge Level	47
4.2. Qualitative Data Analysis	51
4.2.1. RQ5: The Opinions of EFL Teachers Regarding Pre-service Education	52
4.2.2. RQ5: The Opinions of EFL Teachers Regarding Their LAK Level and the Findings of the Scale.....	54
4.2.3. RQ5: EFL Teachers' Needs in Language Testing and Assessment.....	57
5. DISCUSSION	59
5.1. Discussion on the EFL Teachers' General and Skill-based Language Assessment Knowledge Level	59
5.2. Discussion on the Relationship among EFL teachers' Skill-based Assessment Knowledge	61
5.3. Discussion on Certain Background Variables' Impact on LAK Level of EFL Teachers	62
5.4. Discussion on Perceived Self-competency and Actual Language Assessment Knowledge Level	64

5.5. Discussion on The Opinions of EFL Teachers Regarding Pre-service Education and EFL Teachers' Needs in Language Testing and Assessment	65
6. CONCLUSION	67
6.1. Summary of the Study	67
6.2. Practical Implications	68
6.3. Limitations	69
6.4. Recommendations for Further Research	69
REFERENCES	71
APPENDICES	80



LIST OF TABLES

Table 1. Demographic Features and the Number of the Participants	30
Table 2. The Normality Analysis of Research Scales.....	33
Table 3. Statistical Methods Used in Quantitative Analysis.....	34
Table 4. General and Skill-Based Language Assessment Knowledge Level of EFL Teachers	36
Table 5. One Sample T-test Results for the General LAK Level of EFL Teachers ..	37
Table 6. One Sample T-test Results for Skill-Based LAK Level of EFL Teachers ..	37
Table 7. Skill-Based LAK Level of EFL Teachers.....	38
Table 8. The Relationship Among Skill-Based Language Assessment Knowledge .	43
Table 9. LAK based on Years of Experience.....	44
Table 10. LAK based on the BA Program Being Graduated	45
Table 11. LAK based on the Working Place.....	45
Table 12. LAK based on the Grade Level the Participants Teach	45
Table 13. LAK According to the Participants' Attendance to Trainings and Workshops on Testing and Assessment.....	46
Table 14. LAK in Assessing Reading According to the Attendance to Trainings and Workshops	47
Table 15. LAK According to the Experience in Test Preparation/Development and Interpretation	47
Table 16. The Participants' Perceived Self-Competency and the Percentages for Each Variable	48
Table 17. EFL Teachers' Perceived Self-Competency and Their LAK Level in Assessing Reading	48

Table 18. EFL teachers' Perceived Self-Competency and Their LAK Level in Assessing Listening.....	49
Table 19. EFL Teachers' Perceived Self-Competency and Their LAK Level in Assessing Writing	50
Table 20. EFL Teachers' Perceived Self-Competency and Their LAK Level in Assessing Speaking.....	50
Table 21. Themes and Codes of the EFL teachers' Opinions on Language Testing and Assessment	51



LIST OF FIGURES

Figure 1. The Frequency Table of the Participants' Scores Regarding Their General LAK	42
---	----



LIST OF ABBREVIATIONS

EFL	: English as a Foreign Language
LAK	: Language Assessment Knowledge
LAKS	: Language Assessment Knowledge Scale
LAL	: Language Assessment Literacy
LTA	: Language Testing and Assessment

1. INTRODUCTION

1.1. Background of the Study

Just as there is a need for different pieces that complement each other for a puzzle to be a whole, different pieces are needed for the teaching and learning environment to be an effective whole. Assessment, which is there as an indispensable part for an effective teaching and learning environment, is one of those pieces, not only for disciplinary areas such as math, science, literature etc. but also for language. The term testing, which was widely used until a few years ago, has recently undergone a change in usage, and the term assessment has taken its place (Inbar- Lourie, 2008). Brown (2003, p.15) pointed out that the terms “testing” and “assessment” are different from each other by stating that assessment is a continuous process that spans a considerably larger domain in contrast to testing, which is tied to administrative concerns and informs students that they will be assessed. In teaching and learning process, assessment is essential (Lam, 2014). The reason why it is described as indispensable is that it lets teachers think about learning processes critically and get benefited from this evaluation to improve their instruction and students’ learning (Harris, Irving, & Peterson, 2008).

As a part of globalization, people need to communicate with each other no matter in which part of the world they are for some reasons such as work, education, socialization etc. Considering that many people need to use a foreign language rather than their native language in order to fulfil the requirements of globalization, learning a foreign language is a top priority in the 21st century. As communicating in another language effectively is crucial, it requires teachers to take on a greater responsibility, especially in assessment (Purpura, 2016). As a result of an effective assessment process, teachers have a chance to see how effective their instructions are and change them accordingly if it is needed and find out the points students need more help and support them for a better learning process. As teachers have a crucial role in this

process, the value of their professional knowledge in assessment becomes prominent and it makes it necessary for teachers to be an assessment literate.

The term "assessment literacy," first used by Stiggins (1995), refers to understanding the purposes and methods of assessment, being aware of potential assessment issues and how to address them, as well as being knowledgeable about the consequences of making mistakes in assessment. In order to become assessment literate, teachers must have a strong understanding of assessment, both in theory and in practice, as well as the ability to administer, analyze and make decisions on assessments. As the importance of assessment grows in the educational context, these developments have also had an impact on language assessment. According to Purpura (2016, p.191), language assessment refers to a procedure to elicit test and non-test to draw conclusions or make assertions about certain traits of a person related to language. Language assessment literacy (LAL) has been given many different definitions in literature. The most comprehensive definition of LAL was developed by Fulcher (2012, p.125) as the knowledge of a language teacher who is competent enough to develop, analyze and interpret tests, familiar with the assessment process and has awareness of principals including ethics.

1.2. Statement of Problem

The importance of assessment in the teaching and learning process cannot be understated (Lam, 2015, p. 196). One of the most important factors in assessing students is teachers (Vogt & Tsagari, 2014, p. 375). Even though teachers' understanding of assessment is crucial for determining the quality of assessment and testing (Price, Rust, O'Donovan, Handley, & Bryant, 2012), Plake (1993) stated that many teachers don't have the adequate knowledge of it. While instructors are supposed to assess students' progress, many of them still lack the necessary understanding of even the most fundamental key terms (Popham, 2009).

If language teachers lack the necessary proficiency in language assessment, their incompetence will significantly influence the results of the entire assessment process (Ölmezer-Öztürk, 2018). Considering these concerns, EFL teachers' language assessment knowledge (LAK) level has recently started to receive increasing attention (Farhady, 2019, p. 2). However, the findings of a significant number of research conducted worldwide showed a low level of LAK (Jannati, 2015, p. 35; Semiz &

Odabaş, 2016, p. 66; Xu & Brown, 2017, p. 133). It could make it challenging to develop and administer suitable classroom-based assessment assignments to track students' improvement (Doğru, 2020). As Taylor (2009) stated lack of training in language testing and assessment is one of the factors contributing to the low level of LAK. In order to create a more effective teaching and learning environment, it is essential to examine EFL teachers' language assessment knowledge and analyze their needs. However, there isn't much study in the LAK field or in the Turkish context because it's a relatively new area of interest.

Regarding the research studies conducted in Turkey to find out LAK level of EFL teachers, there are some limitations. The very limited sample size is the primary limitation of the research studies (Ballıdağ, 2020; Sarıyıldız, 2018; Tamerer, 2019; Ölmezer-Öztürk, 2018). Even though some studies were conducted with enough participants, they were only carried out either at universities or with pre-service teachers. There are relatively limited studies that have been conducted to investigate LAK level of EFL teachers working at K12 schools in Turkey (Ballıdağ, 2020). Moreover, the studies conducted before are mostly the perceptions studies which are based on participants' opinions on how qualified they are in language testing and assessment. There is definitely a need to search EFL teachers' qualifications based on a performance scale rather than perception. Due to the aforementioned constraints, further research is clearly required to identify their strengths and weaknesses.

1.3. Purpose of the Study and Research Questions

Considering the insufficient data on EFL teachers' language assessment knowledge levels, the purpose of the current study mainly is to explore their general and skill-based LAK level. It also aims to find out whether their level of skill-based language assessment knowledge relate to one another or not. While effects of demographic features are intended to be explored, analyzing teachers' opinions on pre-service education, the possible reasons of findings, and training needs are other important items which are examined in the current study.

In line with these goals, it is aimed to find out the answers for the following questions:

1. What are the general and skill-based Language Assessment Knowledge (LAK) level of EFL teachers working at K12 schools in Turkey?
2. Do their levels of skill-based language assessment knowledge relate to one another?
3. Do the following background variables change the participants' LAK level?
 - a. years of experience,
 - b. the BA program being graduated,
 - c. workplace
 - d. grade level
 - e. experience in testing
 - f. attending trainings and workshops on testing and assessment
4. Does their LAK level change depending on how competent they feel they are in assessing each language skill?
5. What are the participants' opinions on Language Testing and Assessment in relation to following titles?
 - a. Pre-service education
 - b. Their LAK level and the findings of the scale
 - c. Their needs in language testing and assessment

1.4. Significance of the Study

The current study is expected to contribute to the language assessment literacy literature for three main reasons. Firstly, language assessment literacy is considered as a relatively new term and there is a research gap in the field (Fulcher, 2012). Even though there are some studies conducted at universities either with instructors or pre-service teachers, there are very limited studies conducted with EFL teachers working in K12 schools (Ballıdağ, 2020). Secondly, most of the language assessment literacy studies are perception based, and there aren't enough studies exploring the EFL teachers' competence rather than their perceptions (Fulcher, 2012; Inbar-Lourie, 2008; Malone, 2013; Scarino, 2013; Baker & Riches, 2017). There is a strong need for studies focusing on teachers' competence so that they can be supported to be more

assessment literate. That's why, it is very important to explore what teachers know or don't know in language assessment to for a better teaching and learning environment. (Ölmezer-Öztürk & Aydın, 2018). Thirdly, as the participants are interviewed, this study is also expected to provide insights for authorities considering EFL teachers' professional development needs. Due to these reasons, this study is expected to close the gap by focusing on the level of language assessment knowledge of EFL teachers working in K12 schools and contribute to the literature by analyzing their opinions on language assessment literacy regarding various aspects such as pre-service education, findings of the study and their training needs.

1.5. Definitions of Terms

Assessment

It is defined as a range of tasks that teachers employ on a regular basis to track students' development and progress (Coombe et al., 2007, p.8) It provides feedback about areas students need help to improve learning (Fleming, 2007, p.2)

Testing

It is one type of assessment that consists of a collection of questions and is given at regular intervals to check whether our assessment is reliable or not (Douglas, 2014, p. 1).

Assessment Literacy

It relates to understanding assessment principles and using assessment procedures appropriately in addition to knowing the theoretical and philosophical underpinnings of assessing learners' learning (Popham, 2004).

Language Assessment Literacy (LAL)

It is described as being aware of the language related assessments as well as knowing why and how to utilize them (Inbar-Lourie, 2008, p. 390).

2. LITERATURE REVIEW

2.1. Assessment, Evaluation and Testing

Assessment, evaluation and testing play a crucial role in an effective learning and teaching environment. As they help instructors find out the needs of students and change their instructions accordingly, education cannot be taught without these components. However, even though these words are sometimes used interchangeably, they actually mean different things and refer to different concepts.

Evaluation, in a broader sense, is being able to make judgements about all aspects of teaching and learning based on a criterion such as the results of forms of assessment (Genesse, 2001). It goes beyond student achievement and looks at all factors affecting the learning process such as materials, objectives, course design (Harris & McCann, 1994). Brindley (1989) defines evaluation as “conceptualized as broader in scope and concerned with the overall program (p.3)”.

Assessment, in a narrower sense, is a part of evaluation, and it is described as a way of collecting information about student achievement. It is mainly related to students and what they do (Brindley, 1989). Quizzes, presentations, portfolios, tests might be good examples for assessment. Tests are a more specific subcategory of assessment. Bachman (1990) defines it as “a measurement instrument designed to elicit a specific sample of an individual’s behavior” (p. 20).

To sum up, evaluation is an umbrella term including assessment and test; and it is not only about students but also about the whole course. The information is collected from different sources including learners. However, assessment is about learners and their achievement while testing is a specific assessment technique to measure learner’s achievement (Coombe et al., 2007).

2.2. Types of Assessment in Language Education

Assessment tasks to be used in classrooms are carefully chosen according to the purpose of assessment. It is crucial for teachers to know the assessment types as

their knowledge affects their decisions related to education process (Boraie, 2012). There are many different types of assessment, as seen below, depending on the purpose of the assessment.

2.2.1. Formative Assessment and Summative Assessment

In formative assessment, the primary focus is on students' ongoing language improvement. It is more process based. Feedback is the key element in formative assessment (Ramaprasad, 1983). It is aimed to increase students' awareness in their own learning by giving them feedback. Both teachers and students take active part in formative assessment; students are expected to figure out that their purpose is not getting high scores about improving learning (Trumbull & Lash, 2013). However, summative assessment typically occurs at the end of a course and aims to check what students has learned. Contrary to formative assessment, summative assessment is used to analyse student achievement as well as the effectiveness of teachers and institutions (Green, 2018).

2.2.2. Direct Assessment and Indirect Assessment

Direct assessment requires learners to perform the skill intended to be measured. If we want to assess learners' pronunciation skills, we ask them to speak, or if we want to assess their writing composition skills, we ask them to write (Hughes, 1989). According to Hughes (1989), direct assessment has some advantages. Firstly, it is easier to create conditions for judgements as it is clear which skills we want to assess. Secondly, regarding productive skills, the assessment and analysis of it are straightforward. Lastly, as it requires the practice of the related skills, it has a backwash effect. On the other hand, in indirect assessment, "it is aimed to measure the abilities that underlie the skills in which we are interested (Hughes, 1989, p.18)". For example, Lado (1961) suggested an assessment method of pronunciation, which requires learners identify the words that rhyme with one another by using paper and pencil. It can be considered as an indirect assessment method. However, as it isn't always certain that, in indirect assessment, the ability we are assessing is exactly assessing the skill we are interested in, indirect assessment is mostly questioned (Hughes, 1989).

2.2.3. Discrete-point Assessment and Integrative Assessment

Discrete-point assessment assumes that language can be broken into smaller elements and those elements can be assessed separately at a time. According to that view, a language proficiency test must include different skills and as many different linguistic points as possible to be tested (Brown, 2014). However, Oller (1979) stated that language competence must be thought as integrated units which cannot be assessed separately. Hence, this brought integrative assessment into play. According to Hughes (1989), integrative assessment is learners' combining different elements to complete a task. Writing a composition or completing a cloze test are considered as examples of integrative assessment.

2.2.4. Norm-referenced Assessment and Criterion-referenced Assessment

In norm-referenced assessment, learners scores are calculated in relation to a mean, median, standard deviation or percentile rank (Brown, 2003). The scores of test takers are interpreted according to other test takers' scores instead of a predetermined criterion (Richard & Schmidt, 2002). However, in criterion-referenced assessment, test takers' scores are interpreted in relation to predetermined criterion. If the students manage to reach the predetermined criteria, they pass; if they don't, they fail. Hence, in criterion-references assessment, other test takers' scores aren't important to decide on learners' achievement (Richard & Schmidt, 2002).

To finalize, according to the purposes of assessments, the assessment types can be grouped as formative-summative assessment, direct-indirect assessment, discrete point-integrative assessment and norm referenced-criterion referenced assessment.

2.3. Alternative Methods in Language Assessment

A number of alternative assessments have gained popularity in recent years (Brown & Hudson, 1998). The changes in education have caused some changes in assessment methods as well. It was seen that offering just pen and pencil-based tests to assess students' language performance doesn't meet the needs, and they aren't enough to assess each language skill (Lynch, 2001). As alternative assessment is more authentic and needs meaningful contexts, it is seen as a supplement to traditional assessment (Brown, 2003). In alternative assessment methods, content validity is

higher because learners are assessed while performing the target linguistic acts (Brown, 2014). Some of the alternative assessment methods are presented below.

To begin with, Parsons (1998) defined portfolio as “a collection of personal work that a creator selects and displays for the purpose of allowing another person to evaluate the quality of his or her own work” (p.1). It is also presented as students’ organized collections which can show us their development, efforts and achievements (Graves & Sunstein (1992). As it allows students to discover their weaknesses and strengths, they become responsible for their own learning. Hence, using portfolios help instructors create a more student-centered classroom environment rather than teacher-centered one. As a disadvantage, it can be said that a great deal of time is needed for preparation, applying and scoring (Stecher, 1998). Even though it increases interaction between teachers and students, and is considered as beneficial for raising students’ awareness, because of time burden, teachers might not want to apply this method in their classrooms.

Similarly, as a self-reflection tool, journals help learners express their opinions and feelings by writing without any attention on grammatical rules (Brown & Abeywickrama, 2010). Students have a chance of expressing themselves by using their own words in their journals without worrying about being judged. There are different types of journals available such as self-assessment reflections, language-learning logs, response to reading etc. (Brown, 2003).

Self-assessment is another alternative assessment method which let learners make decisions for their own learning and discover their weaknesses and strengths (Andrade & Du, 2007). It requires active student participation and improves critical thinking (Spiller, 2012). Similarly, peer-assessment also make students responsible for their own learning as well as letting students assess others (Caspary & Boothe, 2017).

Observation is an indispensable part of teaching profession. Nevertheless, when it comes to do an assessment through observations, it requires a more systematic procedure. First, the aims are clearly specified beforehand and then the data are collected accordingly. Having a rubric also makes it more systematic and objective (Ballıdağ, 2020).

2.4. Assessment of Language Skills

Language assessment requires teachers to be able to assess four basic language skills which are reading, listening, speaking and writing. When language teachers are aware of methods and techniques which they can use for each skill separately, they can monitor progress better and tailor their instruction accordingly. The detailed information about assessment of these four skills is provided below.

2.4.1. Assessing Reading

Reading is considered as the most necessary skill for all educational context, and it has the highest importance in assessment of language proficiency. Assessment of reading is not only about comprehension of texts. Specific strategies and techniques for full understanding should also be included in assessing learners (Brown, 2003). Even though reading is everywhere in education, learners cannot simply learn necessary skills for reading comprehension without instructions. In order to be effective readers, learners must master some skills and strategies. Firstly, bottom-up strategies for comprehending simple letter, words, phrases, and top-down strategies for focusing on larger parts of a reading text need to be mastered by learners. Secondly, developing a schemata, which is background information, is needed to be able to interpret reading texts effectively (Brown, 2003). Reading skills were divided as macro-skills and micro-skills by Hughes (1989). While micro-skills are about recognizing words and word classes, pattern, rules etc., macro-skills are about recognizing the communicative functions of written texts, inferring context, skimming, scanning activating schemata for text interpretation etc. These skills and strategies are significant for the assessment of reading ability. While thinking about assessing reading related issues such as text genre, specific skills and strategies, it shouldn't be forgotten that the process or a specific product of reading cannot be seen or observed by anyone. For that reason, reading assessment is supposed to be done through inference (Brown & Abeywickrama, 2010).

In an effective assessment of reading skill, the selection of texts is crucial. Harris (1969) stated that the length of texts should be kept brief but also, they should include enough content to be able to ask six-seven comprehension questions. According to him, subject matter should also be taken into consideration. The subject must be carefully chosen according to the purpose of the test, and texts shouldn't

require outside subject-matter knowledge for understanding. Besides, texts shouldn't include universally known knowledge it might cause readers answer questions without focusing on the text. What he meant by the style and treatment of the subject is that tests should include various materials of different types and styles. Lastly, tests shouldn't include very difficult lexical items and sentence structures. However, oversimplification must be avoided because there might be a risk of not discriminating between students of different levels.

Regarding the selection of reading texts, Brown (2003) came up with four types of reading texts which are perceptive, selective, interactive and extensive reading. In perceptive reading, learners are expected to understand discourse components such as punctuation, letters, words etc. Bottom-up strategy is applied here. Some suggested tasks for perception reading are multiple choice, reading aloud, written response and picture-cued items. In selective reading, it is required to recognize lexical, grammatical and discourse features of a language. The tasks which are used for this are multiple-choice, matching, editing tasks, picture-cued tasks and gap-filling. Both bottom-up and top-down strategies are utilized here. In interactive reading, top-down processing is used, and learners are expected to bring their schemata for full understanding and interact with the text. Common reading types used in interactive reading are anecdotes, excerpts from longer texts, announcements, recipes, directions etc. Lastly, extensive reading refers to texts more than one page such as short stories, articles, essays, books that learners read outside classroom. In extensive reading, it is aimed to assess global understanding of a text rather than focusing on little details. That's why, top-down strategies are used.

2.4.2. Assessing Listening

Hughes (1989) argued that it might be weird to assess listening separately from speaking as they are mostly used together in interaction. Even though there are many speaking tests available such as Test of Spoken English, Oral Proficiency Inventory, and Phone Pass, it is very difficult to find one just for listening as it is considered as a component of speaking (Brown, 2003). It was stated that as listening is a receptive skill, the features of assessing listening are in line with assessing reading which is another receptive skill Hughes (1989). However, as it is very complicated process, it is crucial for teachers to know how it works if they want to assess it (Buck, 2001).

For an effective assessment of listening, listening passages should be chosen according to test objectives (Hughes, 1989). If it is intended to see how learners deal with native speakers, authentic listening passages chosen from radio, TV, internet might be used. However, if it is aimed to see whether learners can deal with non-native speakers or not, recordings or specific teaching materials might be used. As it was stated by Hughes (1989), while writing items, it is important to put the items far apart in the passage because test-takers might not catch up second item when they are close to each other. Also, they should be given enough time to familiarize themselves with items before they start. According to Hughes (1989), some suggested techniques that can be used for assessing listening are partial dictation, gap filling, multiple choice, information transfer, note taking and short answer.

Brown (2003) came up with four listening types that allow different tasks to be grouped under. These are intensive, responsive, selective and extensive listening types. Intensive listening is for comprehension of smaller units such as phonemes and words. In responsive listening, learners are required to listen short units of language such as greeting or command in order to give a short response similarly. Selective listening requires learners to listen stretches of discourse such as monologues to look for certain information. The aim isn't looking for global meaning. Lastly, in extensive listening, the aim is to listen for global meaning as different from selective listening. Listening for main idea or making inferences are among extensive listening tasks.

2.4.3. Assessing Writing

While, in the past, daily life was mostly carried out orally, today writing ability has become a significant part of daily life. It has become crucial to assess writing skills in order to monitor students' progress and improve their success. Teachers must decide on their goals and criterion, knowing that listening assessment is not an easy task, and these goals can be assessed through different tasks (Brown, 2003). As similar to other skills, Brown (2003) presented four types of writing, which are imitative, intensive, responsive and extensive. Firstly, in imitative writing, learners master mechanics of a language such as spelling, and in this type, context and meaning come after form. Copying, listening cloze selection tasks, picture-cued tasks, form completion tasks, spelling tests, matching phonetic symbols are among suggested activities. Secondly, in intensive writing, learners are required to produce suitable vocabulary, idioms,

collocations and grammatical structures within a context. Dictation, picture-cued sentence writing, picture description, reordering words in a sentence, limited response writing tasks are some activities of related writing genre. Thirdly, responsive writing requires discourse level production which is creating clearly connected paragraphs. Short reports, descriptions, summaries of articles and brief narratives are among suggested genre of responsive writing. There is a strong emphasis on context along with meaning. Lastly, extensive writing requires writing up to the length of an essay, a long research paper, or maybe a thesis. Writers focus on organizing and developing ideas rather than focusing on form. They write multiple drafts before final product, and focusing on grammatical form is limited to these drafts.

Hughes (1989) stated the points which test writers need to be careful about for eliciting a valid sample of writing ability. He clearly argued that learners should be offered as many different tasks because possible as people's performance change even on the same task. Also, "just testing their writing ability, not something else" was another point which needs careful attention. As Students' creativity or general knowledge aren't teachers' interest, tasks should be carefully designed and they shouldn't measure these things. Lastly, writing tasks should be carefully described and learners should be restricted which means they shouldn't be writing about non-related things and know the limits. Considering the scoring of writing tasks, Hughes (1989) also came up with two different methods which are holistic and analytic. While holistic scoring is about scorer's overall impression of the task, analytic scoring is about assessing different aspects of tasks by assigning some different points. When they are compared, it can be said that holistic scoring is less-time consuming but analytic scoring is more reliable.

2.4.4. Assessing Speaking

Listening and speaking skills are described as closely interrelated skills (Brown, 2003). According to Harris (1969), any other language skills aren't as difficult as speaking to assess, and it is a very complex skills which requires people benefit from some different abilities at the same time. Similarly, O'Sullivan (2012) stated that assessing speaking is the most difficult to administer. Therefore, it is very important for language teachers to be aware of the whole assessing process, criterion to design and administer assessment tasks accordingly.

Types of oral production were divided into five categories by Brown (2003), which are imitative, intensive, responsive, interactive and extensive tasks. To start with imitative tasks, they are solely about phonetic level of oral production. In these activities, test takers basically just imitate a word or phrase. Pronunciation is emphasized, and whether learners can convey meaning, interact with others or not aren't focused on. With intensive type, test-takers are required to produce relatively short stretches of oral language in order to show their competence. Interaction with test administrator is limited in this. Reading aloud, limited picture-cued tasks, sentence level translation, directed response tasks and sentence completion are some of the suggested tasks. Regarding responsive tasks, it can be said that they require interaction and understanding of the test. Short conversations, small talk, simple requests can be listed in this type of oral production. Question and answer tasks, giving instructions and directions, paraphrasing are some suggested tasks for this type. The next type is interactive oral production. The difference between interactive and responsive oral production is that interactive production is longer and more complex than responsive production. The last type is extensive oral production. This type includes oral presentations and storytelling. As it is planned, language is more formal and deliberative.

Harris (1969) classified types of oral tests into three different categories which are paper and pencil objective tests of pronunciation, relatively unstructured interviews and highly structured speech samples. He argued that paper and pencil objective tests are supposed to be only used carefully because the validity of these tests are unproven. Rhyme words, word stress and phrase stress are common items in this group of tests. In highly structured speech samples, it is stated that test-takers are given unequal tasks and assessed because of limited guidelines. It is considered as a weakness of this type or oral test. Sentence repetition, reading passages, sentence conversion, sentence construction, response to pictorial stimuli are among the tasks preferred for highly structured speech samples. Highly structured speech samples are described as the method used most frequently to assess oral production. One or more raters interview test takers and evaluate them according to some criteria. The biggest weakness of this type is described as being subjective which cause raters give totally different scores for different people and the same task, or different scores for even the same person and activity.

To conclude, even though they might seem a bit complicated, it is crucial for teachers to know the assessment process for each skill as it affects language instruction and learning significantly. By learning the criteria for assessing each skill, teachers can design tasks which allow them to evaluate their students' improvement, determine their strengths and weaknesses, and change their instruction accordingly.

2.5. Assessment Literacy

There are many different definitions offered for assessment literacy in the literature. However, the very first definition was proposed in 1990 by the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. They came up with 7 standards needed to define a teacher' assessment literacy which are choosing and developing appropriate assessment methods, administering, scoring, and interpreting the results, using assessment results to make decisions, developing reliable grading procedures, communicating assessment results to students or parents, and being aware of unethical assessment methods (pp.31-32). Webb (2002) also described it as teachers' knowledge on evaluating the development of students and how to benefit from the assessment data to promote learning and improve instruction. For Purpura (2016), it is "teachers' understandings of assessment and assessment processes related to the identification and narrowing of learning gaps in instruction through formative assessment" (p.201). Popham (2009) stated that assessment literacy is a must for teachers rather than being something to be forgotten soon, and he also described assessment literacy as knowing the assessment practices used in classrooms.

In the same vein, according to McMillan (2000), assessment process involves the use of instructors' professional judgment in items such as the creation of test items, scoring, creating rubrics, and analyzing results; and it isn't only about knowing the terminology but also about choosing the best one to promote learning. Similarly, Falsgraf (2005, p.6) described it as "the ability to understand, analyze and apply information on student performance to improve instruction". Thus, it can be concluded that being assessment literate requires theory and practice together.

Xu and Brown (2017) argued that being assessment literate is significant as having insufficient assessment literacy may cause poor judgements and ill-formed decisions for learning and teaching environment. According to Mertler and Campbell

(2005), for an effective assessment, determining the assessment goals in relation to the related program objectives is a prerequisite, and assessment literate teachers may specify these goals and choose appropriate assessment techniques. If this is the case, assessment practices represent students' performance correctly. Stiggins (2002) also argued that teachers must be aware of the goals of the program and inform their students about them beforehand, which requires them to be assessment literate (as cited in Sevilen, 2021, p.23).

To conclude, as it can be seen above, "the definitions of the term assessment literacy abound in the literature" (Coombe, 2012, p.2). It is crucial for teachers to use assessment techniques matching with program objectives, and teachers must be capable of applying those techniques in classrooms rather than just knowing them theoretically. Moreover, it is a consensus among scholars that rather than being a luxury for teachers, it is a skill every teacher must have (Ölmezer-Öztürk, 2018).

2.6. Language Assessment Literacy

Despite the fact that language assessment literacy (LAL) and assessment literacy (AL) have many common components, 'language' is the key construct of LAL (Giraldo, 2018). Fulcher (2012, p.125) describes it as "the knowledge, skills and abilities required to design, develop, maintain or evaluate large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice". According to Inbar-Lourie (2008, p.390), it is about knowing the use of correct language assessment methods and being aware of the reasons of using them. For Purpura (2016), it is a term which is about "a systematic procedure for eliciting test and non-test data in order to make inferences about people's language related characteristics" (p.191). As it can be seen, literature offers different definitions for LAL.

For O'Loughlin (2006: p. 71), LAL is now a crucial skill for language instructors to possess as it helps teachers to improve their self-awareness and discover their perceptions and provides insights about the nature of language assessment (Scarino, 2013: p. 311). However, as teachers aren't born testers (Jin, 2010: p. 556), they need to be provided with sufficient and efficient training about the essentials and basics of LAL (Odo, 2016). Without having sufficient assessment knowledge on

language skills which are reading, writing, speaking, and listening, it is difficult for language teachers to assess language-related abilities effectively (Ölmezer-Öztürk, 2018, p.9). Even though there are much research conducted on assessment, there is an insufficient number of research studies on LAL (Malone, 2013). As language teachers need sufficient language assessment literacy to apply language related assessment practices appropriately, many more research studies are needed regarding LAL to find out their needs and improve their language assessment literacy levels.

2.7. Studies on Language Assessment Literacy of EFL Teachers

Even though the number of the studies regarding the assessment literacy levels of teachers from various subjects is high, studies conducted about EFL teachers' assessment literacy levels are limited in number. This section presents the research studies carried out on language assessment knowledge both abroad and in Turkey. The LAK studies conducted focused on different aspects such as teachers' perceptions in LAL, the level of language assessment literacy of EFL teachers and teacher candidates, perceived further training needs and the quality of assessment courses given at universities or teaching training programs. Regarding those topics, the LAK studies conducted abroad as well as in Turkey are presented below.

2.7.1. Studies Conducted Abroad

The EFL teachers' language assessment knowledge and perceived training needs have been searched by many researchers. As one of those, a comprehensive study was conducted by Hasselgreen, Carlsen, and Helness (2004) to explore the teachers' language testing and assessment literacy levels and more specifically their needs for further training. A total number of 914 participants took part in the study including language teachers, teacher trainers and experts in the field of testing. The questionnaire used in the study were in 3 parts which are classroom-focused assessment, aims of assessment and content and concepts of assessment. Questions in three aspects, which were whether they participated in any type of language testing and assessment, how much training they got in certain areas, and how much further training needed they think, were presented in those 3 parts. Teachers and teacher trainers were asked the same questions while the experts were asked different questions such as item writing, test developing etc. The results showed that, for the

classroom-focused assessment part, teachers carried out some assessment practices in the questionnaire such as “using ready-made tests, giving feedback and using informal, continuous assessment” without any training. However, their perceived training need for using ready-made tests was presented as very little whereas their perceived training need for the items of “preparing tests, interpreting results, giving feedback, using self-assessment” was presented as around 60%. Also, there were mostly “yes” answers for the item “using portfolios” for the perceived training need. Regarding the aims of assessment part, it was stated that it was easier for teachers to apply some assessment practices for the purposes of “awarding certificates” and “placing students” without any training at all compared to “giving grades” and “finding out what needs to be taught”. However, it was found out that the teachers need training in all items. As for the content and concepts of assessment part, “assessing productive skills” was the assessing item that was the least likely to be applied without any training compared to other items in the questionnaire. The significant need for training for all items in the related part was also revealed in the study. All in all, overall results showed that teachers were applying assessment practices mostly without sufficient training and they perceived a need for training in most areas such as using portfolios, preparing and interpreting tests, assessing integrated skills, establishing reliability and validity, writing items, statistics etc.

Similarly, to get a deeper understanding of the teachers’ perceptions on their language assessment literacy and their perceived training needs, Vogt and Tsagari (2014) conducted a study. The data were collected from seven different countries in Europe (Cyprus, Germany, Greece, Italy, Macedonia, Poland, Turkey). Mixed-methods study approach was implemented. The quantitative data were collected from 853 participants by means of a questionnaire which was adapted from The Hasselgreen et al. (2004) questionnaire, and for the qualitative data, 63 participants across Europe were interviewed. No differences were found across countries. Moreover, the results showed that the literacy level of teachers in language testing and assessment is not well-developed no matter which country they are from. It was revealed that the participants received a little or no training in the area. The "purposes of testing" section of the questionnaire, where 42.4% of respondents on average reported receiving no training at all, appears to be the least developed of the three LTA components. Therefore, as the findings from the interview data clearly revealed, key components of

teachers' language testing and assessment literacy, such as "giving grades," "placing students," and "awarding certificates," are undeveloped and most likely picked up on the job. Overall, 33.7% and 34.6% of teachers received either no training or little training, which left them underdeveloped in terms of "classroom-focused assessment" practices such as the teachers' capacity to create and compile tests and other assessment procedures. The teachers' literacy in "content and concepts of assessment" part is similarly underdeveloped. It was shown that the participants either didn't get any training (29.4%) or got only a little training (35.1%) in this aspect. However, this area clearly showed a need for further training, with almost half of the respondents (41.4%) expressing this need. Overall results were consistent with the findings of Hasselgreen et al. (2004)' study and indicated that not sufficient training is given to the pre-service teachers at universities, and they aren't preparing students for their future job. It was also stated that there is definitely a need for practical work opportunities in-service training.

In 2019, Muhammed and Bardakçı also attempted to investigate EFL teachers' assessment literacy working in Iraq. The data were collected through Classroom Assessment Literacy Inventory (CALI) from 101 teachers working in secondary or preparatory schools. The results showed that less than 15 out of 35 questions were correctly answered by Iraqi English language teachers. It was presented that the participants got the lowest score on the item "recognizing unethical, illegal, and inappropriate assessment methods" whereas they got the highest score on the item "using assessment results when making educational decisions". The interesting thing was that even though 77% of the teachers described themselves as well trained for assessment, the results indicated the opposite. Hence, it was recommended that pre-service and in-service trainings be revised.

In similar vein, Tao (2014), tried to explore the assessment literacy development of English language teachers. Hence, he developed four different scales which were Classroom Assessment Knowledge, Innovative Methods, Grading Bias and Quality Procedure. The first scale was a multiple-choice one to explore the assessment knowledge of EFL instructors whereas the others were to explore the participants' assessment related personal beliefs. 108 instructors took part in the study and 6 of them were interviewed. The findings indicated that the participants had poor assessment literacy which affected their implementation in classrooms negatively. The

study provided direct implications for designing assessment policies in higher education settings as well as curricula for pre- and in-service teacher education programs in developing nations.

Another study belongs to Fulcher (2012) who developed a survey to find out the needs of language teachers in language testing and assessment. 278 participants from different countries took part in the study. After a comprehensive analysis, it was revealed that the participants were totally aware of their assessment related training needs. They stated that the areas which they mostly needed training were statistics, reliability and validity. Additionally, it became clear that a balance between large-scale and classroom evaluation was needed (p.116). It was also indicated that assessment practices and its principles should be dealt with by taking into consideration the historical or social context they are in (p.122).

Chan (2016) conducted a study to find out the beliefs of EFL teachers working at elementary schools on their use of multiple assessment. The data were collected from 520 teachers by means of a questionnaire comprising of a self-report Likert scale, multiple choice and open-ended questions. The findings indicated that the participants were aware of what multiple assessment is and what it includes. Additionally, they agreed on the success of multiple assessment, especially the use of portfolio. It was also presented that a big deal of the teachers preferred using formative assessment over traditional assessment such as pen and paper tests. Besides this, the findings showed a meaningful connection between the participants' experiences and their assessment-related beliefs.

In order to learn more about the assessment literacy of 891 English teachers working in China, Xu and Brown (2017) also conducted a study. The data were collected through a questionnaire adapted from Plake et al.'s (1993). The results showed that EFL teachers' levels of assessment literacy were insufficient. It was also demonstrated that demographic characteristics of the instructors, such as age, years of experience, assessment training experience etc., had no meaningful impact on teachers' assessment literacy. A clear need for further training in language assessment was concluded as a result.

In a different study, Jannati (2015) explored the beliefs and practices of English language teachers on assessment by means of semi-structured interviews. In the study,

it was revealed that the participants were aware of what the assessment was and its essential concepts. They were described as assessment literate. However, even though they were familiar with the essentials of assessment, it was stated that they couldn't reflect this in their practices. In other words, the study showed that the participants had difficulty in transferring their knowledge into the practice.

For the purpose of exploring the effectiveness of assessment education and assessment literacy levels of preservice teacher, Volante and Fazio (2007) carried out a study. 69 preservice English language teachers responded to a questionnaire comprising of four aspects such as "self-described level of assessment literacy, main purposes of assessment, utilization of different assessment methods and need for further training, and suggested methods for promoting assessment literacy in university and practice teaching settings" (p.753). Regarding the first aspect, the findings of the study revealed that preservice teachers' level of self-efficacy was relatively low regardless of which year they were in. As for the second aspect, most of the participants stated that they used traditional summative assessment purposes rather than formative assessment. For the third aspect, it was revealed that the participants needed further training in more authentic assessment practices such as portfolio and performance assessment rather than traditional methods. As for the final aspect, it was indicated that courses at universities weren't sufficient and efficient to provide a deeper understanding of language assessment and to prepare students for their future job. Even though they were trained for more traditional approaches, they strongly argued that they were lacking information in some areas such as formative assessment. In order to increase their competency in language assessment, it was indicated that specific courses designed according to needs preparing students for in-class practices for their future job were needed.

Similarly, Deluca and Klinger (2010) looked at teacher candidates' perspectives on the assessment education program in Canada and examined the participants' level of confidence in their research study. The data were collected from 288 preservice teachers through a questionnaire to identify the participants' perceived confidence level in assessment practice, theory and philosophy. In addition, they stated their opinions on the assessment topics needed to be included in an assessment course at university level. It was found out that preservice teachers had higher self-confidence in theory and practice compared to philosophy. As different from Volante and Fazio's

(2007) research findings, in this study it was revealed that the courses at preservice teacher training program had a significant impact on their confidence in assessment practices. The findings of the current study also indicated a need for further training in certain areas such as validity and reliability.

Jin (2010) carried out a study to explore preservice teachers' training in assessment in relation with the assessment courses provided in China. The data were collected from 86 instructors through a survey which was developed to get a deeper understanding of the content of these courses and the students' opinions on how necessary and useful the courses were. It was concluded that the courses were sufficient enough to cover the basics and essential parts of language testing and assessment. However, they weren't effective for bridging the gap between theory and practice. It was stated that classroom practice wasn't getting enough attention in the courses in preservice education. The overall results suggested to focus on the under addressed aspects of assessment courses.

In another study conducted by Lam (2014), it was aimed to investigate the language assessment training in Hong Kong against the reforms in assessment practices in primary and secondary schools. More specifically, it was attempted to search the effects of assessment courses at universities improving or inhibiting the level of assessment literacy. The overall results of the study revealed that assessment courses at the institutions in Hong Kong weren't sufficient and effective in that teachers were unable to transfer their theoretical knowledge into practice. It was also stated that social aspects such as ethics and fairness weren't included in assessment courses. Besides, as he assessment related courses was reported as too academic, the gap between theory and practice widened. For those reasons, it was suggested by Lam (2014) that preservice teachers assessment practices might be included and evaluated during their teaching practicum for developing assessment literacy.

Regarding the assessment practices at tertiary level, in their article, Davin and Heineke (2016) focused on how preservice teachers' learning of classroom-based language assessment was integrated into practice. They argued that utilizing practice-based approach to prepare teacher candidates was a way which enabled teachers to discover authentic ways of using language assessment for supporting student learning. It was also indicated that teacher candidates must be in field for assessment practices besides spending hours in classroom to learn about the theoretical aspects.

2.7.2. Studies Conducted in Turkey

Yetkin (2015) carried out a study to examine the perceptions of preservice English teachers on the use of assessment strategies and the effectiveness of courses at universities for assessment literacy development. 30 preservice English teachers took part in the study and the data were collected through a questionnaire comprising of 5 open-ended and 3 closed-ended questions. 5 of the participants were interviewed in the second phase to support the quantitative findings. Regarding the participants' perceptions on the use of assessment methods, the overall results indicated that they were all aware of the importance of assessment as they described assessment as a tool to observe students' development and their teaching skills, to change their instruction if it was needed and increase the students' awareness. As for the effectiveness of courses at universities, the participants stated that the assessment related courses were beneficial as they prepared preservice teachers for both theory and practice. They also mentioned the effect of their practicum on their assessment practices. The results on their preferred methods of assessment showed that observation technique was the most preferred method while selected response and constructed response were the least favored. Additionally, the participants expressed that they needed further training in observation techniques, personal communication and performance assessment.

In a study conducted in 2010, Hatipoğlu tried to find out the perceptions of trainee teachers on the effectiveness of testing and assessment course offered at Middle East Technical University. The study was carried out with 81 students by means of questionnaires and interviews. In the questionnaire, trainee teachers were supposed to answer two questions. The 1st question was to list 5 things which they covered in the testing and assessment course, and they considered as beneficial for their future practices while the 2nd question was to list the things which must be changed in relation to this course. The findings of the study revealed that the trainee teachers were satisfied with just three sections, which were reliability, validity and testing skills, out of seven. Most of the students stated that other four sections, which were teaching and testing, kinds of tests, kinds of testing and writing multiple-choice items, weren't related to their future career. Besides, the participants mentioned that they didn't have any chance to practice assessment strategies in class or at home as a project. Last but not least, it was also revealed that preservice teachers thought that just one testing and assessment course throughout four years wasn't enough and it didn't give them enough

time to cover all aspects and practice. According to the overall results, Hatipoğlu (2010) suggested relevant changes regarding the implementation of testing and assessment courses.

Öz and Atay (2017) aimed to explore the perceptions of EFL instructors on in-class language assessment, and its impact on practice. In addition, examining the relation between experience and their perceptions was another purpose of this study. 8 female and 4 male instructors participated in the study. Their experience varied from 1 to 15 years. The data were collected through semi-structured interviews. The findings showed that even though the participants were aware of the basics of assessment, they couldn't apply it in classroom. Their assessment literacy and classroom practice weren't matching. It was also revealed that there was no relation between their experience and perceptions on classroom assessment. With regard to these results, it was suggested to include more assessment courses to balance theory and classroom practice.

Cirit (2015) studied the preservice ELT teachers' perceptions towards different assessment methods such as alternative, online and traditional, and examined the changes in their attitudes towards different methods after Web 2.0 tools were included. The study was carried out with 40 preservice ELT teachers in 14 weeks. In this period, 6 tasks with 7 different Web 2.0 tools were implemented. Pre-survey, reflection papers, post-survey and semi-structures interviews were used to collect data. The findings showed that the participants had a positive attitude towards Web 2.0 tools before implementation and it went more positive after. The participants also stated that alternative assessment was more motivating, gave more practical feedback, increased interaction and improved critical thinking skills. Hence, it was revealed that instead of traditional or online assessment, they preferred alternative assessment in general. Results of qualitative and quantitative data were found as consistent with one another.

Yüce (2015) aimed to explore the assessment conceptions of preservice English language teachers as well as finding out the assessment practices the participants would use when they started their professional lives. The data were collected by through an adapted version of a scale named as "Teacher Conceptions of Assessment Scale" which was developed by Brown (2001). It consisted of 27 items in four aspects as improvement, school accountability, student accountability and irrelevance. 133 preservice English teachers from two universities in Konya

participated in the study. The findings showed that some preservice English language teachers considered assessment as a tool to improve the quality of teaching and learning, find out the schools' quality and certify students on their learning while most of them regarded assessment was irrelevant. Additionally, the study's correlation results revealed a substantial association between "Improvement, School Accountability, and Student Accountability" as well as a moderate correlation between these variables and "Irrelevance." It was also stated that preservice English teachers, when they started their professional careers, tended to use alternative assessment methods mostly as different from their own language teachers using traditional methods.

Regarding the training needs and practices of EFL teachers, Mede and Atay (2017) conducted a study with 350 teachers working at universities in Turkey. Mixed-method approach was implemented in order to collect data. The data were gathered by means of the adapted version of Vogt and Tsagari's (2014) questionnaire as well as semi-structured interviews. The overall findings of the study showed that the participants had poor assessment literacy and needed further training in many areas regarding the essentials of assessment. It was also found out that the participants weren't qualified for testing productive and receptive skills while they reported themselves as qualified for testing grammar and vocabulary. The study also indicated that training courses needed to focus on the assessment related classroom practices which would improve assessment literacy level.

In the study of Sarıyıldız (2018), it was aimed to explore preservice EFL teachers' language assessment literacy, perceptions regarding theoretical and practical aspects of assessment courses at universities, and also their perceived training needs. Besides, how preservice EFL teachers evaluated their school practicum course in relation to language testing and assessment course was another purpose of the study. The data were collected from 101 preservice teachers studying at Middle East Technical University by means of a questionnaire. Later, 25 of them were interviewed for the qualitative phase of the study. The results revealed that the participants perceived their assessment training at the university as insufficient and expressed a need for further training. In the qualitative phase, it was also stated that the gap between theory and practice was a lot, and the participants didn't have much chance

to practice assessment methods effectively in their teaching practicum even so they were aware of the importance of assessment courses for their professional careers.

Similarly, Tamerer (2019) studied the training needs and the literacy levels of preservice language teachers at Kocaeli University. Both a questionnaire, which developed by Vogt and Tsagari (2014), and semi-structured interviews were implemented to collect data. A total of 30 preservice EFL teachers participated in the study. After comprehensive analysis of the data, it was revealed that participants didn't have sufficient levels of language assessment literacy and needed further training. They reported that the area they needed training most was using statistic with 54.3% while the area they needed training least in assessing grammar and vocabulary with 40%. Moreover, 70% of the participants stated that they weren't happy with the training they received, and they had wished to get more practical training rather than theory while other 30% expressed that they were happy with their assessment training.

As one of the milestones of studies conducted on language assessment literacy, Ölmezer-Öztürk (2018) developed a "Language Assessment Knowledge Scale (LAKS)" and right after that conducted a study to explore the language assessment literacy levels of EFL instructors working at universities in Turkey for her PhD dissertation. As different from other studies, the researcher used a scale assessing knowledge rather than using perception scales. In LAKS, there were 60 items in four aspects as assessing reading, assessing writing, assessing listening and assessing speaking. The data were collected from 542 EFL instructors working at universities, and 11 of them were interviewed. The LAK level of the instructors was found as 25 out of 60. It was also explored that the instructors got the highest average in assessing reading while they got the lowest average in listening. Additionally, demographic features' effect on LAK level was found as insignificant except for working in a testing office. Along with supporting the findings of quantitative phase, qualitative findings also indicated that trainings in both pre-service and in-service years wasn't efficient and sufficient, and teachers needed further training specifically in each skill.

In conclusion, both abroad and in Turkey, there are some research studies conducted by focusing on language assessment literacy of teachers from various perspectives such as teachers' perceptions in LAL, their perceived competence, training needs, pre-service and in-service EFL teachers' level of LAL or the quality of language assessment courses in pre-service education. However, as the studies

mentioned above indicated, most of them are listed under the perception studies, and there aren't enough research studies focusing on the competence level of EFL teachers (Fulcher, 2012; Iarino, 2013; Baker & Riches, 2017; Ölmezer-Öztürk, 2018). Moreover, most of these studies are conducted at universities either with instructors or pre-service teachers. The number of studies conducted in K12 schools is very limited (Ballıdağ, 2020). In the light national studies, it was revealed that teachers had a low level of language assessment literacy, and pre-service education wasn't sufficient to provide teachers with necessary information and practices on language assessment. As for the needs, it was revealed that they needed more practice in pre-service education rather than theory, and they needed more trainings on formative assessment rather than summative assessment. Similarly, according to international studies, it was explored that EFL teachers' level of LAK was insufficient. They also pointed out the imbalance between theory and practice in pre-service education, and they expressed a need for more practice in-preservice education and more training on alternative assessment methods.

3. METHODOLOGY

In this chapter, the methodological structure of the study is described. It starts with the explanation of the research design. After, the information related to the setting and the participants are given. Then, the data collection instruments are explained along with the data collection procedure. Last but not least, the data analysis procedure and methods used in this study are identified.

3.1. Research Design

In this study, it was aimed to explore the general and skills-based language assessment literacy levels of EFL teachers working at K12 school setting in Turkey as well as their opinions on the findings of the scale, pre-service education and training needs. The effects of background variables such as years of experience, the BA program being graduated, workplace, grade level, experience in testing, and trainings and workshops in testing and assessment were also intended to be explored. In order to investigate them, it was decided to use a mixed-method research design.

Mixed-method research design is the one in which quantitative research design is used for one phase of the study whereas qualitative research design is used for another phase for the purpose of understanding a research problem comprehensively (Creswell, 2005). Rather than being competitive, quantitative and qualitative methods complement each other in a mixed-method research design as the findings of one method are used to expand the findings of the other method (Beck, 2005). Mixed-method research design consists of two categories which are sequential and concurrent (Creswell, Plano, Gutmann & Hanson, 2003). In sequential design, the first type of data is collected and later the second type of data is collected whereas in concurrent design both are collected at the same time. In this study, sequential design was adopted as quantitative data collection preceded qualitative.

In accordance with this research design, the language assessment knowledge scale (LAKS), which was developed by Ölmezer-Öztürk in 2018, was used to collect and evaluate the quantitative data initially regarding the language assessment literacy

of the participants. Even though Dörnyei (2007) states that the results of a qualitative study cannot be generalized because of the limited sample size, qualitative research design helps researchers to reach more in-depth perspectives of participants (Johnson & Christensen, 2004). Due to this, in the second phase, semi-structured interviews were conducted to get more in-depth and personalized data on the findings of quantitative part and explore the participants' opinions and needs on the phenomena investigated. It is claimed that when the interviews are conducted in participants' native language, participants express themselves better and it prevents any problems which might occur due to language barrier and affect the quality of answers (Mackey & Glass, 2005). That's why, the interview questions were offered in the participants' native language, which is Turkish, as well as in English to be able to get more comprehensive answers. Finally, the findings of both parts were analyzed and interpreted for the purpose of answering the research questions.

3.2. Setting and Participants

In the initial phase of the study, the purpose was to collect the quantitative data by means of language assessment knowledge scale (LAKS) developed by Ölmezer-Öztürk (2018). LAKS was shared as a Google form, and the participants responded online along with the required demographic questions and informed consent form. For the participant selection, convenience sampling, which is a type of non-probability sampling, was used. According to Dörnyei and Taguchi (2009), it is the most common sampling method in language studies. It is described as a way of sampling which "occurs when researchers select individuals to study because they are available and convenient, and meet some criteria or characteristics that the investigator seeks to study" (Plano- Clark & Creswell, 2015, p. 236). At the end, 272 EFL teachers working in state and private K12 schools in Turkey participated in the quantitative research phase. The distribution of participants regarding workplace was equal in percentage, which means 136 participants were a teacher in a private K12 school while the other half was in state K12 school. Beside this similarity, there were also some differences among participants regarding their demographic features. The descriptive information of the participants is presented in Table 1 below.

On the other hand, in the qualitative phase of the study, for participant selection, voluntary response sampling was utilized, and the participants stating in

LAKS that they wanted to take part in qualitative phase voluntarily were contacted regarding the semi-structured interviews. Even though the number of participants giving permission to be contacted to take part in interviews was high in number, many of them didn't reply to the e-mail in the first run. As, in this study, it was also intended to describe different perspectives varying according to workplace, equal numbers of participants from both private and state K12 schools were chosen on purpose. In total, 5 EFL teachers working in state and 5 EFL teachers working in private K12 schools took part in one-on-one semi-structured interviews conducted face to face or on Zoom platform according to both participants and researcher's convenience.

Table 1. Demographic Features and the Number of the Participants

Demographic feature	Group	Number of the Participants	Percentage
Gender	Male	71	26,10
	Female	201	73,90
Years of experience	1-5 years	96	35,29
	6-10 years	62	22,79
	11-15 years	45	16,54
	16-20 years	36	13,24
	More than 21 years	33	12,13
The BA program graduated	Non-ELT	67	24,63
	ELT	205	75,37
Educational background	BA	207	76,10
	MA	65	23,90
Workplace	State school	136	50,00
	Private school	136	50,00
Grade level	Primary school	77	28,31
	Middle school	114	41,91
	High school	81	29,78
Attended any trainings on language testing/assessment	No	67	24,63
	Yes	205	75,37
Experience in test preparation/development and interpretation	No	135	49,63
	Yes	137	50,37

3.3. Data Collection Instrument

As the present study has a mixed-method design, both quantitative and qualitative data collection instruments were utilized. The quantitative data was

obtained from language assessment knowledge scale (LAKS) by Ölmezer-Öztürk (2018). Also, to address the related research questions and elaborate the findings of LAKS, semi-structured interviews accompanied the quantitative phase of the study. The detailed information regarding the data collection instruments can be found below.

3.3.1. Language Assessment Knowledge Scale (LAKS)

The questionnaire developed and validated by Ölmezer-Öztürk (2018) was adopted and used in the present study. According to reliability analysis reported to ensure internal consistency, the Cronbach Alpha coefficient of LAKS in total was found as .91 in the original study, which means it has a statistically high reliability. Similarly, in the current study, the Cronbach Alpha coefficient of LAKS was found as .71 which is higher than .60 and reasonably reliable as recommended by Tavşancıl (2006). The questionnaire consists of two parts. The first part includes demographic questions such as gender, years of experience, educational background, the BA program being graduated, workplace, grade level, experience in testing and trainings and workshops in testing and assessment whereas the second part of the questionnaire includes four subscales which are assessing reading, assessing writing, assessing listening, and assessing speaking for the purpose of measuring participants' skill-based assessment knowledge. In the first part, a question about how competent participants see themselves in each skill was also added to be able to compare their perceived competence and actual competence obtained via LAKS. The second part of the questionnaire includes 60 items in total. There are 15 items, for each skill, which are designed to measure participants' skill-based assessment knowledge. Each item has three different options available such as "True", "False" and "Don't know". Participants get "1" point for each correct answer whereas they get "0" points when they give wrong answers or choose "Don't know" option. For this reason, maximum point that can be taken at the end of this study is 60, which means 15 points for each skill. The questionnaire can be seen in Appendix 1.

3.3.2. Semi-structured interviews

Semi-structured interviews are the most common types of interviewing in social sciences, and moreover, they are sometimes the only type included in textbook related to qualitative research (Flick, 2002). As different from structured interviews, semi-structured interviews let interviewers use knowledge-producing potentials of

dialogues better as they are more visible in process by directing ongoing interviewing process into the parts seemed as more important and being more flexible in asking questions (Brinkmann & Kvale, 2005).

Within the scope of the present study, as it was aimed to get more detailed data regarding participants' opinions on the quantitative findings of the questionnaire, pre-service education and their perceived needs in language testing and assessment; semi-structured interviews which consist of six main questions and many follow up questions were conducted. While the first question with its follow up questions was about participants' in-dept thoughts on pre-service education, the next four questions were about the participants' understandings of the findings of LAKS. Finally, the last question was about the participants' perceived assessment related needs (see Appendix 2). In total, 10 participants were interviewed and duration for each of them was between 20-30 minutes.

3.4. Data Collection Procedure

The quantitative data of the present study were collected from the end of October till the last day of December in 2022-2023 academic year. As the first step, LAKS was created as a google form document to share it easily and reach many participants from different cities of the country. The consent form for voluntary participation was also included as the first page. Then, the questionnaire was shared online with many colleagues on professional development groups and social media. It was also directly emailed to many teachers through the use of professional platforms such as LinkedIn. As participation was very less on some days, the reminders were shared regularly until reaching out target number of participants. The quantitative data collection procedure lasted two months and at the end 272 participants took part in the study.

For the qualitative data, first the participants who gave consent to be contacted for interviews, were listed and they were sent an email again to check whether they still wanted to be interviewed or not. However, although 108 participants ticked the box in the questionnaire accepting to be interviewed, just a few of them responded to the email. Reminder emails were sent again until finding enough participants which was decided as 10 teachers comprising of 5 EFL teachers working in state K12 schools and 5 EFL teachers working in private K12 schools. As soon as they confirmed their

participation in semi-structured interviews, an appointment was created for each participant according to both the participants and the researchers' convenience. While two participants were interviewed face to face as they asked for it, the other eight participants were interviewed online on Zoom. Before starting, they were informed about consent form and guaranteed to keep their names confidential. The interviews were conducted in Turkish to create a stress-free environment and let the participants express themselves in a better and detailed way. The interviews were recorded in audio format and transcribed by the researcher to be analyzed in later stages. Also, they were translated into English for the data analysis procedure.

3.5. Data Analysis

As the present study utilizes mixed method research design to seek answers for the research questions, different methods were used to analyze and interpret the data in each step.

The quantitative data in the first step were analyzed by means of SPSS program version 26. Firstly, normality tests were conducted to see whether parametric or non-parametric tests will be used. For this purpose, Skewness and Kurtosis values of the research scale were examined. The values for each subscale were found to be as below:

Table 2. The Normality Analysis of Research Scales

	Skewness	Kurtosis
Reading	-0,57	1,30
Listening	-0,13	0,12
Writing	-0,52	0,52
Speaking	-0,77	1,07
Total	-0,98	1,52

George and Mallery (2012) stated that a value of Skewness and Kurtosis between ± 2.0 demonstrates a normal distribution. As it can be seen in Table 2, all values for each subscale are between ± 1.5 . As Skewness and Kurtosis values confirmed the normality, and parametric tests were decided to be utilized in the data analysis. The significance level of 0.05 was used as a criterion in interpreting whether the obtained values were significant or not.

For the first research question, the participants' general and skill-based language assessment knowledge were analyzed through descriptive statistics focusing

on the mean scores rather than total scores as cut-off points were provided in the previous studies as “1” for each correct answer and “0” for the other options. One-sample t-test was used to see whether there is a statistically significant difference between the participants’ mean scores and the reference point. For the second research question, through Pearson correlation, the relationship of the participants’ skill-based knowledge was examined and it was explored whether the skills and total LAK level have a positive or negative correlation. For the third research question, the effects of demographic features on participants’ assessment knowledge level were analyzed through inferential statistics which are independent-samples T-test and one-way ANOVA. Independent-samples T-test was used with the demographic features which include two independent groups such as the BA program being graduated, workplace, the attendance to trainings, the experience in test preparation/ development and interpretation whereas one-way ANOVA was used to analyze the demographic features which include three or more independent groups such as years of experience and grade level. With regard to fifth research question, for the analysis of the participants’ perceived self-competency and LAK level, one-way ANOVA test was run. The statistical methods used in analysis are presented in Table 3 below.

Table 3. Statistical Methods Used in Quantitative Analysis

The research questions	The statistical method
RQ 1: What are the general and skill-based Language Assessment Knowledge (LAK) level of EFL teachers working at K12 schools in Turkey?	Descriptive statistics (mean, percentage, standard deviation, etc.) One sample t-test
RQ 2: Do their levels of skill-based language assessment knowledge relate to one another?	Pearson Correlation
RQ 3: Do the following background variables change the participants’ LAK level?	Inferential statistics (Independent samples T-test, one-way ANOVA)
RQ 4: Does their LAK level change depending on how competent they feel they are in assessing each language skill?	One-way ANOVA

The qualitative data in the second step were analyzed through content analysis. For the fifth research question, Creswell's (2008) bottom-up approach was utilized to the process of qualitative data analysis. After collecting the data, the first thing the researcher did was to prepare the data for analysis by transcribing and translating into English. The printed copies of transcripts were prepared, and the data analyzed by hand. After a few times of reading the data, the researcher explored the data to get the general sense of it. Then, the data was divided into segments, and the segments were labelled with codes. Then, the codes were classified into the pre-determined themes. To prevent subjectivity, the data was also analyzed by another researcher holding an MA degree in ELT. After the re-evaluation of the data by the second coder, some codes were revised and they were grouped under the pre-determined themes in a more organized way. Last but not least, the results were presented with supporting quotes of the participants taking part in the interviews.

4. FINDINGS

The following chapter presents the results of both quantitative and qualitative analyses conducted by addressing related questions. It starts with the information on the participants' general and skill-based language assessment literacy. Then, the results of the second research question, which tries to explore whether their levels of skill-based language assessment knowledge relate to one another or not, are presented. Thirdly, the effects of demographic features are explained one by one. Next, the results on the participants' perceived competency and their language assessment knowledge are demonstrated. Lastly, the qualitative findings, which are about the participants' opinions on pre-service education, their LAK level and the findings of the scale and their needs in language testing and assessment, are presented.

4.1. Quantitative Data Analysis

4.1.1. RQ1: EFL Teachers' General and Skill-based Language Assessment Knowledge Level

In the first research question, it was aimed to explore the general and skill-based language assessment knowledge level of EFL teachers working at K12 schools in Turkey. The results obtained from 272 participants are presented below in Table 4.

Table 4. General and Skill-Based Language Assessment Knowledge Level of EFL Teachers

N = 272	M	SD
Reading	9.58	2.225
Listening	6.82	2.087
Writing	6.88	2.288
Speaking	8.31	2.600
TOTAL	31.59	6.366

The responses from participants indicated that the mean score of EFL teachers' general LAK level was 31.59 out of 60. The mean scores for each skill were found to be as follow: 9.58 in assessing reading, 8.31 in assessing speaking, 6.88 for assessing writing and 6.82 for assessing listening over 15. Assessing reading was ranked as the area the participants are the most knowledgeable in while assessing listening was ranked as the area the participants are the least knowledgeable in. As the overall point which can be taken from the whole scale is 60, the reference point for 50% success rate was identified as 30. According to the results, it can be said that the mean score of EFL teachers' general LAK level, which is 31.59, is slightly higher than the reference point, which shows average success. In order to confirm this hypothesis, one-sample t-test was applied. The results are presented in Table 5.

Table 5. One Sample T-test Results for the General LAK Level of EFL Teachers

Mean difference	df	t	P
1,59	271,00	4,12	0,00*

*p<.05

As a result, it was revealed that the mean difference between the participants mean score ($M = 31.59$, $SD = 6.36$, $t(271) = 4.12$, $p < .05$) and the score which shows 50% success rate (30 out of 60) was statistically significant. However, to see how significant it is, Cohen's d effect size was determined. The effect size, as measured by Cohen's d, was $d = 0.25$, indicating a small effect (Cohen, 1988). That can be interpreted as EFL teachers' LAK level in general is only slightly higher than 50% success rate. One-sample t-test was also run for each skill to see whether the difference between the participants' mean score for each skill and the score which shows average success (7,15 out of 15) is significant or not. Moreover, Cohen's d effect size for each skill was also calculated.

Table 6. One Sample T-test Results for Skill-Based LAK Level of EFL Teachers

	Mean difference	df	t	P
Reading	2,08	271,00	15,42	0,00*
Listening	-0,68	271,00	-5,38	0,00*
Writing	-0,62	271,00	4,45	0,00*
Speaking	0,81	271,00	5,13	0,00*

*p<.05

The results in Table 6 indicated that there was a significant difference between the participants' mean score in the scale for each skill and the half of the maximum score (7,5). The participants' mean scores in reading ($M = 9.58$, $SD = 2.225$, $t(271) = 15.42$, $p < .05$) and speaking ($M = 8.31$, $SD = 2.600$, $t(271) = 5.13$, $p < .05$) are higher than the half of the total score whereas the participants' mean scores in listening ($M = 6.82$, $SD = 2.087$, $t(271) = -5.38$, $p < .05$) and writing ($M = 6.88$, $SD = 2.288$, $t(171) = 4.45$, $p < .05$) are lower than the half of the total score. The effect size for reading, as measured by Cohen's d , was $d=0.93$, indicating a large effect whereas the effect size for listening ($d=0.32$), writing ($d=0.27$) and speaking ($d=0.31$) indicated a medium effect (Cohen, 1988). From these results, it can be concluded that the participants' LAK level in assessing reading is significantly higher than half of the total score with a large effect size, and their LAK level in assessing speaking is slightly higher than half of the total score with a medium effect size. Also, their LAK level in both assessing listening and writing is slightly lower than half of the total score with a medium effect size.

In addition to general interpretation of the findings, the participants' skill-based LAK levels were also investigated in a detailed way. The results regarding each skill are presented below in Table 7.

Table 7. Skill-Based LAK Level of EFL Teachers

ITEMS	N	True	False	Don't know	Mean	SD
ASSESSING READING	(The participants with correct answers are shown in bold)					
1. Asking learners to summarize the reading text is a way of assessing their reading skills.	272	243	22	7	.89	.309
2. When asking several questions about a reading text, all the questions are independent of each other.	272	92	167	13	.34	.474
3. Cloze test is used for assessing the main idea of the text.	272	88	151	33	.56	.498
4. In a reading exam, using a text learners have encountered before is not a problem.	272	120	109	43	.40	.491
5. One reading text is enough to be included in a reading exam.	272	87	156	29	.57	.495
6. The language of the questions is simpler than the text itself.	272	170	67	35	.62	.485
7. Errors of spelling are penalized while scoring.	272	100	141	31	.52	.501
8. Taking vocabulary difficulty into	272	222	40	10	.82	.388

consideration is necessary in assessing reading skills.

9. Including not stated/doesn't say along with true/false items has advantages over true/false items.	272	182	42	48	.67	.471
10. The more items a reading text is followed, the more reliable it becomes.	272	170	53	49	.62	.485
11. Using the same words in the correct option as in the text is not a problem.	272	71	181	20	.67	.473
12. Simplification of reading texts is avoided.	272	75	162	35	.60	.492
13. Reading texts in a reading exam include various genres (essay, article, etc.).	272	244	20	8	.90	.304
14. In top-down approach, assessment is on overall comprehension of the reading text.	272	194	26	52	.71	.453
15. Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.	272	189	57	26	.69	.461

READING-TOTAL	272				9.58	2.225
----------------------	------------	--	--	--	-------------	--------------

ASSESSING LISTENING

16. Using reading texts for listening purposes poses a problem.	272	115	120	37	.42	.495
17. Including redundancy (e.g. what I mean to say is that....) in a listening text poses a problem.	272	78	159	35	.58	.494
18. Any type of listening text is used for note-taking.	272	129	116	27	.43	.495
19. Spelling errors are ignored in scoring the dictation.	272	100	155	17	.37	.483
20. Errors of grammar or spelling are penalized while scoring.	272	144	101	27	.37	.484
21. A listening cloze test is a way of selective listening.	272	226	24	22	.83	.376
22. Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative testing.	272	148	37	87	.14	.343
23. Scoring in note-taking is straightforward.	272	136	63	73	.23	.423
24. In discrete-point testing, comprehension is at the literal/local level.	272	154	30	88	.57	.497
25. Using dictation diagnostically in assessing listening skills does not pose a problem.	272	106	99	67	.39	.489
26. Giving learners a transcript of the listening text is a valid way of assessing listening skills.	272	72	178	22	.65	.476
27. Dictation is a kind of discrete-point testing.	272	143	46	83	.17	.376
28. Inference questions based on intelligence are avoided in listening tests.	272	140	85	47	.51	.501
29. Asking learners to listen to names or numbers is called intensive listening.	272	144	78	50	.29	.453
30. In selective listening, learners are	272	236	20	16	.87	.339

expected to look for certain information.

LISTENING-TOTAL	272				6.82	2.087
ASSESSING WRITING						
31. Giving two options to learners and asking them to write about one ensure reliable and valid scoring.	272	198	44	30	.16	.369
32. Analytic scoring is used to see the strengths and weaknesses of learners.	272	215	18	39	.79	.408
33. The parts of a scoring scale and the scores in each part do not change for different levels of learners.	272	88	154	30	.57	.497
34. When there is a disagreement between the scores of the two raters, they score the written work again.	272	194	33	45	.12	.327
35. Learners are required to write about at least two tasks in the exam rather than one task.	272	104	113	55	.38	.487
36. Giving restrictive prompts/guidelines to learners for the writing task is avoided.	272	111	128	33	.47	.500
37. Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.	272	209	41	22	.15	.358
38. Using visuals which guide learners for writing poses a problem.	272	58	185	29	.68	.467
39. Holistic scoring is used to see whether the learner is proficient or not at the end of the term.	272	155	37	80	.57	.496
40. Analytic scoring leads to greater reliability than holistic scoring in writing.	272	164	29	79	.60	.490
41. In controlled writing, learners have the chance to convey new information.	272	123	99	50	.36	.482
42. Classroom evaluation of learning in terms of writing is best served through analytic scoring rather than holistic scoring.	272	123	59	90	.45	.499
43. Irrelevant ideas are ignored in the assessment of initial stages of a written work in process writing.	272	120	108	44	.40	.490
44. Providing a reading text for writing is a way of assessing writing skills.	272	141	88	43	.52	.501
45. Mechanical errors (e.g. spelling and punctuation) are dealt with in the assessment of later stages of a written work.	272	178	60	34	.65	.476
WRITING-TOTAL	272				6.88	2.288
ASSESSING SPEAKING						
46. When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.	272	158	94	20	.35	.476
47. Giving learners one task is enough to assess speaking skills.	272	50	196	26	.72	.450
48. Interlocutors' showing interest by verbal and non-verbal signals poses a problem.	272	78	162	32	.60	.492
49. When it becomes apparent that the	272	100	124	48	.37	.483

learner cannot reach the criterion level, the task is ended.						
50. Using holistic and analytic scales at the same time poses a problem.	272	84	98	90	.36	.481
51. Reading aloud is a technique used to assess speaking skills.	272	117	133	22	.43	.496
52. In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.	272	199	32	41	.73	.444
53. In interactive tasks, more than two learners pose a problem.	272	91	142	39	.33	.473
54. The interlocutor gives the score when the learner is in the exam room.	272	96	134	42	.49	.501
55. In a speaking exam, production and comprehension are assessed together.	272	212	30	30	.78	.415
56. Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills.	272	124	119	29	.46	.499
57. Discussion among learners is a way of assessing speaking skills.	272	240	14	18	.88	.323
58. A checklist is a means of scoring oral presentations in in-class assessment.	272	211	24	37	.78	.418
59. When the focus is to assess discourse, role plays are used.	272	199	27	46	.73	.444
60. In peer interaction, random matching is avoided.	272	83	153	36	.31	.461
SPEAKING-TOTAL	272				8.31	2.600
LAKS-TOTAL	272				31.59	6.366

The results revealed that the participants had the highest mean score in reading (9,58 out of 15). Among 15 items, “Reading texts in a reading exam include various genres (essay, article, etc.) (TRUE)” was the one which the participants mostly answered correctly whereas “When asking several questions about a reading text, all the questions are independent of each other (TRUE)” was the least correctly answered. Even though, the mean score is the highest in assessing reading skills with the average 9,58, the participants still have knowledge gap in assessing this skill as the overall score is 15. Regarding assessing listening, the item stating that “In selective listening, learners are expected to look for certain information (TRUE)” had the highest mean score. However, “Phonemic discrimination tasks (e.g., minimal pairs such as sheep-ship) are examples of integrative testing (FALSE)” was the item which had the lowest mean score. Considering the results in assessing listening, 9 items over 15 were answered incorrectly by most of the participants. Assessing listening was found as the least knowledgeable skill with the average of 6,82. The items between 31 and 45 were for exploring assessing writing skills. It was revealed that “Analytic scoring is used to

see the strengths and weaknesses of learners (TRUE)” was answered correctly by most of the participants and had the highest mean score while “When there is a disagreement between the scores of the two raters, they score the written work again (FALSE).” had the lowest mean score. The last 15 items were about assessing speaking. “Discussion among learners is a way of assessing speaking skills (TRUE)” was identified as the item which was answered correctly by most of the participants whereas “In peer interaction, random matching is avoided (TRUE)” was the item which was answered incorrectly with the lowest mean score.

In addition, the participants’ range of frequency based on their mean score is presented in Figure 1 below.

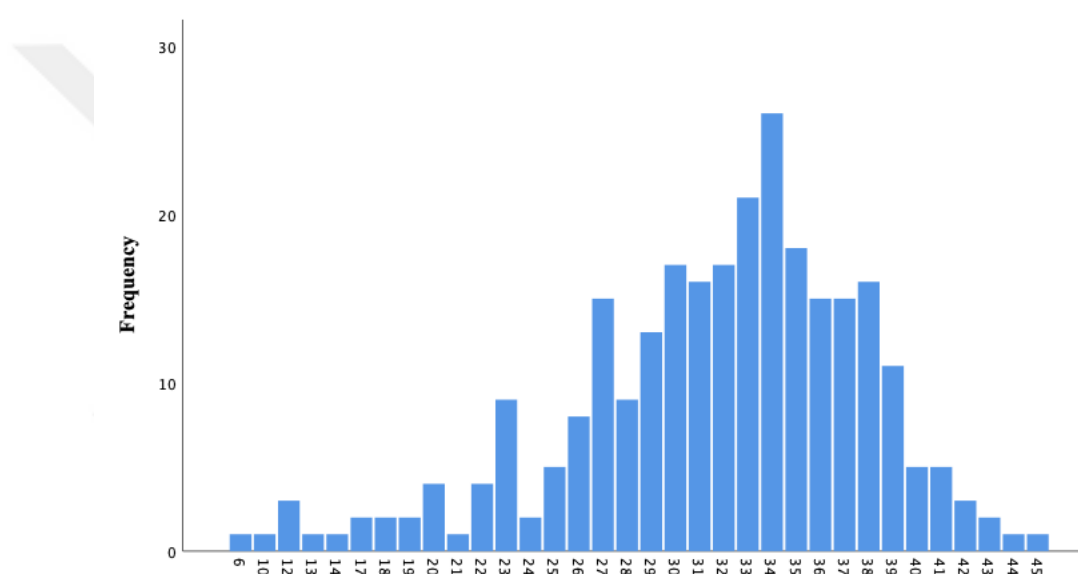


Figure 1. The Frequency Table of the Participants’ Scores Regarding Their General LAK

As it can be seen in Figure 1 above, the lowest LAK level in general is 6 and the highest LAK level is 45. There is no one who shows full success in the scale. Moreover, the number of the participants who got 45 or over, which shows 75% success, is only 1. The participants who scored 30 or over, as a sign of 50% success, are 188 EFL teachers over 272. The ones who scored less than 30 aren’t less in number. 83 EFL teachers had a poor performance in the scale with the scores less than 30. Among them, 7 participants even scored less than 15. As it can be seen, even though most of the participants got average score, there aren’t many participants who can be considered as high achievers. To conclude, it can be said that the participants still have

to develop their assessment knowledge even in the skills they are described as knowledgeable considering the mean scores of each skill.

4.1.2. RQ2: The Relationship among EFL teachers' Skill-based Assessment Knowledge

To investigate the relationship among the participants' skill-based assessment knowledge, Pearson correlation was employed. The findings are presented in Table 8.

Table 8. The Relationship Among Skill-Based Language Assessment Knowledge

	LAK	Reading	Listening	Writing	Speaking
LAK	1,00	.628**	.645**	.762**	.723**
Reading		1,00	.259**	.301**	.209**
Listening			1,00	.350**	.247**
Writing				1,00	.447**
Speaking					1,00

**Correlation is significant at the 0.01 level (2-tailed), N=272

The correlation degrees are defined as follows: $\pm 1 \leq r \leq \pm 0.7$ indicates a strong relationship, $\pm 0.7 \leq r \leq \pm 0.3$ indicates a moderate relationship, and $\pm 0.3 \leq r \leq \pm 0.1$ indicates a weak relationship (Newbold, 2000). According to the findings above, it can be said that all correlational values among the variables are significant. It was found that assessing writing and assessing speaking were highly and positively correlated with language assessment knowledge (LAK) in general whereas assessing reading and assessing listening were moderately and positively correlated. It was discovered that there were significant positive correlations - ranging from high to low - between all types of skill-based knowledge. Among them, the highest correlational level was found between writing and speaking (.447), whereas the lowest was between reading and speaking (.209). It can be inferred that productive skills, which are writing and speaking, have more common features compared to reading and speaking skills which appear to be less interrelated. All in all, as all correlations are positive, an improvement in one of the skills causes an increase in other skills. They might be all considered as interrelated elements.

4.1.3. RQ3: Certain Background Variables' Impact on LAK Level of EFL Teachers

Whether the demographic features such as years of experience, the BA program being graduated (ELT or non-ELT), working place (at a private or state school), grade level the participants teach, attending trainings and workshops on testing and assessment, experience in test preparation/development and interpretation affect LAK level of EFL teachers or not was aimed to be explored with the third research question. The findings can be examined in the tables below.

Table 9. LAK based on Years of Experience

Years of experience	N	M
1-5 years	96	32,08
6-10 years	62	31,45
11-15 years	45	30,67
16-20 years	36	32,47
more than 21 years	33	30,73

	Sum of Squares	df	Mean Square	F	p
Between Groups	115.50	4	28.874	.709	.586
Within Groups	10868.21	267	40.705		
Total	10983.70	271			

Firstly, one-way ANOVA was run to be able to investigate the impact of years of experience on LAK level of the participants. As it can be seen in Table 9, the number of participants decreased in more experienced groups respectively. The highest mean score of LAK ($M=32.08$, $SD=6.607$) was obtained by the participants who were 1-5 years experienced. However, the lowest mean score of LAK ($M=30.73$, $SD=6.625$) was obtained by the participants who were experienced more than 21 years. When the mean scores for each group were examined, no significant difference was found among the groups as a result of one-way ANOVA analysis ($F(4,267)=.709$, $p=.586$). In light of this, it can be inferred that years of experience didn't have an important impact on LAK levels of the participants.

Table 10. LAK based on the BA Program Being Graduated

BA Graduation	N	Mean	Std. Deviation	Std. Error Mean
English Language Teaching	205	31.51	6.568	.459
Non-ELT	67	31.84	5.746	.702
Mean diff.	df	t	p	
.324	270	.361	.72	

Secondly, to see the effects of the BA program being graduated, independent sample t-test was applied. The number of the participants who were ELT graduates was 205 whereas 67 of them were non-ELT graduates. The results showed that there was no significant difference between ELT and non-ELT graduates in terms of their LAK level.

Table 11. LAK based on the Working Place

Workplace	N	Mean	Std. Deviation	Std. Error Mean
State School	136	31.38	7.321	.628
Private School	136	31.81	5.260	.451
Mean diff.	df	t	p	
-.434	270	-.561	.58	

It was also intended to see the effects of the workplace on LAK levels of the participants. For that reason, independent sample t-test was utilized. The number of participants were equal for each group. The mean score of the participants working in state K12 schools was 31,38 whereas it was 31,81 for the participants working at private K12 schools. Based on the data above, it was revealed that there was not a significant difference between these two groups. In other words, whether the participants work in a private or state K12 school has no effect on their language assessment knowledge.

Table 12. LAK based on the Grade Level the Participants Teach

Grade level	N	M
Primary School	77	30,69
Middle School	114	31,41
High School	81	32,70

	Sum of Squares	df	Mean Square	F	p
Between Groups	166.671	2	83.336	2.072	.128
Within Groups	10817.031	269	40.212		
Total	10983.702	271			

To see the impact of the grade level being taught on LAK, one-way ANOVA was applied. The number of the participants working in primary school was 77 with the mean score 30.69, middle school was 117 with the mean score of 31,41 whereas the number of the participants working in high school was 81 with the mean score of 32,70. The highest mean score was obtained from high school group while the lowest mean score was obtained from primary school group. However, as a result of one-way ANOVA test, it was explored that there was no significant difference among the groups. Given this information, the grade level being taught has no significant effect on the participants' LAK level.

Table 13. LAK According to the Participants' Attendance to Trainings and Workshops on Testing and Assessment

Attending any trainings in LTA	N	Mean	Std. Deviation	Std. Error Mean
Yes	205	31.52	6.274	.438
No	67	31.81	6.684	.817
Mean diff.	df	t	p	
.284	270	.316	.75	

Independent sample t-test was run to explore the effects of attendance to trainings and workshops on testing and assessment. 205 EFL teachers stated that they attended trainings in LTA whereas 67 of them stated that they didn't attend to any trainings or workshops. The mean score was found as 31.52 for the first group while the mean score was found as 31,81 for the second group. As it can be seen from the table above, the results indicated that there was no significant difference between these two groups in terms of their LAK. However, to see whether attendance to trainings and workshops affects skill-based LAK or not, independent sample t-test was applied for each skill. The only significant difference was found in assessing reading. In other words, the attendance to trainings and workshops played a significant role in participants' LAK in assessing reading. The results can be seen in Table 14 below.

Table 14. LAK in Assessing Reading According to the Attendance to Trainings and Workshops

Attending any trainings in LTA	N	Mean	Std. Deviation	Std. Error Mean
Yes	205	9.73	2.091	.146
No	67	9.12	2.556	.312
Mean diff.	df	t	p	
.612	270	1.965	.05	

Lastly, in the third research question, it was aimed to explore the effects of experience in test preparation, development, and interpretation on the participants' LAK level. As it can be seen in Table 15, among all participants, 137 EFL teachers stated that they were experienced in test preparation and interpretation whereas other 135 stated that they weren't experienced. The mean score for the first group was found as 32.24, and it was found as 30.93 for the second group. As a result of the independent sample t-test, it was revealed that there wasn't a significant difference between these two groups, and experience in test preparation/development and interpretation didn't play a significant role in the participants' LAK level.

Table 15. LAK According to the Experience in Test Preparation/Development and Interpretation

Experience in LTA	N	Mean	Std. Deviation	Std. Error Mean
Yes	137	32.24	5.938	.507
No	135	30.93	6.732	.579
Mean diff.	df	t	p	
1.308	270	1.699	.90	

4.1.4. RQ4: Perceived Self-competency and Actual Language Assessment Knowledge Level

With the fourth research question, it was aimed to explore whether their LAK level change or not according to how competent they feel for each skill. For that reason, firstly, they were asked to evaluate themselves as an assessor in reading, writing, listening, and speaking. They had four options for each skill as very competent (1), competent (2), not very competent (3) and not competent (4). Then, one-way ANOVA

test was conducted to check whether there is a significant difference among different perception groups or not. The detailed information on participants' perceived self-competency and the percentages for each variable are presented in Table 16.

Table 16. The Participants' Perceived Self-Competency and the Percentages for Each Variable

Perceived Self-Competency	Mean	Very Competent %	Competent %	Not Very Competent %	Not Competent %
Reading	1,48	56,3	39,3	4,4	0
Listening	1,67	41,5	50	8,5	0
Writing	1,72	39,3	48,9	11,8	0
Speaking	1,77	36	51,1	12,9	0
TOTAL	1,66	43,275	47,325	9,4	0

As it can be seen in Table 16, there are no participants describing themselves as “not competent” for any skills. Regarding general LAK, it was found out that more than half of the participants (90,6%) stated that they were “very competent” or “competent” in each skill whereas just 9,4% of them stated that they were “not very competent.” It was also the same for each skill separately. In other words, for each skill, most of the participants' perceived self-competency was either “very competent” or “competent.” The skill which the participants thought as they were most competent in was found as assessing reading ($M=1.48$, $SD=0.582$), and least confident in as assessing speaking ($M=1.77$, $SD=0.661$).

The results which were derived from one-way ANOVA analysis for each skill are presented below for each skill. Also, descriptive analysis was also performed to get a detailed understanding of the data and demonstrated for each skill in the following tables.

Table 17. EFL Teachers' Perceived Self-Competency and Their LAK Level in Assessing Reading

Assessing Reading	N	M
very competent	153	9.80
competent	107	9.41
not very competent	12	8.33

	Sum of Squares	df	Mean Square	F	p
Between Groups	28.928	2	14.464	2.963	.053
Within Groups	1312.292	269	4.882		
Total	1342.221	271			

In Table 17, the descriptive analysis demonstrated that 153 EFL teachers evaluated themselves as “very competent” in assessing reading, and their mean score was 9.80 (SD=2.178) whereas 12 of the participants described themselves as “not very competent” and their mean score was 8,33 (SD=3.114). There was no one saying that they were “not competent” in assessing reading. Even though it isn’t a perfect score to describe them as high achievers, the highest mean score was obtained from the group describing themselves as “very competent”. On the other hand, the lowest mean score was obtained from the group describing themselves as “not very competent.” The mean scores for each group were close to each other. According to the results of one-way ANOVA test, there was no significant difference at the $p < .05$ level among EFL teachers who evaluated themselves as very competent, competent, and not very competent in assessing reading ($F(2,269) = 2.963$, $p = .053$).

Table 18. EFL teachers’ Perceived Self-Competency and Their LAK Level in Assessing Listening

Assessing Listening	N	M
very competent	113	7.21
competent	136	6.70
not very competent	23	5.61

	Sum of Squares	df	Mean Square	F	p
Between Groups	53.152	2	26.576	6.343	.002
Within Groups	1127.021	269	4.190		
Total	1180.173	271			

As a result of one-way ANOVA test, it was revealed that there was a significant difference among perception groups at the $p < .05$ level in terms of their LAK level in assessing listening ($F(2,269) = 6.343$, $p = .002$). The results suggests that the differences among the three different perception groups' mean scores in assessing listening are unlikely to be due to random variation. The perception groups significantly differ in assessing listening. The descriptive analysis findings demonstrated that in assessing listening, most of the participants perceived themselves as “very competent” or “competent”. Only 23 of the participants stated that they were “not competent”, and

their general LAK was 5.61 (SD=2.126) over 15. The general LAK level of the participants who perceived themselves as “very competent” was found as 7.21 (SD=1.989), and it was found as 6,70 (SD=2.081) for the ones who perceived themselves as “competent”. Those mean scores were too low in order to describe themselves as “very competent” or “competent” as they were even less than average success score which is 7.5 over 15. Therefore, it can be said that perceived self-competency of the participants in assessing listening aren’t in line with their actual LAK in assessing listening.

Table 19. EFL Teachers’ Perceived Self-Competency and Their LAK Level in Assessing Writing

Assessing Writing	N	M
very competent	107	7.03
competent	133	6.74
not very competent	32	6.97

	Sum of Squares	df	Mean Square	F	p
Between Groups	5.042	2	2.521	.480	.619
Within Groups	1413.193	269	5.254		
Total	1418.235	271			

Considering assessing writing, no significant difference was found at the $p < .05$ level among the participants who perceived themselves as “very competent”, “competent” or “not very competent” ($F(2,269) = .480, p = .619$). The descriptive analysis demonstrated that the mean score of the participants who perceived themselves as “very competent” in assessing writing was 7,03 (SD=2.288) which isn’t even half of the total score 15. Moreover, the mean score of the participants who perceived themselves as “competent” ($M = 6.74, SD = 2.211$) was even less than the mean score of the participants who perceived themselves as “not very competent” ($M = 6.97, SD = 2.621$). Therefore, it can be said that the participants might have different perceptions of their own abilities, which don’t align with their actual performance.

Table 20. EFL Teachers’ Perceived Self-Competency and Their LAK Level in Assessing Speaking

Assessing Speaking	N	M
very competent	98	8.26
competent	139	8.47
not very competent	35	7.83

	Sum of Squares	df	Mean Square	F	p
Between Groups	11.861	2	5.930	.876	.417
Within Groups	1820.198	269	6.767		
Total	1832.059	271			

The findings demonstrated that the LAK level of the participants who perceived themselves as “very competent” was 8.26 (SD = 2.141), the ones who felt “competent” had a mean score of 8.47 (SD = 2.767), and the ones who felt “not very competent” had a mean score of 7.83 (SD = 3.063). The descriptive analysis also showed that the mean scores of the participants who perceived themselves as “competent” was higher than the ones who perceived themselves as “very competent”. One-way ANOVA test results revealed that there was no significant difference at the $p < .05$ level for three different perception groups in terms of their LAK level in assessing speaking ($F(2,269) = .876, p = .417$). This indicates that the participants’ perceptions of their self-competency don’t align with their actual performance.

4.2. Qualitative Data Analysis

In order to find out the participants’ opinions on language testing and assessment, semi-structured interviews were conducted with 10 EFL teachers working in either private or state K12 schools. The data obtained from the interviews were analyzed via content analysis. The answers were coded and categorized under three different themes which are pre-service educated related issues, the results of the scale and EFL teachers’ needs in LTA. The findings are presented in the following table.

Table 21. Themes and Codes of the EFL teachers’ Opinions on Language Testing and Assessment

Themes	Codes	<i>f</i>
Pre-Service Education Related Issues	Full of theoretical lessons	8
	Limited number of practices	7
	Limited lesson hours	5
	No proper feedback given	4
	Crowded classes	2
	Effective LTA courses	1
	Lack of knowledge	6

	Incompetence of EFL teachers	7
The Results of the Scale	Not being able to get any in-service trainings	6
	Lack of self-criticism	5
	Not getting feedback	2
	Lack of motivation	3
	Having limited time to search or study	8
	Impracticality	4
	Testing and evaluation departments at schools	5
EFL Teacher's	Regular trainings with up-to-date information	9
Needs in LTA	More lessons in pre-service education with a chance of practicing	5
	Less workload to be able to focus on assessment practices	8
	Institutions supporting teachers for formative assessments	6

4.2.1. RQ5: The Opinions of EFL Teachers Regarding Pre-service Education

As a result of the semi-structured interviews, it was revealed that the participants mostly weren't happy with the pre-service education regarding language testing and assessment. As it can be seen from Table 21, five participants expressed that they had limited lesson hours for LTA and it wasn't enough to go into detail in the area. The following comment was made by one of the participants.

"To be honest, as far as I remember, I had just one LAT course in my last year at the university. Other than that, I didn't have any, and it was just a course which introduced us LAT in a broader sense. However, after I started teaching, I realized that student evaluation is a continuous process, and it deserves more attention."

Eight of the participants mentioned that pre-service education includes too much theoretical knowledge while other seven participants were saying that they didn't have any chance for practising LTA activities. In this regard, the following comments were made:

"In our lessons, we mostly focused on theory. We did not prepare many tests and exams, which means we could not put the theoretical knowledge into practice. Therefore, right now, it is very difficult for me to do some assessment related activities in a real classroom setting. In addition, our focus at university

was just to learn the information and pass the exams. As we didn't think that we would use this information in the future, we easily forgot everything. The theory didn't help me a lot actually because after the exams, I forgot everything."

"While 90% of the course was theoretical, 10% was practical. We mostly learnt basic terminology, and issues such as reliability and validity. We didn't have many chances of practice. I think it would have been easier if we practiced testing and assessment related activities because theory took such long time to learn and understand."

Among the participants, there were four people mentioning that they didn't get any proper feedback from the instructors in LTA courses while two participants were mentioning crowded classes as a reason affecting the effectiveness of LTA courses. Regarding this issue, two of the participants made the following comments:

"Even though we practised some assessment related activities in the class, I don't think that they helped me a lot on this issue because I didn't get any feedback from my instructor. Yes, he told us to prepare some things but at the end I never learnt what I did wrong because he just announced the scores directly rather than giving some feedback. That's why, I don't think that they helped me improve myself."

"As there were so many students taking that LTA course, it wasn't possible for the instructor to create an effective discussion environment to examine sample LTA activities in a detailed way. Just following the presentations didn't help me a lot as it was just theory. Also, I just couldn't follow the explanations of the instructor effectively as there were many people distracting me in the class."

As different from these, there was only one participant who mentioned the effectiveness of the pre-service education in terms of LTA courses. Regarding this issue, she made the following comment:

"Actually, I had "Language Testing and Assessment" course in practice and it was very effective. In that course, we were supposed to prepare an exam and there were a lot of mistakes in it and then every week, we discovered our mistakes with the help of theoretical knowledge and discussed on the exams."

Everyone in the class criticized and corrected the mistakes in the exams. Even now, I can remember my own mistakes, and I try to be careful about them while preparing an exam. Therefore, I think it prepared me well to test learners.”

4.2.2. RQ5: The Opinions of EFL Teachers Regarding Their LAK Level and the Findings of the Scale

The participants were also asked about their opinions on the findings of the scale. When they were asked to comment on the general LAK level of the participants, which was found as 31.59, and the possible reasons behind it, seven of them mentioned the teachers’ lack of knowledge in LTA as a result of the pre-service education related issues mentioned above. Regarding this, some of the comments made by the participants can be seen below:

“I think it shows that teacher training departments failed to provide student teachers with what they need when they are in active teaching. The reason could be what you learn in college might fade away in time. It isn’t permanent. To be honest, when I was at the university, I used to know more about testing, evaluation and assessment.”

“Maybe, if they had the course on the basis of theory and didn’t practice this knowledge, it may not have been understood clearly. Therefore, it is hard to remember. Also, when we lack practice over time, we easily forget what we are taught.”

As a result of the quantitative part of the study, in terms of assessing the skills, the highest knowledge level was found in assessing reading (9.58 out of 15) whereas the lowest level was in assessing listening (6.82 out of 15). The knowledge level in assessing other skills was found as 8.31 in speaking and 6.88 in writing. When the participants were asked to share their opinions on this, it was revealed that EFL teachers’ incompetency and impracticality of assessing some skills were the reasons behind. The following comments were made by some participants on this issue:

“As I mentioned in the previous question the infrastructure at most schools isn’t convenient to assess certain skills. There is no chance to actively do listening practices at most schools especially the state ones. The allocated time or the number of lessons isn’t enough to practice many skills and some teachers aren’t capable of teaching those skills. Teachers become teachers thanks to

their level of reading, how we are prepared to become a teacher actually gives us the answer. Many teachers with different background in high school learn English with written materials and they are hardly ever encountered any authentic materials. Most of us never heard anything in English or write anything, so not until the time of university we actively start using our productive skills. It cannot be expected from teachers to be good at teaching skills that they don't even acquire earlier."

"I think in Turkey we give great importance to reading in English. We know how to assess it because of our own language learning experience. This could be the reason why reading is the highest. Testing listening needs certain equipment and environment which is sometimes hard to have in some schools. So, teachers may ignore to test this skill. Another reason could be that language teachers may not be so competent enough to assess listening."

Also, in the quantitative part of the study, it was revealed that different demographic characteristics that are years of experience, educational background (BA – MA- PhD), the BA program being graduated (ELT or non-ELT), working at a private or state school, experience in test preparation and attending trainings on testing and assessment had no influence on the participants' language assessment knowledge. Teachers' lack of motivation for assessment related issues, not being able to get any in-service trainings and having limited time to search, study and improve themselves in LTA were mentioned as the reasons behind. Some quotes supporting these codes can be seen below:

"Even such situations do not have an effect on language assessment literacy, it shows that LTA has not been sufficiently emphasized, researched and given importance. For example, if someone with experience has never paid attention to this issue in these years, or, if someone who has participated in assessment and evaluation training only attended with the intention of getting a certificate, or if the trainings given in a more general sense are insufficient in this regard, I mean if they aren't motivated enough to improve themselves on this topic, it seems to me that the result of the research is very normal."

"Especially in private schools, teachers' workload is a lot. As it will require extra time to search and study LTA, they don't prefer doing it. They aren't

motivated enough and they don't have time for this. ELT graduates mostly learn theoretical knowledge. So, I think they are the same with non-ELT graduates when it comes to practice. Also, because of economic situation, many teachers don't have chance to pay and get some trainings. As most of the schools don't offer in-service trainings, teachers only have limited chance to improve themselves in this area. When we think about all these things, I think it is very normal that these features don't have any impact of EFL teachers' assessment knowledge."

The quantitative data also showed that there were no participants describing themselves as "not competent." The participants stated that they were very competent, competent or not very competent for all skills. As a result of the analysis, it was found that there was no significant difference between different perception groups except for listening skill, and most of the participants' actual score didn't match with their perception. When they were asked about to comment on this, the participants mentioned lack of self-criticism and not getting feedback as the reasons behind.

"Yes, I was expecting this result, actually, but I was surprised that no one perceived himself as not competent. Unfortunately, in our society, no one criticizes himself. We are too weak for this. There is also a phenomenon called "teacher arrogance" in our society. People have the perception that the teacher has to know everything. That's how the students see teachers. For example, sometimes when students ask me word, I don't know it at that moment, or I can't remember. I think this is very normal. However, due to high expectations, teachers do not accept that they don't know it, and they don't criticize themselves. Unfortunately, we are a very weak in terms of self-criticism."

"When you work for a long time, you may not see your deficiencies as long as not having feedback on your work. One reason could be that schools of each levels fail to evaluate assessments process. Maybe there should be more strict rules and guidelines to follow to set the standards."

"It is clear that we teachers don't want to see our flaws and there aren't many chances to check our progress in teaching. After we graduate from the faculty, we barely get any constructive feedback from our colleagues and there seems

no need to improve ourselves. As long as the students pass their exams, we imagine that we teach them what is on the curriculum, and we are the best and the most loved teachers. In addition to that, we are not accustomed to evaluating skills such as speaking, listening, and writing therefore it is less likely to see our actual level.”

4.2.3. RQ5: EFL Teachers’ Needs in Language Testing and Assessment

As a result of the semi-structured interviews, participants opinions on their language testing and assessment related needs were revealed. Nine of the participants stated that they need regular trainings with up-to-date information. The following comment was made by one the participants.

“I think I need trainings according to the curriculum updated regularly, and also there should be a continuity on this. For example, if I attend an assessment related training, I shouldn’t trust it for coming ten years. There needs to be a necessity to update ourselves.”

Five participants mentioned the need for a specific testing and evaluation department at schools while eight of them mentioned teachers’ need for less workload to be able to focus on assessment practices. The following comments were made by some participants:

“We really need some schools that do not tire the teachers after long working hours and a system that encourages teachers to join those in-service trainings willingly.”

“There should be a separate testing and assessment departments at each school, and that department should provide some trainings to the teachers. There should be some guidelines prepared by them on language testing and assessment so that we can consult to if necessary to be able to increase our knowledge. It would be really helpful especially for newly graduated teachers.”

By six of the participants, it was also mentioned that there is a need for institutions prioritizing formative assessments rather than aiming getting high scores in summative assessments. The participants said the followings on this issue:

“Teachers should be given more time to focus on formative assessments. Workload could affect badly teachers’ performance on these. Instead of grades and marks, assessment should focus on the whole language journey of students like portfolios. The system should change and teachers should be supported on this.”

“In my institution or any place in Turkey, the existence of summative is seen as the most crucial assessment type, this should be avoided and one-time tests should become less important. Not every student is capable of perform well during the day of test and this should be left out as the main criterion to understand the knowledge of a student. Predetermined tasks with clear objectives that require students to carry out exercises in all skills can be more beneficial than summative. According to my beliefs, exclusion of summative assessment could be the better option which allows teachers to become more knowledgeable in terms of other skills teaching and assessing.”

Lastly, in terms of the needs in pre-service education context, it was stated that pre-service teachers need more than one lesson on language testing and assessment with a chance of practicing rather than just focusing on theory. The following comment was made by one of the participants:

“You know last year of the university is very late to take an LTA course. I think we need to start taking these lessons starting from first year at university by the end of it regularly. Also, we need to have a chance for practicing. Otherwise, it won’t be permanent, and just theory won’t help us in real classroom setting.”

5. DISCUSSION

5.1. Discussion on the EFL Teachers' General and Skill-based Language Assessment Knowledge Level

In the first research question, it was aimed to investigate the general and skill-based LAK levels of the participants working at K12 schools. According to the quantitative results, it was seen that general LAK level of the participants ($M=31.59$, $SD=6.366$) was slightly higher than the half of the total score (30 over 60) indicating that the participants had average success in the current research study and there are some areas that needs improvement. The lowest mean score was found to be as 6 and the highest mean score was found to be as 45. The standard deviation also showed a considerable among of variation among the participants' scores.

This result can be discussed in two ways regarding both contradiction and alignment with other studies conducted before. The existing studies using the same scale in the literature indicated a contradiction with the current study. To start with, Kaya (2020) and Sevilen (2021) found out that the literacy levels of the participants were significantly high indicating that the participants were knowledgeable in language assessment. Within this context, the current study just stands in the middle showing an 50% average success. However, it doesn't mean that the participants were very knowledgeable regarding assessment related issues. It still indicates their insufficient knowledge in the area and supports the existing literature. In her research study, Ölmezer-Öztürk (2018) found the mean score of the participants' general LAK considerably low ($M=25.19$, $SD=11.39$). Similarly, Çetin-Argün (2020) came up with an average (47%) which indicated insufficient knowledge of the participants with a low mean score. In their study which was conducted with 350 teachers by using an LTA scale adapted from Vogt and Tsagari (2014), Mede and Atay (2017) yielded results indicating that the participants had limited knowledge in assessment regarding four skills. Hatipoğlu (2015) also shed light on the insufficient knowledge of the participants in testing and assessment as a result of her research study conducted with 124 pre-service EFL teachers. Likewise, as a result of their study conducted with 891

EFL teachers working in China, Xu and Brown (2017) revealed that the participants had a very basic level of language assessment literacy. The current study shows parallelism with these results as the mean score 31,59 still indicates insufficiency of the participants in testing and assessment.

Regarding four skills, in the current study, the highest mean score was found in assessing reading ($M=9.58$, $SD=2.225$) followed by assessing speaking ($M=8.31$, $SD=2.600$) and assessing writing ($M=6.88$, $SD=2.288$). The lowest mean score was in assessing listening ($M=6.82$, $SD=2.087$). The results are in line with the findings of the studies conducted by Ölmezer-Öztürk (2018), Çetin-Argün (2020), Sevilen-Yılmaz (2021). In all of these studies, the highest mean score belongs to assessing reading. It is followed by assessing speaking, assessing writing and the lowest mean score belongs to assessing listening.

There might be several reasons for reading to have the highest mean score. Prioritization of reading might be one of them. As it was stated by Hubley (2012), most of the language input comes from reading sources around us. It is assumed that people need to read as a natural consequence. That's why, reading is considered as the most important skills, and as a result, it is given so much importance. Also, practicality and easiness of classroom practices might be another reason. It is not difficult for teachers to assess reading as there are lots of pre-prepared reading materials and clear rubrics (Backlund, Brown, Gurry, & Jandt, 1980).

For assessing listening to be the least succeeded one among other skills, complexity of listening skill and assessing might be one of the reasons. As it isn't directly observable and requires specific techniques and materials, it challenges teachers. Alderson and Bachman (2001) described listening as the skill which is least understood and developed in testing and assessment. The difficulty of preparing listening texts, finding suitable materials including different accents and intonations, or not having enough time and equipment cause teachers to ignore this skill. Therefore, it may end up with being the last among other skills. Regarding assessing writing to be among the skills which needs improvement, underestimation of the difficulty of assessing it and not being aware of the required techniques to evaluate written work might be the reason. As Weigle (2012) stated, just giving random topics to students, and trying to grade them without following any rubrics or so is not considered as a good way of assessing writing.

It can be said that in the current study, the result of assessing speaking is surprising. However, it is the second skill which has got the highest score among others in other studies, too. The reason of this might be related to the groups of participants. Sevilen-Yılmaz (2021) and Ölmezer-Öztürk (2018) conducted their studies with EFL teachers teaching in preparatory programs. As students there need to be assessed in terms of their speaking skills to be able to skip preparatory programs, teachers working there need to employ related classroom activities. As it is considered as the most difficult skill to assess by many (Bachman & Palmer, 1981), it might be given much more attention by teachers, and this affects its rank among other skills. However, in the current study, the participants were the ones working at K12 schools. The reason of such a result again might be about the participants. Especially in private schools, the communicative skills of students have started to receive much attention because of the increasing of globalization. Instead of learning grammar rules, students are expected to communicate fluently in English. As a result, teachers are trying to do speaking-related assessment activities more.

To conclude, the results of the current study both contradicts and align with the studies in literature in different ways. Even though numerical results contradict with some of the earlier studies, the implications of them are in line with the existing literature. The general mean score 31.59 over 60 implies that the teachers working at K12 schools need to improve their language assessment literacy.

5.2. Discussion on the Relationship among EFL teachers' Skill-based Assessment Knowledge

The second research question aimed to find out the relationship among skill-based assessment knowledge and general LAK. The findings demonstrated that assessing writing and assessing speaking were highly and positively correlated with general LAK whereas assessing reading and assessing listening were moderately and positively correlated. Moreover, the findings also presented that there were significant positive correlations between all types of skill-based knowledge. Assessing writing and assessing speaking had the highest correlational level which is a moderate one whereas assessing reading and assessing speaking had low level correlation. Within this context, it can be inferred that due to these positive correlations among skill-based knowledge and general LAK, language assessment must be considered as a

comprehensive and interconnected process because an improvement in one of the skills affects other skills. This result is also supported by Ölmezer-Öztürk (2018) describing LAK as “a holistic phenomenon with its own interrelated elements (p.91).”

5.3. Discussion on Certain Background Variables’ Impact on LAK Level of EFL Teachers

The third research question aimed to explore the effect of certain background variables on LAK level of EFL teachers. Those variables were years of experience, the BA program being graduated (ELT or non-ELT), working place (at a private or state school), grade level the participants teach, attending trainings and workshops on testing and assessment, experience in test preparation/development and interpretation. The findings demonstrated that none of them have an effect on participants’ LAK levels.

To start with “years of experience”, it was seen that there is no difference among the LAK levels of the most experienced teachers and less experienced ones. It can be concluded that language assessment knowledge isn’t a phenomenon which increases gradually according to years of experience in teaching. This finding shows parallelism with the results of another study conducted by Jannati (2015) with the purpose of investigating the years of teaching experience’s effect on LAK. Likewise, Tao (2014), Büyükkarcı (2016), Öz and Atay (2017) conducted similar studies. The results of their studies also demonstrated that there was no significant difference among participants’ LAK levels with different years of teaching experience. Therefore, the current study aligns with the existing literature. As mentioned and supported in semi-structured interviews, there might be several reasons of this finding such as lack of knowledge, lack of motivation, lack of trainings, having limited time to search, study and improve.

The second background variable was “the BA program being graduated.” The effect of being a graduate of ELT or non-ELT departments on LAK was investigated. The results demonstrated that there was no significant difference between these two groups. As different from this finding, Tao (2014) and Sevilen-Yılmaz (2021) explored in their studies that ELT graduates’ LAK levels were significantly higher than non-ELT graduates. That’s why, the current study contradicts with these results. Nevertheless, this finding also supports some of the studies in the existing literature.

Ölmezer-Öztürk (2018) and Kaya (2020) found out that being an ELT graduate or non-ELT graduate doesn't affect participants' LAK level. As it was mentioned in the semi-structured interviews, insufficient LTA lessons in pre-service education might be one of the reasons of this. In her study, Hatipoğlu (2015) also stated that pre-service education doesn't provide learners with sufficient and efficient lessons in LTA.

With regard to working place, it was explored that working at a private or state K12 schools doesn't affect the participants LAK level, namely, working in a private school doesn't make the teachers more assessment literate. Even though private K12 schools look like providing more trainings in LTA and have higher professional expectations from their teachers, the results show that it doesn't make any difference. Relatedly, semi-structured interviews provided some information about the possible reasons of this finding, and it was revealed that even though they are offered more trainings in private schools, teachers' lack of motivation for assessment related issues, not being able to get any in-service trainings depending on the school they work in, and having limited time to search, study and improve themselves in LTA because of workload were mentioned as the reasons behind.

With respect to the variable "grade level the participants teach", it was aimed to explore whether the participants LAK level change according to the grade level they teach in such as primary, middle and high school levels. The findings demonstrated that there was no significant difference among the groups, which means the grade level being taught didn't affect the participants' LAK level. Lastly, other variables "attending trainings and workshops on testing and assessment" and "experience in test preparation/development and interpretation" were found as not affecting LAK level. These findings also shed light on the importance of the quality of the trainings as well as quantity because even the groups of participants stating that they participated in trainings in LTA weren't different from the ones stating that they didn't. Malone (2008) stated that just a training itself is not enough to improve LAK. It needs to equip teacher with the necessary practical information which allows in-class practices. As stated by Mede and Atay (2017), the number of trainings offered to teachers are limited. Trainings offered regularly in long term needs to be provided for a better understanding of LTA. The only significant difference was found in assessing reading in terms of the effect of "attending trainings and workshops on testing and assessment". It was seen that the attendance to trainings and workshops played a vital

role in the participants' LAK in assessing reading. The reason of this might be about increased familiarity of assessing reading tools and frequent use of them in class. After being introduced the essentials of assessing reading and different assessment methods in reading in trainings and workshops, the participants become more familiar with assessing reading. The more they are familiar with it, the better they apply them in class, and as a result they could develop a better understanding of assessing reading, which causes the only significant difference on the effects of attending trainings and workshops. Also, even being experienced in test preparation/development and interpretation didn't make any difference. This was also mentioned and supported in semi-structured interviews by participants with the reason of getting insufficient feedback on their LTA practices.

5.4. Discussion on Perceived Self-competency and Actual Language Assessment Knowledge Level

The fifth research question aimed to examine whether the participants' LAK level change according to their perceived self-competency or not. When the descriptive analysis results were examined, it was seen that no one perceived themselves "not competent". Almost all of the participants stated they were either "very competent" or "competent" in assessing each skill. While the highest mean score belongs to the perception of group of "very competent" in assessing reading and assessing listening, it belongs to the perception group of either "competent" or "not very competent" in assessing writing and assessing speaking. Also, the numerical scores for some perception groups and skills aren't high enough to be able to define them as "competent". That's why, it can be said that their actual LAK level and their actual performance are far from each other. According to the results of one-way ANOVA, no significant difference was found among different perception groups for each skill except for assessing listening. These findings are supported by Jannati (2015) demonstrating that the knowledge of the participants related to testing and assessment was enough, namely, the participants were competent in LTA. The reason of this inconsistency might be because of teaching emphasis over some skills. For instance, while traditionally there is a strong emphasis on reading and writing skills, there isn't such an emphasis on listening, which leads receiving less attention in training programs and ends up with having greater variance in competence. Also, as supported

in semi-structured interviews, people have lack of self-criticism and tend to overestimate their abilities. However, having more realistic views on assessing listening as a result of receiving less attention compared to other skills might lead that significant difference. These findings highlight the importance of providing more efficient and comprehensive professional development trainings to improve assessing listening skills of teachers.

5.5. Discussion on The Opinions of EFL Teachers Regarding Pre-service Education and EFL Teachers' Needs in Language Testing and Assessment

The last research question addressed the participants opinions regarding pre-service education the findings of the scale and EFL teachers' needs in LTA. The semi-structured interviews provided deeper understanding of the participants views on these themes.

With regard to pre-service education, except for just one participant, they were all unhappy with the quality of education. Only one of the participants stated that LTA course offered in pre-service education was very effective and helped the participant a lot to improve himself / herself in the field. However, all other participants had some negative views about it. Eight of the participants out of ten criticized pre-service education in LTA by saying that full of theoretical knowledge is offered while the other seven participants criticizing having limited number of practices. The participants also stated that they had just one LTA course in their senior year and it wasn't enough to cover all information needed and give enough chance to learners to improve themselves. That's why, having limited lesson hours was another item mentioned in the interviews. Also, some participants complained about the ineffectiveness of LTS courses as no proper feedback was given to students so that they could see their mistakes and improve themselves. Lastly, two participants criticized crowded classes as they didn't allow for an effective teaching and learning environment. To conclude, as it can be seen, among 10 participants, just one EFL teacher provided positive opinions regarding pre-service education in LTA. All other nine participants weren't happy with the courses offered in LTA, which highlights the emergent need of a change in pre-service education curriculum. These results are in line with the findings of Mertler (2003) and Hatipoğlu (2010), which showed in the insufficiency of the LTA course in pre-service education.

Regarding the needs of EFL teachers in LTA, nine of the participants stated that sustainability of trainings in the field is very important, and they need regular training with up-to-date information, namely, attending a training once and using the information learnt there for years shouldn't be the case. This finding aligns with the findings of Herrera and Macias (2015) which highlights the importance of providing teachers with ongoing trainings with updated information. As mentioned in pre-service education part, five of the participants came up with a need of more lessons in pre-service education with a chance of practicing by criticizing the limited chance of practicing and less lesson hours. Likewise, Lam (2015) mentioned the importance of including theory and practice in trainings. Almost all of the participants were complaining about the workload in the institutions they work in. They stated that they were so busy that they didn't have any time or motivation to focus on assessment practices. Therefore, they strongly desired for a working environment which has less workload so that they can have a chance to improve themselves in the related area. Relatedly, five participants remarked that there must be separate testing and evaluation departments at schools so that they can provide necessary trainings for better implementation of assessment practices. Lastly, the interviews shed light on the participants' need of supportive institutions for formative assessments. Six of the participants just stated that their schools' priority is providing students with higher scores which makes parents happy rather than focusing on their strengths and weaknesses by applying appropriate assessment practices. That's why, they stated that it would be really helpful to focus on the process by using some formative assessment practices rather than aiming to collect higher scores with the help of supportive institutions. To conclude, the participants' need cannot be only based on pre-service education such as more lesson hours and chance of practicing as their needs are also institutional based such as specific testing and evaluation departments at schools, less workload and supportive institutions. Both pre-service and in-service training needs of teachers need to be taken into consideration for better assessment practices.

6. CONCLUSION

6.1. Summary of the Study

The current study aimed to explore general and skill-based language assessment literacy levels of EFL teachers working at K12 schools in Turkey. Additionally, it was aimed to find out whether their level of skill-based language assessment knowledge relate to one another or not. Also, the effect of certain background variables such as years of experience, the BA program being graduated (ELT or non-ELT), working place (at a private or state school), grade level the participants teach, attending trainings and workshops on testing and assessment, experience in test preparation/development and interpretation affect LAK level of EFL teachers or not was a matter of research. In the last part of the quantitative study, the relationship between their perceived self-competency and general LAK level was explored. Finally, in the qualitative part, the current study aimed to analyze the participants' opinions on LTA in relation to pre-service education, the findings of the scale and their needs in LTA.

The findings of the quantitative part indicated that the participants demonstrated an average success which indicates a need for some improvements in LTA. Regarding skill-based LAK, assessing reading was ranked as the area the participants are the most knowledgeable in while assessing listening was ranked as the area the participants are the least knowledgeable in. It was also explored that there were significant positive correlations between all types of skill-based knowledge and general LAK. As for certain background variables, it was found out that there was no significant difference between groups, and these variables didn't affect the participants' LAK level. The results revealed that the participants mostly perceived themselves either "very competent" or "competent" in assessing these four skills. There was no one choosing "not competent" option for description. No significant difference was found between different perception groups except for listening.

The findings in the qualitative part shed light on the reasons of the findings of the quantitative part. The codes such as lack of knowledge, incompetence of EFL teachers, not being able to get any in-service trainings, lack of self-criticism, not getting feedback, lack of motivation, having limited time to search or study and impracticality were emerged as the possible reasons of the findings of the scale. Regarding pre-service education issues, except for one participant, all the others had negative opinions. Full of theoretical lessons, limited number of practices, limited lesson hours, no proper feedback given, crowded classes were emerged codes showing reasons of having negative opinions about pre-service education. Lastly, regarding the participants' needs analysis, it was revealed that there is a need for trainings both in pre-service education and in-service. Their needs in LTA were listed as testing and evaluation departments at schools, regular trainings with up-to-date information, more lessons in pre-service education with a chance of practicing, less workload to be able to focus on assessment practices, institutions supporting teachers for formative assessments.

6.2. Practical Implications

The current study aimed to explore the language assessment literacy levels of EFL teachers working at K12 schools in Turkey and their opinions on pre-service education, the findings of the scale and their needs in LTA. Based on the findings of the qualitative and quantitative studies, some implications are suggested.

First of all, as the general LAK level of the participants indicated, the participants are in need of improving themselves in LTA. EFL teachers could be provided with professional development trainings related to LTA. As it was suggested in the interviews, they must be ongoing trainings with recent innovations. However, as it was also seen in the quantitative findings, attending trainings didn't affect participants' LAK level. That's why, the content of trainings could be carefully designed and the number of trainings to be attended could be decided attentively.

Secondly, there could be specific testing and evaluation departments at school. Thus, they can prepare some guidelines to help teachers with their assessment practices and prepare trainings and workshops on regular basis. Also, teachers could be offered less workload so that they can focus on their assessment practices better and have a chance of attending trainings.

Lastly, as it can be seen in the qualitative findings, the participants have negative opinions about the effectiveness of pre-service education in LTA. The quantitative findings also showed that there is no significant difference between ELT and non-ELT graduates, which shows the ineffectiveness of pre-service education. That's why, the curriculum could be revised and the number of the lesson could be increased.

To conclude, the suggested implications above might improve the quality of teaching and learning by providing teachers with a better understanding of assessment practices. That's why, they might be critical for language teachers, teacher educators, curriculum developers, policy makers and administrators.

6.3. Limitations

The first limitation of the current study is about the sample size. The current study was conducted with relatively small number of participants compared to all EFL teachers working at K12 schools. The number of participants in quantitative part was 272. Also, it was just 10 in the qualitative part. That's why, it may not represent the entire population of EFL teachers. Involving more participant would have yielded a more in-depth understanding of the data. Another limitation of the study is about the length of the research scale used. As the scale includes four subscales including 15 items in each, it takes some time to be able to answer all the items in the scale, which might be demanding. Because of the scale's length, the participants might have lost attention at some parts, and this might affect their responses' reliability.

6.4. Recommendations for Further Research

Based on the findings and the limitations mentioned above, some recommendations can be proposed. Firstly, in order to provide a broader understanding of LTA, the current study can be conducted with a larger sample size. Secondly, because of the extended length of the scale, further research can be conducted by focusing on either only productive skills or receptive skills. Also, as the third part of the study, participants assessment practices might be examined to check to what extent theory and practice match. Lastly, it might be perfect to examine the relationship

between teachers' assessment literacy and student outcomes to see whether it causes an improvement in students' performance or engagement in the lesson or not.



REFERENCES

- Alderson J. C., Bachman L. F. (2001). Series editors' preface. In Buck G., *Assessing listening* (pp. x–xi). Cambridge: Cambridge University Press.
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-Assessment. *Assessment and Evaluation in Higher Education*, 32 (2), 159-181.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., and Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Backlund, P., Brown, K., Gurry, J., and Jandt, F. (1980). Evaluating speaking and listening skill assessment instruments: Which one is best for you? *Language Arts*, 57 (6), 621-627.
- Baker, B. A., & Riches, C. (2017). The development of EFL Examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581. <https://doi.org/10.1177/0265532217716732>
- Ballıdağ, S. (2020). *Exploring the language assessment literacy of Turkish in-service EFL teachers*. (Publication No. 627937) [Master's Thesis, Kocaeli University]. Council of Higher Education Thesis Center.
- Beck, K. A. (2005). Ethnographic decision tree modeling: A research method for counseling psychologists. *Journal of Counseling Psychology*, 52, 243–249.
- Boraie, D. (2012, September 12). *Formative vs. summative assessment: Does it matter?*. TESOL Connections. <http://newsmanager.commpartners.com/tesolc/issues/2012-09-01/3.html>
- Brindley, G. (1989). *Assessment achievements in the learner-centered curriculum*. Sydney: National Centre for English Language Teaching and Research, MacQuarie University.
- Brinkmann, S. & Kvale, S. (2005). Confronting the ethics of qualitative research. *Journal of Constructivist Psychology*, 18, 157-181. <https://doi:10.1080/10720530590914789>.
- Brown, J. D., & Hudson, T. (1998). The Alternatives in Language Assessment. *TESOL Quarterly*, 32(4), 653–675. <https://doi:10.2307/3587999>

- Brown, H. D. (2003). *Language assessment principles and classroom practices*. Pearson Education.
- Brown, H. D. (2014). *Principles of language learning and teaching* (6th ed.). Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Büyükkarci, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Egitim ve Ogretim Dergisi*, 6(3), 333-346.
- Caspary, M., & Boothe, D. (2017). Assessment of EFL Through the Process of Problem-Based Learning. In R. Al-Mahrooqi, C. Combee, F. Al-Maamari, V. Thakur (Eds.), *Revisiting EFL assessment* (pp.9-27). https://doi:10.1007/978-3-319-32601-6_7
- Chan, Y. (2016). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, 7 (1), 37-62.
- Cirit, N. C. (2015). Assessing ELT Pre-Service Teachers via Web 2.0 Tools: Perceptions toward Traditional, Online and Alternative Assessment. *The Turkish Online Journal of Educational Technology*, 14(3), 9-19.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Coombe, C., Folse, K. S., & Hubley, N. J. (2007). *A practical guide to assessing English language learners*. University of Michigan.
- Coombe, C. A., Folse, K. S., & Hubley, N. J. (2019). *A practical guide to assessing English language learners*. Langara College.
- Coombe, C. (2012). Second language assessment. In C. Coombe, P. Davidson, B. O'Sullivan, and S. Stoyloff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 1-5). Cambridge University Press, USA.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced Mixed Methods Research Designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 209-240). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

- Creswell J. W. (2008). *Educational research: Planning, conducting, and evaluating qualitative and quantitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Çetin-Argün, B. (2020). *Language Assessment Knowledge of Preservice Teachers of English as a Foreign Language*. (Publication No. 615901) [Master's Thesis, Çag University]. Council of Higher Education Thesis Center.
- Davin, K. J., & Heineke, A. J. (2016). Preparing Teachers for Language Assessment: A Practice-Based Approach. *TESOL Journal*, 7(4), 921-938.
- Deluca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438. <https://doi.org/10.1080/0969594x.2010.516643>
- Doğru, D. (2020). *Assessment literacy conceptions and practices EFL teachers in higher education*. (Publication No. 614678) [Master's thesis, Bahçeşehir University]. Council of Higher Education Thesis Center.
- Douglas, D. (2014). *Understanding language testing* (2nd ed.). New York, NY: Routledge.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.
- Falsgraf, C. (2005, April). Why a national assessment summit? New visions in action. National Assessment Summit. Meeting conducted in Alexandria, Va. Retrieved from <https://files.eric.ed.gov/fulltext/ED527580.pdf>.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Farhady, H. (2019). A Cross-Contextual Perspective on EFL Teachers' Assessment Knowledge. *Edu 7: Yeditepe Üniversitesi Eğitim Fakültesi Dergisi*, 8 (10), 1-19.
- Fleming, M. (2007, November). *Languages of schooling within a European framework for language of education: Learning, teaching, assessment – Report*. Report of the Intergovernmental conference of the same name, Prague, 8-10 November 2007. Strasbourg, France: Council of Europe, Language Policy Division.
- Flick, U. (2002) *An introduction to qualitative research* (2nd ed.). London. England: Sage.

- Fulcher, G. (2012). *Assessment Literacy for the Language Classroom*. *Language Assessment Quarterly*, 9(2), 113–132. doi:10.1080/15434303.2011.642041
- Graves, D. H. & Sustain, B. S. (1992) *Portfolio portraits*. London: Heinemann.
- Genesee, F. (2001), Evaluation, In Carter, R., Nunan, D. (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 144-150). Cambridge: Cambridge University Press.
- 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>.
- Green, A. (2018). Assessment of learning and assessment for learning. *The TESOL Encyclopedia of English Language Teaching*, 1–6.
<https://doi.org/10.1002/9781118784235.eelt0352>
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.
- Harris, L., Irving, S. E., & Peterson, E. (2008). *Secondary Teachers' Conceptions of the Purpose of Assessment and Feedback*. [Paper presentation]. The Australian Association for Research in Education Conference, Brisbane.
- Harris, M., & McCann, P. 1994. *Assessment*. Oxford: MacMillan Heinemann.
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs. Part one: General findings*. Retrieved from <http://www.ealta.eu.org/resources.htm>
- Hatipoğlu, C. (2010). Summative Evaluation of an English Language Testing^[1] and Evaluation Course for Future English Language Teachers in Turkey. *English Language Teacher Education and Development (ELTED) Journal*, 13, 40-51.
- 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.
- Herrera, L. and Macias, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colomb. Appl. Linguist. J.*, 17(2), 302-312.
- Hubley, N. N. (2012). Assessing reading. In C. Coombe, P. Davidson, B. O'Sullivan, and S. Stoyhoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 211- 217). Cambridge University Press, USA.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402.
<https://doi.org/10.1177/0265532208090158>

- Jannati, S. (2015). ELT Teachers' Language Assessment Literacy: Perceptions and Practices . *The International Journal of Research in Teacher Education* , 6 (2) , 26-37.
- 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.
- Johnson, R. B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*. Boston, MA: Pearson Education, Inc.
- Kaya, T. (2020). *Exploring Language Assessment Literacy of EFL Instructors in Language Preparatory Programs*. (Publication No. 614695) [Master's Thesis, Bahçeşehir University]. Council of Higher Education Thesis Center.
- Lado, R. (1961). *Language Testing*. London: Longman.
- Lam, R. (2014). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/10.1177/0265532214554321>
- Lynch B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351–372.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, N. J.: Lawrence Erlbaum.
- Malone, M. E. (2008). Training in language assessment. In E. Shohamy and N. Hornberger (Eds.), *Encyclopedia of language and education, Vol. 7: Language testing and assessment* (2nd ed.), (pp. 225-233). New York: Springer Science and Business Media.
- Malone, M. E. (2013). The Essentials of Assessment Literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. <https://doi.org/10.1177/0265532213480129>
- Mede, E., and 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.
- Muhammad, F. H. N., Bardakçı, M. (2019). Iraqi EFL Teachers' Assessment Literacy: Perceptions and Practices. *Arab World English Journal*, 10 (2) 420-430. <https://dx.doi.org/10.24093/awej/vol10no2.33>
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research, and Evaluation*, 7(1), 8.

- Mertler, C. A., and Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Research Association, Montreal, Quebec, Canada. Retrieved from <https://eric.ed.gov/?id=ED490355>.
- Newbold, P., Carlson, W. L., & Thorne, B. (2000). *Statistics for Business and Economics* (6th ed.). Prentice Hall.
- Odo, D. M. (2016). An investigation of the development of pre-service teacher assessment Literacy through individualized tutoring and peer debriefing. *Journal of Inquiry & Action in Education*, 7(2), 31-61.
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, 1, 71–85.
- Oller, John W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- O'Sullivan, B. (2012) Assessing speaking. In C. Coombe, P. Davidson, B. O'Sullivan, and S. Stoyanoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 234-246). Cambridge University Press, USA.
- Ölmezer-Öztürk, E. (2018). *Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers*. (Publication No. 524164) [Doctoral thesis, Anadolu University]. Council of Higher Education Thesis Center.
- Ölmezer-Öztürk, E., & Aydin, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia*, 8(1). <https://doi.org/10.1186/s40468-018-0075-2>
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Parsons, J. (1998). Portfolio assessment: Let us proceed with caution. *Adult Learning*, 9 (4), 28-29.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assesment of students. *Mid-Western Educational Researcher*, 6 (1), 21- 27.
- Plano-Clark, Vi. L., & Creswell, J. W. (2015). *Understanding research: A consumer's guide* (2nd ed.). Pearson Education Inc
- Popham, W.J. (2004). Why Assessment Illiteracy Is Professional Suicide. *Educational Leadership*, 62(1), 82-83.

- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Price, M., Rust, C., O'Donovan, B., Handley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. The Oxford Centre for Staff and Learning Development.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13.
- Richard, JC & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. London: Pearson Education.
- Sarıyıldız, S. (2018). *A study into language assessment literacy of preservice English as a foreign language teachers in Turkish context*. (Publication No. 515699) [Master's Thesis, Hacettepe University]. Council of Higher Education Thesis Center.
- Scarino, A. (2013). Language assessment literacy as self-awareness: understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327. <https://doi.org/10.1177/0265532213480128>
- Sevilen, Ç. (2021). *Exploring the Language Assessment Literacy of EFL Instructors and Students' Perceptions about Assessment in English Preparatory Program*. (Publication No. 697053) [Master's Thesis, Bahçeşehir University]. Council of Higher Education Thesis Center.
- Spiller, D. (2012). Assessment Matters: Self-Assessment and Peer Assessment. *The University of Waikato*, 13, 2-18
- Stecher, B. (1998). The local benefits and burdens of large scale portfolio assessment. *Assessment in Education*, 5 (3), 335-351.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765. <https://doi.org/10.1177/003172170208301010>
- Semiz, Ö., Odabaş, K. (2016). Turkish EFL teachers' familiarity of and perceived needs for language testing and assessment literacy. *Proceedings of the 3rd International Linguistics and Language Studies Conference*, 3(1), 66-72.

- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238.
- Tamerer, R. (2019). *An investigation of Turkish pre-service EFL teachers' language assessment literacy*. (Publication No. 566738) [Master's Thesis, Kocaeli University]. Council of Higher Education Thesis Center.
- Tao, N. (2014). *Development and validation of classroom assessment literacy scales: English as a Foreign Language (EFL) teachers in a Cambodian Higher Education Setting*. Unpublished PhD dissertation, Victoria University, Australia.
- Tavşancıl. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayın Dağıtım.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. <https://doi:10.1017/S0267190509090035>
- The American Federation of Teachers, National Council on Measurement in Education, National Education Association (AFT, NCME, NEA). (1990). The standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Trumbull, E., & Lash, A. (2013). Understanding formative assessment: Insights from learning theory and measurement theory. San Francisco: WestEd.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770.
- Webb, N. (2002, April). *Assessment literacy in a standards-based urban education setting*. In annual meeting of the American Educational Research Association, New Orleans.
- Weigle, S. C. (2012). Assessing writing. In C. Coombe, P. Davidson, B. O'Sullivan, and S. Stoyanoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 236- 246). New York: Cambridge University Press.
- Xu, Y., Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Language Testing and Assessment*, 6(1), 133-158.
- Yetkin, C. (2015). *An Investigation on ELT Teacher Candidates' Assessment Literacy*. (Publication No. 405422) [Master's Thesis, Çığ University]. Council of Higher Education Thesis Center.

Yüce, Z. (2015). *Pre-service English language teachers' conceptions of assessment and assessment practices*. (Publication No. 379999) [Master's Thesis, Çağ University]. Council of Higher Education Thesis Center.



APPENDICES

Appendix 1. Language Assessment Knowledge Scale - LAKS

PART I: DEMOGRAPHIC INFORMATION

1. Gender a) male b) female

2. Years of experience

a) 1-5 years b) 6- 10 years c) 11- 15 years d) 16- 20 years e) more than 21 years

3. The BA program you graduated from

a) English Language Teaching (ELT) b) non- ELT

4. Educational background

a) BA degree b) MA degree c) PhD degree

5. Where are you working at now? a) private school b) state school

6. Which grade are you teaching? a) pre-school b) primary school c) middle school d) high school

7. Have you attended any professional development programs/ courses/ training on language assessment? a) yes b) no

8. Do you have any experience in test preparation/development and interpretation?
a) yes b) no

9. How do you evaluate yourself as an assessor in the following areas/subskills?

a) reading (1) very competent (2) competent (3) not very competent (4) not competent

b) listening (1) very competent (2) competent (3) not very competent (4) not competent

c) writing (1) very competent (2) competent (3) not very competent (4) not competent

d) speaking (1) very competent (2) competent (3) not very competent (4) not competent

10. Would you like to take part in online one-on-one interviews later?

a) Yes b) No

If YES; Please write your e-mail address:

PART II: LANGUAGE ASSESSMENT KNOWLEDGE SCALE

ITEMS	True	False	Don't Know
ASSESSING READING			
1. Asking learners to summarize the reading text is a way of assessing their reading skills.	X		
2. When asking several questions about a reading text, all the questions are independent of each other.	X		
3. Cloze test is used for assessing the main idea of the text.		X	
4. In a reading exam, using a text learners have encountered before is not a problem.		X	
5. One reading text is enough to be included in a reading exam.		X	
6. The language of the questions is simpler than the text itself.	X		
7. Errors of spelling are penalized while scoring.		X	
8. Taking vocabulary difficulty into consideration is necessary in assessing reading skills.	X		
9. Including not stated/doesn't say along with true/false items has advantages over true/false items.	X		
10. The more items a reading text is followed, the more reliable it becomes.	X		
11. Using the same words in the correct option as in the text is not a problem.		X	
12. Simplification of reading texts is avoided.		X	
13. Reading texts in a reading exam include various genres (essay, article, etc.).	X		
14. In top-down approach, assessment is on overall comprehension of the reading text.	X		
15. Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.	X		
ASSESSING LISTENING			
16. Using reading texts for listening purposes poses a problem.	X		
17. Including redundancy (e.g. what I mean to say is that) in a listening text poses a problem.		X	

18. Any type of listening text is used for note-taking.		X	
19. Spelling errors are ignored in scoring the dictation.	X		
20. Errors of grammar or spelling are penalized while scoring.		X	
21. A listening cloze test is a way of selective listening.	X		
22. Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative testing.		X	
23. Scoring in note-taking is straightforward.		X	
24. In discrete-point testing, comprehension is at the literal/local level.	X		
25. Using dictation diagnostically in assessing listening skills does not pose a problem.	X		
26. Giving learners a transcript of the listening text is a valid way of assessing listening skills.		X	
27. Dictation is a kind of discrete-point testing.		X	
28. Inference questions based on intelligence are avoided in listening tests.	X		
29. Asking learners to listen to names or numbers is called intensive listening.		X	
30. In selective listening, learners are expected to look for certain information.	X		

ASSESSING WRITING

31. Giving two options to learners and asking them to write about one ensure reliable and valid scoring.		X	
32. Analytic scoring is used to see the strengths and weaknesses of learners.	X		
33. The parts of a scoring scale and the scores in each part do not change for different levels of learners.		X	
34. When there is a disagreement between the scores of the two raters, they score the written work again.		X	
35. Learners are required to write about at least two tasks in the exam rather than one task.	X		
36. Giving restrictive prompts/guidelines to learners for the writing task is avoided.		X	
37. Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.		X	
38. Using visuals which guide learners for writing poses a problem.		X	
39. Holistic scoring is used to see whether the learner is proficient or not at the end of the term.	X		
40. Analytic scoring leads to greater reliability than holistic scoring in writing.	X		
41. In controlled writing, learners have the chance to convey new information.		X	
42. Classroom evaluation of learning in terms of writing is best served through analytic scoring rather than holistic scoring.	X		
43. Irrelevant ideas are ignored in the assessment of initial stages of a written work in process writing.		X	
44. Providing a reading text for writing is a way of assessing writing skills.	X		

45. Mechanical errors (e.g. spelling and punctuation) are dealt with in the assessment of later stages of a written work.	X		
---	---	--	--

ASSESSING SPEAKING

46. When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.		X	
47. Giving learners one task is enough to assess speaking skills.		X	
48. Interlocutors' showing interest by verbal and non-verbal signals poses a problem.		X	
49. When it becomes apparent that the learner cannot reach the criterion level, the task is ended.	X		
50. Using holistic and analytic scales at the same time poses a problem.		X	
51. Reading aloud is a technique used to assess speaking skills.	X		
52. In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.	X		
53. In interactive tasks, more than two learners pose a problem.	X		
54. The interlocutor gives the score when the learner is in the exam room.		X	
55. In a speaking exam, production and comprehension are assessed together.	X		
56. Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills.	X		
57. Discussion among learners is a way of assessing speaking skills.	X		
58. A checklist is a means of scoring oral presentations in in-class assessment.	X		
59. When the focus is to assess discourse, role plays are used.	X		
60. In peer interaction, random matching is avoided.	X		

Appendix 2. Semi-Structured Interviews – Guiding Questions

1. How do you evaluate pre-service education in terms of an LTA course based on your experiences?
 - 1.1 How do you think that your undergraduate “Language Testing and Assessment” course prepared you in your practices?
 - 1.2 Which aspects (theoretical or/and practical) were focused in your courses?
 - 1.3 Did you actively prepare tests and exams, and establish reliability and validity of their productions in pre-service-education?
 - 1.4 What specific aspects of your Language and testing assessment course do you convey to your current teaching and testing practices?
2. According to "Language Knowledge Assessment Scale" used in this study, language assessment knowledge level of EFL teachers was identified as 31,59 out of 60. How do you evaluate this situation? What might be the underlying reasons of this situation?
3. There are four sections in the scale, assessing reading, listening, writing and speaking, each consisting of 15 questions. In terms of assessing the skills, the highest knowledge level was found in assessing reading (9.58 out of 15) whereas the lowest level was in assessing listening (6.82 out of 15). The knowledge level in assessing other skills was found as 8.31 in speaking and 6.88 in writing. How do you evaluate this situation? What are the possible reasons of this?
4. In the study, whether language assessment knowledge of the teachers changed according to certain background variables that are years of experience, the BA program being graduated(ELT or non-ELT), working at a private or state school, experience in test preparation and attending trainings on testing and assessment was investigated, and it was seen that **none of them had an influence on their knowledge**. How do you evaluate this?
5. The relationship between the participants’ perceived self-competency and their actual knowledge level was searched, and it was seen that most of them perceived themselves as competent or very competent although their actual score was 31.59 out of 60. How do you evaluate this difference? What can be the potential reasons of it?
6. What kind of an in-service / pre-service training module do you think will meet teachers’ needs in LTA?