



**AN INVESTIGATION OF EFL TEACHERS' LANGUAGE
ASSESSMENT KNOWLEDGE AND PRACTICE: THE CASE OF K-12
SCHOOLS IN TURKEY**



Gamze Satgun

**MASTER THESIS
ENGLISH LANGUAGE TEACHING PROGRAM**

**GAZI UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES**

MAY 2023

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 2 (iki) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Gamze

Soyadı : Satgun

Bölümü : İngiliz Dili Eğitimi

İmza :

Teslim Tarihi :

TEZİN

Türkçe Adı: İngilizce Öğretmenlerinin Dilde Ölçme Değerlendirme Bilgilerinin ve Uygulamalarının İncelenmesi: Türkiye’deki K-12 Okullarının Durumu

İngilizce Adı: An Investigation of EFL Teachers’ Language Assessment Knowledge and Practice: The Case of K-12 Schools in Turkey

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazarın Adı Soyadı: Gamze Satgun

İmza:

JÜRİ ONAY SAYFASI

Gamze Satgun tarafından hazırlanan “An Investigation of EFL Teachers’ Language Assessment Knowledge and Practice: The Case of K-12 Schools in Turkey” adlı tez çalışması aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Gazi Üniversitesi Yabancı Diller Eğitimi Ana bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Asuman AŞIK

İngiliz Dili Eğitimi Ana Bilim Dalı, Gazi Üniversitesi

Üye: Doç. Dr. Burçak YILMAZ YAKIŞIK

İngiliz Dili Eğitimi Ana bilim Dalı, Gazi Üniversitesi

Üye: Doç. Dr. Aydan IRGATOĞLU

Yabancı Diller Ana bilim Dalı, Hacı Bayram Veli Üniversitesi

Tez Savunma Tarihi: 27/03/2023

Bu tezin İngiliz Dili Eğitimi Ana Bilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban ÇETİN

Eğitim Bilimleri Enstitüsü Müdürü



Dedicated to my mother

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my supervisor Assoc. Prof. Asuman AŞIK for her professional guidance, gracious thoughts, and encouragement throughout every process of this research. This thesis wouldn't have been completed without her constructive feedback and inspiring statements.

I also would like to express my appreciation to my dearest friends Ayşenur Ağaoğlu, Talha Altuntaş, Özge Pelin Nazlı Turgut, Aslıgül Köyük, who have been a great support in this academic journey.

Lastly, my deepest gratitude and appreciation to my family. My sisters Özge Satgun Sinir and Gaye Satgun, my father Şerafettin Satgun, and finally my lovely mother Kıymet Satgun, who have never stopped supporting me throughout this journey. I am always indebted for their endless affection.

**İNGİLİZCE ÖĞRETMENLERİNİN DİLDE ÖLÇME
DEĞERLENDİRME BİLGİLERİNİN VE UYGULAMALARININ
İNCELENMESİ: TÜRKİYE’DEKİ K-12 OKULLARININ DURUMU
(Yüksek Lisans Tezi)**

Gamze Satgun
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Mayıs 2023

ÖZ

Bu çalışma Türkiye’de K-12 bağlamındaki okullarda çalışmakta olan İngilizce öğretmenlerinin yabancı dilde ölçme değerlendirme bilgi düzeylerinin ve uygulamalarının incelenmesini amaçlamaktadır. Ayrıca öğretmenlerin bilgi düzeylerinin değerlendirme uygulamalarına ne derece yansıdığı ve öğretmenlerin değerlendirme uygulamalarında karşılaştıkları problemlerin belirlenmesi bu araştırmanın amaçları arasındadır. Araştırma açıklayıcı sıralı karma araştırma desenine göre tasarlanmıştır. Araştırmanın nicel verileri Ölmezer-Öztürk ve Aydın (2018) tarafından tasarlanan Dilde Ölçme Değerlendirme Bilgi Ölçeği (LAKS) ile ilkökul, ortaokul ve lise düzeyinde çalışmakta olan 202 İngilizce öğretmeninden elde edilmiştir. Araştırmanın nitel verileri ise yarı yapılandırılmış görüşmelerle 17 İngilizce öğretmeninden toplanmıştır. Yapılan analizler öğretmenlerin dilde ölçme değerlendirme bilgi düzeylerinin genel olarak anlamlı derecede yüksek olduğunu gösterirken okuma ve konuşma becerilerinin değerlendirilmesinde yüksek, yazma ve dinleme becerilerinin değerlendirilmesinde ise düşük olduğu ortaya çıkmıştır. Öğretmenlerin bilgi düzeylerinin demografik özelliklerine göre değişiklik göstermediği görülmüştür. Buna ek olarak, öğretmenlerin yazma ve dinleme becerilerine dair öz yeterlik algıları ve gerçek bilgi düzeyleri arasında tutarsızlık görülmüştür. Çalışmanın nitel verilerinin analizinde içerik analizi yöntemi kullanılmış ve yapılan analizler sonucunda öğretmenlerin değerlendirmeyi süreç içerisinde sürekli olarak ve süreç sonunda öğrenme derecesini ve eksikliklerini belirlemek; ayrıca sınıf uygulamalarını bu yönde düzenlemek amacıyla kullandıkları anlaşılmıştır. Ders içerisinde en çok okuma, dinleme, konuşma, dil bilgisi, kelime ve yazma becerilerinin değerlendirildiğini belirtmelerine rağmen konuşma ve dinleme becerilerinin kimi zaman göz ardı edildiği anlaşılmıştır. Ayrıca tüm öğretmenlerin

yazılı sınavlar gibi geleneksel yöntemleri benimsediği görülse de gözlem, portfolyo ve performans değerlendirmesi gibi alternatif yöntemleri de benimsedikleri anlaşılmıştır. Öğretmenlerin değerlendirme sonuçlarını sınıf içi uygulamalarını uyarlamak ve öğrencilere dönüt vermek amacıyla kullandığı ve genellikle yazılı sınav sonuçlarına hızlı dönüt verirken sınıf içinde benimsedikleri değerlendirme uygulamalarına ise tüm sınıfa yönelik sözlü dönüt verdikleri ortaya çıkmıştır. Öğretmenler değerlendirme uygulamalarını planlarken en çok müfredatı ve hedef kazanımları baz alan öğretmenler sıklıkla meslektaşlarından da yardım almıştır. Ayrıca planlama sürecinde ölçme değerlendirme ilkelerini ve öğrenci ihtiyaçlarını göz önünde bulunduran öğretmenlerin öğrencileri planlama sürecine aktif olarak dahil etmediği anlaşılmıştır. Değerlendirme alanında verilen hizmet içi eğitimlerin hem nicelik hem de nitelik açısından yetersiz olduğunu belirten öğretmenler farklı alanlarda eğitime ihtiyaç duymuştur. Ayrıca öğretmenlerin bilgi düzeylerinin değerlendirme uygulamalarına belirli bir dereceye kadar yansıdığı görülmüş; öğretmenlerin karşılaştıkları problemlerin yetersiz ders saatleri, sınav odaklı eğitim sisteminin olumsuz etkisi ve kalabalık sınıf ortamları olduğu anlaşılmıştır. Sonuçlara dayanarak öğretmenlere ve karar alıcılara önerilerde bulunulmuştur.

Anahtar Kelimeler : Dilde Ölçme Değerlendirme Bilgisi, Dilde Ölçme Değerlendirme Uygulamaları, Dilde Ölçme Değerlendirme Okuryazarlığı
Sayfa Adedi : xviii + 193
Danışman : Doç. Dr. Asuman Aşık

**AN INVESTIGATION OF EFL TEACHERS' LANGUAGE
ASSESSMENT KNOWLEDGE AND PRACTICE: THE CASE OF K-12
SCHOOLS IN TURKEY
(M.A. Thesis)**

**Gamze Satgun
GAZI UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
May 2023**

ABSTRACT

This study aims to investigate the language assessment knowledge level and practices of the EFL teachers in the K-12 context in Turkey. In addition, to what extent their knowledge is reflected on their assessment practices and the problems they encounter with regarding language assessment are aimed to be examined. Explanatory sequential mixed methods design is adopted in the study. The quantitative data is collected with the Language Assessment Knowledge Scale (LAKS), developed by Ölmezer-Öztürk and Aydın (2018), from 202 EFL teachers working at primary, middle, and high school levels. Semi-structured interviews are conducted with 17 EFL teachers for the qualitative data collection. The results of the analyses revealed that the teachers' LAK in general was significantly high. In addition, the teachers' knowledge level in assessing reading and speaking was significantly high while they had significantly lower levels of knowledge in assessing writing and listening. The impact of the demographic factors on LAK was not observed. Additionally, a discrepancy was observed between the teachers' perceived self-competency and their actual knowledge level in assessing writing and listening skills. Based on the results of the content analysis of the qualitative data, teachers assessed their learners for both formative and summative purposes to determine the degree of learning and learners' weaknesses in addition to adjusting their classroom practices accordingly. Even though teachers expressed that assessed reading, listening, speaking, grammar, vocabulary, and writing skills typically, it was revealed that listening and speaking assessment can be ignored in the process. Besides, all the teachers adopted traditional methods such as written exams in addition to the alternative methods such as classroom observations, portfolio and performance assessment. It was demonstrated that the teachers made use of assessment results for planning and

adjusting their classroom practices and giving feedback to learners and provided feedback quickly to learners' performances in written exams while for the classroom activities they adopted for assessment, they gave feedback to the whole class in the form of oral feedback. Teachers used curriculum and course objectives as the basis for their assessment plans in addition to consulting their colleagues. Furthermore, even though taking the assessment principles and learners' needs into consideration while planning assessment, learners were not actively involved in the planning process. Teachers needed trainings in various areas of language assessment since both the quantity and quality of the in-service trainings were inadequate. Moreover, it was revealed that teachers' LAK was reflected on their assessment practices to a certain degree. Finally, the problems that teachers face regarding language assessment was determined to be insufficient teaching hours, the negative impact of exam-oriented education policy, and large class sizes. Based on the results, pedagogical implications are made for the teachers and policy makers.



Key Words : Language Assessment Knowledge, Language Assessment Practices,
Language Assessment Literacy
Page Number : xviii + 193
Supervisor : Assoc. Prof. Asuman AŞIK

TABLE OF CONTENTS

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
JÜRİ ONAY SAYFASI.....	iii
ACKNOWLEDGEMENT	v
ÖZ	vi
ABSTRACT	viii
TABLE OF CONTENTS.....	x
LIST OF TABLES	xv
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS.....	xviii
CHAPTER I.....	1
INTRODUCTION.....	1
1.1. Background of the Study	1
1.2. Statement of the Problem	5
1.3. Aim of the Study.....	6
1.4. Significance of the Study	7
1.5. Limitations of the Study	9
1.6. Assumptions of the Study	9
1.7. Definition of the Terms.....	9

CHAPTER II	11
LITERATURE REVIEW	11
2.1. Testing, Assessment, and Evaluation	11
2.1.1. Test, Testing, and Measurement	11
2.1.2. Assessment	12
2.1.3. Evaluation	13
2.2. Purposes of Assessment	14
2.3. Types of Assessment.....	15
2.3.1. Formative Assessment and Summative Assessment	16
2.3.2. Assessment for Learning (AfL), Assessment of Learning (AoL) and Assessment as Learning (AaL).....	17
2.3.3. Formal Assessment and Informal Assessment	18
2.3.4. Norm-Referenced Assessment and Criterion-Referenced Assessment ..	18
2.3.5. Objective Assessment and Subjective Assessment	19
2.3.6. Direct Assessment and Indirect Assessment.....	19
2.3.7. Discrete Point Assessment and Integrative Assessment	20
2.3.8. Traditional Assessment and Alternative Assessment	21
2.3.8.1. <i>Performance-Based Assessment</i>	22
2.3.8.2. <i>Portfolio Assessment</i>	22
2.3.8.3. <i>Self-Assessment and Peer-Assessment</i>	23
2.3.9. Dynamic Assessment	24
2.4. Principles of Language Assessment.....	24
2.4.1. Reliability	25
2.4.1.1. <i>Student-Related Reliability</i>	25
2.4.1.2. <i>Rater-Reliability</i>	25
2.4.1.3. <i>Test Administration Reliability</i>	26

2.4.1.4. <i>Test Reliability</i>	26
2.4.1.5. <i>Test/Re-test Method</i>	27
2.4.1.6. <i>Parallel Forms Method</i>	27
2.4.1.7. <i>The Split-Half Method</i>	28
2.4.1.8. <i>Kuder-Richardson Formulas</i>	28
2.4.2. Validity	28
2.4.2.1. <i>Criterion-Related Validity</i>	29
2.4.2.2. <i>Content Validity</i>	30
2.4.2.3. <i>Construct Validity</i>	30
2.4.2.4. <i>Face Validity</i>	31
2.4.3. Practicality	31
2.4.4. Authenticity	32
2.4.5. Washback	33
2.5. Assessment of Language Skills	34
2.5.1. Assessing Reading	34
2.5.2. Assessing Listening	36
2.5.3. Assessing Speaking	38
2.5.4. Assessing Writing	40
2.5.5. Assessing Young Language Learners	43
2.6. Assessment Literacy	46
2.7. Language Assessment Literacy	49
2.7.1. Definitions of LAL	50
2.7.2. LAL Models-Components of LAL	52
2.7.3. The Importance of LAL	56
2.7.4. Language Assessment Knowledge and Practice	57

2.8. Studies on Language Assessment Literacy/Knowledge and Practices in the World.....	58
2.9. Studies on Language Assessment Literacy/Knowledge and Practices in Turkey	64
CHAPTER III.....	73
METHODOLOGY.....	73
3.1. Research Design	73
3.2. Research Context	74
3.3. Participants.....	75
3.4. Data Collection Tools.....	79
3.4.1. Language Assessment Knowledge Scale (LAKS).....	79
3.4.2. Semi-Structured Interviews	80
3.5. Data Collection Process	81
3.6. Data Analysis	82
3.6.1. Quantitative Data Analysis.....	82
3.6.2. Qualitative Data Analysis	83
CHAPTER IV	85
FINDINGS AND DISCUSSION	85
4.1. EFL Teachers' General and Skill-based Language Assessment Knowledge Level	86
4.1.1. Assessing Reading	91
4.1.2. Assessing Listening.....	97
4.1.3. Assessing Writing	102
4.1.4. Assessing Speaking.....	108
4.2. Effects of Demographic Features on EFL Teachers' LAK Level.....	113
4.3. EFL Teachers' Perceived Self-competency and Actual LAK Level.....	118

4.4. Classroom Practices of EFL Teachers	120
4.4.1. Characteristics of Language Assessment	121
4.4.2. Assessment of Language Skills.....	127
4.4.3. Assessment Results.....	135
4.4.4. Planning Assessment.....	138
4.4.5. Professional Development.....	144
4.5. Reflection of LAK on EFL' Teachers Classroom Practices	148
4.6. Problems of EFL Teachers in the K-12 Context Regarding Language Assessment	151
CHAPTER V.....	155
CONCLUSION AND IMPLICATIONS.....	155
5.1. Conclusion.....	155
5.2. Pedagogical Implications	163
5.3. Suggestion for Further Research	165
REFERENCES	167
APPENDIX	183
Appendix A. Ethics Committee Approval	184
Appendix B. Permission from the Ministry of National Education	186
Appendix C. Permission E-mail from the Authors of LAKS.....	187
Appendix D. Language Assessment Knowledge Scale (LAKS)	188
Appendix E. Interview Consent Form	191
Appendix F. Semi-structured Interview Questions.....	192

LIST OF TABLES

Table 1. <i>Demographic Features of the Participants in the LAKS</i>	76
Table 2. <i>Demographic Features of the Participants in the Interview</i>	78
Table 3. <i>Statistical Analysis Procedures</i>	84
Table 4. <i>One Sample T-test Results for General LAK Level</i>	86
Table 5. <i>One Sample T-test Results for Skill-based LAK Level</i>	87
Table 6. <i>The Correlation Among the Skill-based LAK</i>	88
Table 7. <i>Descriptive Statistics of Language Assessment Knowledge Level of EFL Teachers in the K-12 Context</i>	88
Table 8. <i>Independent Samples T-test Results According to Gender and LAK</i>	113
Table 9. <i>One-way ANOVA Results According to the Years of Experience and LAK</i>	114
Table 10. <i>Independent Samples T-test Results According to BA Program and LAK</i>	114
Table 11. <i>One-way ANOVA Results According to Educational Background and LAK</i>	115
Table 12. <i>Independent Samples T-test Results According to the Type of the Institution and LAK</i>	115
Table 13. <i>One-way ANOVA Results According to the Educational Level of the Workplace and LAK</i>	116
Table 14. <i>One-way ANOVA Results According to the Geographical Region and LAK</i>	116
Table 15. <i>Independent Samples T-test Results According to Taking Separate Testing Course in BA</i>	117
Table 16. <i>Independent Samples T-test Results According to Attending Additional Trainings in Language Testing and Assessment</i>	117

Table 17. <i>One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Reading</i>	118
Table 18. <i>One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Listening</i>	119
Table 19. <i>One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Writing</i>	119
Table 20. <i>One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Speaking</i>	120
Table 21. <i>Teachers' Views on the Characteristics of Language Assessment</i>	121
Table 22. <i>Teachers' Classroom Practices Regarding the Assessment of Language Skills</i>	127
Table 23. <i>Teachers' Use of Assessment Results</i>	135
Table 24. <i>Teachers' Decision-making Process for Planning Assessment</i>	138
Table 25. <i>Teachers' Professional Development</i>	144
Table 26. <i>Teachers' Language Assessment Related Problems</i>	151

LIST OF FIGURES

<i>Figure 1. Research design of the study</i>	74
---	----



LIST OF ABBREVIATIONS

AL	Assessment Literacy
EFL	English as a Foreign Language
ELT	English Language Teaching
LAK	Language Assessment Knowledge
LAKS	Language Assessment Knowledge Scale
LAL	Language Assessment Literacy
MoNE	Ministry of National Education
SPSS	Statistical Package for Social Sciences

CHAPTER I

INTRODUCTION

1.1. Background of the Study

The purpose of assessment is to improve learners' development and learning by means of systematically gathering information about their knowledge, abilities, and understanding (Coombe, 2018). Similarly, the main purpose of assessment should be to illuminate greater instruction and lead to more fruitful learning (Fulcher & Davidson, 2007) by "collecting information about a given object of interest according to the procedures that are systematic and substantively grounded" (Bachman, 2004, p. 7). Thus, assessment and teaching are believed to have had an integral relation since the very beginning because any kind of teaching is generally followed by some kind of testing (Farhady, 2018). Wilam (2013) explicitly states that assessment has an indisputable function in teaching and learning process since it works as a bridge between teaching and learning to illuminate whether teaching is effective. However, he also discusses that not all teaching leads to learning no matter how thorough we plan and apply it in our classroom practices, and that is why assessment is needed in the first place. Therefore, it can be uttered that assessment informs teachers about their instructional activities if they lead to intended outcomes or not. In other words, assessment is an essential part regarding the process of teaching and learning by virtue of its effect on teachers' instructional choices and learning process. It guides teachers to make accurate adjustments according to the information provided by their assessment practices in order to enhance learning process and improve students' achievement.

Inbar-Lourie (2008) explains that the term 'assessment' has undergone some changes both semantically and theoretically, which indicates the relation between teaching and assessment in that the advancements in testing and assessment concepts have emerged at the same time with the altering approaches in teaching (Farhady, 2018). It is now considered as the

umbrella term that embodies methods and approaches of testing and evaluation. With the changes in teaching and learning theories from behavioristic to more communicative, approaches to testing and assessment have evolved as well from focusing on discrete linguistic elements such as grammar and phonology and geared towards being more integrative and communicative by means of authenticity and simulating the real-world (Brown, 2003). In addition, Inbar-Lourie (2008) depicts that this theoretical shift from behaviouristic approaches to more social-constructivist ones in the language education field has also affected the approaches to assessment to evolve from testing cultures, which is mostly based on the numerical data of learner performance, to assessment cultures, which sees language as a social practice and improves learner engagement in the assessment process as with the focus on both product and process. However, a discrepancy may occur when teachers pursue both cultures at the same time though they are incompatible (Inbar-Lourie, 2008).

Fulcher (2012) illustrates that language teachers' responsibilities regarding testing and assessment have increased substantially at the beginning of the 21st century for three main reasons, which are both external and internal to the field. The first reason is the increasing number of test and assessment use with the aim of accountability. The second reason is the consideration of testing and assessment as a part of national immigrant policy, such as citizenship tests. Both are regarded as external to the field. The last and only reason that is internal to the field is the assessment for learning perspective that has been debated over its importance of integrating it in the classroom practice for its impact on enhancing learning (Black & Wiliam, 1998). In a similar vein, for language teaching to be effective, assessment should be embedded in the instruction (Coombe, Troudi, & Al-Hamly, 2012). Thus, it is critical for teachers to be aware of the various types and purposes of assessment and adopt them in their classroom practices in consideration of their context. To be able to perform sound assessment practices, it is clear that some sort of training in assessment is required for language teachers "if they are engaged in selecting, administering, interpreting, and sharing results of large-scale tests produced by professional testing organizations, or in developing, scoring, interpreting, and improving classroom-based assessments" (Taylor, 2009, p. 24).

The awareness about assessment and its importance has raised significantly throughout the years in that there have been many professional organizations and associations throughout the world, both national and international, that attempted to set the standards of assessment for teachers. The most known one is the 'Standards for teacher competence in educational

assessment of students (1990)' by American Federation of Teachers, National Council on Measurement in Education, and National Education Association. In addition, Common European Framework of Reference (CEFR) also contains the concept and implication of assessment for language education by making the biggest impact across Europe. McNamara and Roever (2006) believes that CEFR is “the most comprehensive example of policy-driven assessment yet seen” (p. 212). In Turkish context, there is a list of qualifications that the teachers of English language need to have determined by the Ministry of National Education (MoNE) (2008): “The Specialized Field Competencies of English Language Teachers”, in which language assessment is presented as a competency that teachers should be equipped with among four others. According to MoNE (2008), language assessment competencies contain four dimensions. Firstly, teachers should be able to specify the purposes of testing and assessment methods and techniques regarding language assessment and aim for a systematic assessment procedure that involves both formative and summative assessment methods in consideration of the educational program and individual differences of the learners. In addition, they should be able to design and implement various assessment tools and techniques for assessing four language skills with respect to assessment principles such as practicality, validity, and reliability and share their knowledge and experience with their colleagues. English language teachers should also be able to make interpretations based on assessment results and provide feedback to the learners and their parents in addition to evaluating their own instruction and assessment practices. As for the final competence, it is critical that teachers utilize assessment results and adjust their classroom practices accordingly. In addition, a teacher guidance manual has been published recently in Turkey by the department of General Directorate of Testing, Assessment and Test Services of MoNE, which is the “Strengthening teacher capacity based on school- and classroom-based assessment: English lesson teacher’s guide booklet” (MoNE, 2020). Since it is indicated in several studies that teachers do not feel competent in classroom-based assessment even though they spend more than one third of the classroom time for assessment-related activities, it has become a necessity that educational programs are adjusted according to the latest advancement in approaches to teaching and assessment (MoNE, 2020). In parallel with these objectives, the manual, embracing a formative approach to assessment, involves the fundamental concepts of assessment and assessment techniques and tools that can be used for formative assessment, sample assessment activities that are designed in association with the current curriculum, and the aspects that should be taken into consideration while communicating with other stakeholders in the formative assessment process. From the

growing rate of preparing guidelines or standards of assessment quality for teachers, one can infer that the teachers are the most prominent agents of assessment. Similarly, Vogt and Tsagari (2014) agree that the teachers engage in several assessment procedures daily; therefore, they are an important element in the process since assessment takes almost half of the teachers' time (Plake, 1993; Stiggins, 1999). The responsibilities of the teachers do not only involve teaching, but also assessing. Thus, teachers do not only transfer knowledge to learners, but they also work as an assessor to monitor or judge their students' learning process. Having the dual role of teaching and assessing, teachers should acknowledge the significance of assessment procedures and act accordingly (Rea-Dickins, 2004). The quality of assessment mostly relies on teachers' shoulders. Therefore, teachers are to be competent about the requirements of good assessing implementations. Plake (1993) indicates that assessment competencies are crucial to teaching profession as they use assessment data gathered from their students to make decisions about their instructions and evaluations about their students. As a result of the above-mentioned discussions about assessment and its influence on everyday classroom practice, teachers need to possess a degree of knowledge on assessment and be familiar with assessment-related procedures, which is referred to as Assessment Literacy (AL).

Assessment literacy is the understanding of basic assessment-related concepts and procedures that may possibly have an impact on educational decisions (Popham, 2011). Deriving from AL in general education, Language Assessment Literacy (LAL) involves many of the assessment literacy features with additional competencies that are specific to language assessment. Adopting a process-based approach to developing LAL, Scarino (2013) sees the goal and criterion as the adjustments that teachers make in how they conceptualize and interpret assessment, and how assessment shape their judgements and decisions by critically observing and evaluating their own assessment procedures. In order to make appropriate modifications in their frameworks, teachers need to be aware of the potential of these frameworks to degrade or promote their perspectives. That is to say, LAL is the process of critical evaluation of teachers' own assessment procedures by bridging the gap between their previous conceptions and diverse theoretical knowledge existent in the field. Coombe et al. (2012) address the significance of LAL by emphasizing that language educators to have assessment literacy is a necessity in today's world where educational organizations devote a great importance to assessment.

1.2. Statement of the Problem

Assessment can be a powerful instrument for learning when used properly (Popham, 2011). It helps teachers to determine the strengths and weaknesses about their instructional choices (Rea-Dickins, 2004) by adjusting them according to the outcomes of the assessment procedures provided with the feedback. Herrera and Macías (2015) explain the value of sound assessment practices as an inspiration for teachers to monitor the effectiveness of a certain teaching method, the pace of the course or the convenience of the content. However, Plake (1993) does not believe that teachers are adequately equipped to carry out the function of assessment at a professional level and she claims that the reason is deficient training in pre-service programs. Likewise, Popham (2011) argues that teachers need more education on assessment due to their lack of knowledge on the matter of assessment or as Stiggins (1991, p. 535) utters, “we are a nation of assessment illiterates”. Additionally, Popham (2011) suggests that teachers, to be able to perform properly, need to become *assessment literate* (p. 267) and assessment literacy of teachers needs to be valued as much as instructional proficiency. Since language teaching is another domain in education, it requires teachers to be equipped with additional competencies that are subject specific, which is called language assessment literacy (Inbar-Lourie, 2008). LAL can be defined as the familiarity of assessment-related procedures such as designing, developing, grading, and evaluating the assessment data and the reflection of this knowledge, skills, and assessment principles on classroom practices guided by the current theoretical conceptions (Malone, 2013; Taylor, 2009; Vogt & Tsagari, 2014). The knowledge base of LAL is considered as a prerequisite for good assessment practices (Xu & Brown, 2016). In other words, the knowledge base is the criteria to make appropriate judgements on assessment practices since it would be hardly possible to evaluate assessment practices properly (Fulcher, 2012). Scarino (2013) argues that even though knowledge base of LAL is indispensable, the sufficiency of the basis for actual development of knowledge and understanding in assessment for language teachers is questionable. Thus, teachers should be aware of the theoretical concepts in different approaches and reflect their knowledge to their teaching practices relative to their context so that they can enhance student learning and evaluate their own instruction and assessment practices.

Foreign language education in Turkey have always been a subject of discussion besides testing assessment practices to be summative and test-based with an emphasis on discrete points such as grammar and vocabulary because of the high-stakes exams students have to

take for placement purposes (Hatipoğlu, 2016). In other words, the impact of these standardized tests that only contain multiple-choice questions regulates the instructional methods as well as assessment practices of the teachers. Similarly, even though EFL teachers have received a training on testing and assessment in pre-service education, they still do not feel competent enough to assess their learners' language skills except for grammar and vocabulary and state that they need further training (Mede & Atay, 2017). The reason behind is related to the lack of trainings and the quality of pre- and in-service education regarding language assessment in that they are predominantly theory-based while disregarding the practical concerns (Ölmezer-Öztürk & Aydın, 2019b). As Inbar-Lourie (2013b, p. 5) puts forth, "language assessment constitutes a body of knowledge comprising both theory and practice"; therefore, to be able to suggest that teachers have high level of LAL, they should put the theoretical knowledge into their practices. However, Öz and Atay (2017) inform that teachers were not able to reflect their knowledge into their classroom practices though they express that they possess a degree of knowledge in assessment. In consideration of these discussions on the significance of LAL, it is important that LAL in terms of language assessment knowledge and the extent of assessment practices EFL teachers is investigated so as to determine the quality of language assessment and enhance the foreign language education in the K-12 context in Turkey. In conclusion, this study addresses the LAL of EFL teachers in Turkey by means of investigating their knowledge level of language assessment and its reflection on classroom practices in K-12 education.

1.3. Aim of the Study

The aim of the study is to investigate language assessment literacy of EFL teachers by means of analyzing their language assessment knowledge (LAK) level in assessing listening, reading, writing, and speaking skills and their language assessment practices in the context of K-12 education in Turkey. Another objective of the study is to discover whether demographic features of the teachers such as experience, workplace, educational background, and professional development have an impact on their LAK level. Finally, after analyzing teachers' language assessment practices, it is aimed to find out the problems that EFL teachers encounter with regarding their assessment practices in the context of primary and secondary education in Turkey. In line with these objectives, this study seeks to answer these research questions:

1. What is the language assessment knowledge (LAK) level of EFL teachers in the K-12 context in Turkey?
2. Does EFL teachers' level of LAK change in terms of;
 - a. gender,
 - b. years of experience,
 - c. the BA program being graduated,
 - d. educational background,
 - e. type of the institution they work at,
 - f. educational level of the workplace,
 - g. geographical region they work in,
 - h. taking a separate testing and assessment course in BA,
 - i. attendance to additional training in testing and assessment?
3. Does EFL teachers' perceived self-competency in assessing each language skill have an effect on their LAK level?
4. What are the language assessment practices of EFL teachers in the K-12 context in Turkey?
5. Is EFL teachers' LAK reflected on their classroom practices?
6. What are the problems that EFL teachers encounter with in the K-12 context in Turkey regarding language assessment?

1.4. Significance of the Study

Even though there are many studies regarding LAL throughout the world, “research into assessment literacy is in its infancy” (Fulcher, 2012, p. 117); therefore, more research is required to illuminate the field of LAL. Since the knowledge base of LAL is a dynamic facet, recent research findings can be utilized for the enhancement of teachers' language assessment knowledge (Coombe, Vafadar, & Mohebbi, 2020). Besides, the role of the context regarding LAL is undeniable. However, in some studies, it is assumed that the conceptualization of LAL applies to all educational levels, such as primary, secondary, and tertiary, neglecting the specific needs and aims of that particular context (Tsagari, 2021).

Thus, a distinctive discussion of LAL in various educational levels are required in order to offer a contribution to LAL literature. Furthermore, Lam (2014) illustrates that LAL research is generally concerned with language instructors at tertiary level or language testers; thus, future research needs to consider classroom-based assessment of in-service language teachers at primary and secondary levels and its impact on LAL development. Likewise, studies on LAL in Turkey focus on the higher education context as well (Mede & Atay, 2017; Ölmezer-Öztürk & Aydın, 2019a; Yastıbaş & Takkaç, 2018). In the meantime, Önalın and Gürsoy (2020) highlight that LAL studies focusing on primary and secondary educational context remain neglected. For that matter, this study seeks to contribute to the field by examining the of EFL teachers' LAK level in assessing four skills; reading, writing, speaking, and listening, and their assessment practices that they implement in the classroom in the context of K-12 education in Turkey.

Furthermore, in their comprehensive review of the studies on language assessment literacy, Coombe, Vafadar, and Mohebbi (2020) suggest that more research is needed in different contexts in order to get better insights about how assessment policies and standards can be implemented. However, even though the impact of contextual factors such as assessment culture and policies, physical condition of the classrooms is believed to shape classroom practices, the data still need to be expanded on this matter (Tuzcu-Eken, 2016; Vogt, Tsagari, & Csépes, 2020). This study is expected to contribute to the LAL literature by providing an understanding of EFL teachers' LAK and practice as well as the problems they experience concerning language assessment in the K-12 context in Turkey.

Reviewing the research in LAL, Tsagari (2021) asserts that there are gaps in the LAL literature. First, a gap between theory and practice of LAL has been observed, yet not explicitly revealed; therefore, components and practices of LAL need to be the priority for future research so that actual assessment practices could be incorporated rather than solely being conceptual. In a similar vein, Sevımel-Şahin and Subaşı (2019) propose that the practical aspects of LAL need to be emphasized since it is apparent that even though EFL teachers have indicated language assessment knowledge to some degree, they are still incapable of putting their knowledge into practice. Examining the studies concerning the education on assessment in pre-service trainings, Xu and Brown (2016) found out that assessment courses are criticized for being theory-based and lacking practical representations even though various studies provide information on fundamental components for more qualified assessment courses and urge to incorporate meticulously

designed content that are professionally standardized into assessment courses that allow teachers to combine theory and practice. Besides, the data concerning the reflection of LAL on classroom practices is scarce to make accurate interpretations (Öz & Atay, 2017). Therefore, this study is believed to make a valuable contribution to the LAL literature by investigating EFL teachers' language assessment knowledge and practices in order to get a deeper understanding of the gap between theory and practice.

1.5. Limitations of the Study

This study has certain limitations:

1. This study is limited to the EFL teachers who work at the primary, secondary, and high schools that are under the Ministry of National Education during the educational year of 2021-2022.
2. The data and interpretations on LAL of EFL teachers are limited to their knowledge base and classroom practices.
3. EFL teachers' language assessment knowledge is limited to data obtained from the Language Assessment Knowledge Scale (LAKS).
4. This study is limited to exploring the language assessment practices of EFL teachers by means of their self-reported answers in the semi-structured interviews.

1.6. Assumptions of the Study

In this study;

1. It is assumed that the participants have sincerely answered the questions in both Language Assessment Knowledge Scale and semi-structured interviews.
2. It is assumed that the participants' self-reported language assessment practices reflect their actual language assessment practices in the classroom.

1.7. Definition of the Terms

Testing: It is “a method of measuring a person's ability, knowledge or performance in a given domain” (Brown, 2003, p. 3).

Assessment: Assessment, a dynamic process, is an umbrella term that covers testing and assessment methods (Brown, 2003) and defined as “the process of gathering, interpreting, recording and using information about pupils’ responses to educational tasks” (Lambert & Lines, 2001, p. 4).

Assessment Literacy (AL): It “consist of an individual’s understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” (Popham, 2011, p. 267).

Language Assessment: “the systematic process of evaluating and measuring collected data and information on students’ language knowledge, understanding, and ability in order to improve their language learning and development” (Coombe, 2018, p. 10).

Language Assessment Literacy (LAL): an individual’s familiarity, understanding, awareness, knowledge, skills, abilities as well as principles and practices about language assessment-related procedures in line with their specific contexts and the current theoretical concerns of foreign language education within broader social frameworks including the skills to project their theoretical knowledge in classroom practices.

Language Assessment Knowledge (LAK): an individual’s knowledge of pedagogical content and language, purposes, methods, and principles of assessment, grading and scoring, feedback, interpreting and communicating the assessment results, student involvement, and ethical concerns (Xu & Brown, 2016) regarding the assessment of reading, writing, speaking, and listening skills.

CHAPTER II

LITERATURE REVIEW

In this chapter, language testing and assessment will be discussed in terms of its purposes, the types of assessment, assessment principles, assessment of language skills and young learners in addition to the assessment literacy, language assessment literacy by means of its conceptualization, definitions, components. Finally, language assessment knowledge and practice will be reviewed. The related studies in Turkey and in the world regarding language assessment literacy, knowledge, and practice will be discussed in the light of the literature.

2.1. Testing, Assessment, and Evaluation

Testing, assessment, and evaluation are among the most fundamental elements, if not central, in education since everything that is taught may not result in learning; thus, these components can be considered as the requisites to improve teaching (William, 2013). However, these components can sometimes be confusing for some people, thereby, used as synonyms though they do not carry the same meaning, which eventually may lead to more confusion. Thus, it is crucial to clarify the meanings and differentiation of testing, assessment, and evaluation at the very beginning.

2.1.1. Test, Testing, and Measurement

Testing necessitates using both tests and measurement in terms of figuring the extent of achievement in learners' performance, knowledge, or skills. Coombe (2018) illustrates that a test is "a set of tasks or activities intended to elicit the samples of performance which can be marked or evaluated to provide feedback on a test takers ability or knowledge" and testing is the "process of seeing how much we match up to a standard or standards" with a guideline or a set of criteria to measure targeted abilities, skills, and knowledge (pp. 40-41). In a similar

vein, in Brown's (2003) description, a test is "a method of measuring a person's ability, knowledge or performance in a given domain" (p. 3). To put it differently, it is "a particular type of measurement that focuses on eliciting a specific sample of performance" (Bachman, 2004, p. 9). As for measurement, it is presenting the test scores of an individual's attributes in numerical or quantifiable forms, or in Bachman's (1990) definition, it is "the process of quantifying the characteristic of persons according to explicit procedure and rules" (p. 18). Another explanation of the term measurement provided by Coombe (2018, p. 29) is "the practice of quantifying a physical feature or an attribute".

It can be deduced from the various explanations of these terms from several scholars that testing is the procedure of measuring learners' knowledge, performance, or ability with a test. However, testing and assessment needs to be differentiated since assessment is a more comprehensible term that involves both formal and informal procedures within the perspective of classroom assessment that has a purpose of improving learning through an ongoing process of collecting information of the learners' performances.

2.1.2. Assessment

Assessment is "the process of gathering, interpreting, recording and using information about pupils' responses to educational tasks" (Lambert & Lines, 2001, p. 4). Clapham (2000) expresses that assessment, a more comprehensive and acceptable term, considered as an umbrella term that covers every method of testing and assessment. According to Brown (2003), unlike testing, which takes place at a predetermined time, assessment is an ongoing process that never stops. Learners are continuously assessed by the teacher, other students or they assess themselves whether intentionally or incidentally, formally, or informally. On the other hand, Coombe (2018) emphasizes the systematized aspect of assessment and describes it as "the systematic process of evaluating and measuring collected data and information on students' language knowledge, understanding, and ability in order to improve their language learning and development" (p. 10). In a similar vein, Bachman (2004) proposes that assessment should have a basis and a method, and he defines assessment as "the process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded" (p. 7).

In conclusion, there have been many attempts to define assessment. According to Bachman (2004), a consensus on what assessment means has not been reached. Nevertheless, most

definitions meet on the common ground of the notion that the assessment is a process of gathering information. Overall, assessment is a term covering a range of techniques and strategies including tests in order to collect information about learners' knowledge, skills and abilities with the main purpose of improving teaching and learning. The information on the learners' performances can be utilized to diagnose, select, and place students, give certificates, in other words, to form a final judgement on learners' performances besides educational programs, namely, to evaluate.

2.1.3. Evaluation

Another term that is highly associated with assessment is evaluation. Evaluation is "the process of making a value judgment about the worth of a student's product or performance" (Nitko & Brookhart, 2011, p. 28). In addition to value judgements, evaluation also involves making decisions about either individuals or programs in the light of the information collected through assessment (Bachman, 2004). The information provided by assessment is typically used in order to make decisions about grading, selection, diagnosis and placement of the learners. With respect to evaluation, grading can be presented both numerically and in the form of letters. The distribution of the scores, whether numerical or letter, demonstrates the state of the students so that teachers can make a judgement on students' performances based on evaluation criteria. Additionally, checklist evaluations by the teachers or the students by themselves, student journals, student-teachers conferences and narrative evaluations can be alternative to traditional forms of evaluation (Brown, 2003).

However, evaluation is not restricted to making decisions about students. Educational programs and syllabi can also be evaluated in terms of their effectiveness (Heaton, 1990). For that matter, Coombe (2018) includes the function of evaluating the programs in her definition by stating that it is "the process of utilizing assessment results to gauge and support learning, teaching and program development" (p. 20). As it was mentioned above, decisions on an individual basis consist of diagnosis, selection, placement, certification and grading/marking, and decisions about programs can be useful to improve, change or continue with the existing program by determining the strengths and the weaknesses of the program (Bachman, 1990).

To sum up, evaluation is a way of utilizing the information from assessment to make decisions or judgements about students/teachers or educational programs and syllabi in order to enhance the quality of teaching and learning through providing feedback.

2.2. Purposes of Assessment

In the field of language education, the term *language testing* was used before, and the term *assessment* has been more commonly preferred recently (Hamp-Lyons, 2016). According to Black and Wiliam (2010, p. 82), assessment refers to “all those activities undertaken by teachers – and by their students assessing themselves...” that help them to adjust these activities or make certain decisions with the feedback provided by assessment. Therefore, assessment carries a different purpose than testing, yet it covers tests and other activities that informs teachers about the teaching/learning process (Hamp-Lyons, 2016). Assessment and teaching/learning go hand in hand; therefore, while planning their instructions, teachers should also plan their assessment (Nitko & Brookhart, 2011). However, both are required to be based on the learning goals, targets or objectives, thereby, a purpose.

According to Hughes (2003), the purposes of assessment are: a) to measure language proficiency of the learners, b) to discover student success and achievement in terms of learning objectives, c) to diagnose learners’ strengths and weaknesses, and to determine what students know and don’t, d) to place students in a stage or a part of a language program based on their abilities. Similarly, Lambert and Lines (2001) express that there are four roles of assessment: *formative, summative, certification and evaluation role*. The *formative role* of assessment stands for the feedback derived from assessment about the learners’ progress in order to promote future learning, while *the summative role* of assessment provides information about the degree of learner achievement at certain points or the end of the program. Another purpose of assessment is to enable to select learners based on their qualifications, which is referred as *the certification role*. Finally, *the evaluation role* contains the purpose of making judgements regarding the quality or the effectiveness on an individual or institutional basis.

The distinction of assessment purposes has been clarified by Carr (2011) and Green (2014) in a similar way. Green (2014) notes that assessment is used to collect two types of data, which are *educational assessment* and *proficiency assessment*. The former one is about language learning and takes places in school settings and educational environments, namely

language programs, whereas the latter one is concerned with assessing the extent of progress towards the learning objectives so that the language proficiency of an individual can be determined based on pre-set standards or criteria.

Carr (2011) suggests that assessment data aids in making decisions, and these decisions can be divided into two categories of curriculum-related and other types. Curriculum-related decisions consist of *admission*, *placement*, *diagnostics*, *progress* and *achievement tests*, and they are based on a particular teaching/learning curriculum. With the other types of decisions, proficiency and screening, there is an option of deciding what it can be based on. However, he emphasizes that the distinction between these types is not clear cut, and one type of test can bear multiple purposes.

Admission tests are the first type of tests that help with the decision of whether to admit an individual to the program or not. Admission tests are generally followed by *placement tests* or occasionally applied together as a single test with the object of putting the students in the most suitable level of a language program according to their abilities. Learners' strengths and weaknesses in specific aspects of language are generally identified with *diagnostic tests*. Thus, teachers can gather evidence on the needs of the learners and what kind of planning or adjustments on the curriculum can be made so that learners can benefit from it to focus on the problematic areas. (Hamp-Lyons, 2016). However, placement and diagnostic tests sometimes cannot be distinguished since a placement test may also have a diagnostic purpose (Brown, 2003). After determining the areas that needs to be covered with diagnostic tests, teachers need to figure out what is being taught and learned. Therefore, *progress tests* assess to what extent learners are meeting the objectives of the course or if they are acquiring the targeted skills and knowledge at all while learning is still taking place. As opposed to progress tests, *achievement tests* need to be done at the end of the course or after a particular time frame. The purpose behind achievement tests is to explore what students have acquired after a period and whether the learning objectives or outcomes of the program have been met.

2.3. Types of Assessment

There are various methods, types and techniques of assessment based on why it is being carried out. They can be applied to classroom practices or embedded into a curriculum by teachers and administrators depending on the objectives of the language program or the

needs of the learners. The types of assessment and testing that are discussed in this section are formative/summative assessment, assessment for learning/assessment of learning/assessment as learning, formal/informal assessment, norm-referenced/criterion-referenced assessment, objective/subjective assessment, direct/indirect assessment, discrete-point/integrative assessment, traditional/alternative assessment in terms of performance-based assessment, portfolio assessment, self-/peer-assessment, and finally, dynamic assessment.

2.3.1. Formative Assessment and Summative Assessment

Assessment, based on its function, can be divided into two as formative and summative (Brown, 2003). Formative assessment is a systematic procedure of collecting evidence about teaching and learning (Heritage, 2007) and when this type of evidence is used to adapt and modify the instruction according to the needs of the learners, it becomes formative (Black & Wiliam, 2010). Heritage (2007) specifies the three possible ways that formative assessment can occur, which are during the course spontaneously, planned beforehand and embedded into the current or future curricula/syllabi. It provides the teachers and learners the data about the progress of learners' knowledge, skills and abilities and what needs improvement (Coombe, 2018). Feedback, portfolio, progress tests and observations, simply the techniques that assess the process of the learners (Brown, 2003), are the types of formative assessment (Coombe, 2018). Interestingly, Wiliam (2013) believes that every assessment has a potential to be formative and he argues that describing an assessment as formative would be a mistake because the term 'formative' applies to the function of the information gather from assessment, not assessment itself. That is to say, an assessment technique that is designed typically for summative purposes can be used formatively.

Summative assessment, contrary to formative assessment, is interested in the product of learning. It typically intends to measure the knowledge and skills of the learners and the content area that is covered in the course after a predetermined time frame or at the end of a program in order to reach a final decision about the learning outcomes (Coombe, 2018). Therefore, summative assessment of the learners may not imply future progress but demonstrate the current state of the learners (Brown, 2003). The most common forms of summative assessment are midterm exams, final exams, quizzes, periodic tests, proficiency exams, and most of the formal tests (Brown, 2003; Coombe, 2018).

2.3.2. Assessment for Learning (AfL), Assessment of Learning (AoL) and Assessment as Learning (AaL)

Assessment for learning, a type of formative assessment, aims to improve learning and student performance by forming and utilizing the feedback (Coombe, 2018). Assessment for learning is sometimes used as synonyms with formative assessment in the literature. However, Stiggins (2002) disagrees with the temptation of equating them and puts forth that assessment for learning is more than just frequent testing or collecting evidence to adjust the classroom practice, but also entails active involvement on the part of the learners. Moreover, assessment for learning necessitates teachers to: a) acknowledge the achievement goals in advance of instruction, b) inform students about these objectives so that they would understand them at the very beginning of the process, c) become assessment literate to apply their expectations in their assessment practices so that learner achievement is reflected accurately, d) build learners' confidence by implementing classroom assessment for learners are responsible for their own learning, e) transform the classroom assessment data into continuous descriptive feedback so as to inform students about how to improve, f) adapt instruction constantly according to the classroom assessment results, g) support learners to take part in self-assessment regularly with established standards in order that they see their growth and feel responsible for their own success, h) actively promote learners to communicate with their teachers and families about their status, achievement and improvement (Stiggins, 2002). Briefly, assessment for learning helps teachers to obtain the information to differentiate and reform the teaching and learning activities; therefore, it requires a thoroughly planned design that engages the learners in the learning process and gain insights of the steps of their learning for further development (Earl & Katz, 2006).

Assessment of learning, also referred to as summative assessment (Green, 2014), is “the use of a task, activity, or instrument to measure a student's level of learning or performance...”, where the evidence is not used to advance learning but to elicit or summarize it (Coombe, 2018, p. 11).

Earl and Katz (2006) suggest another type of assessment based on its purpose, which is *assessment as learning*. It adopts a metacognitive perspective to assessment, where learners are aware and in charge of their own learning in that they evaluate the new information and correlate it to what they have learned before. That is to say, students monitor their process and make use of the feedback to make critical adjustments and adaptations in their understandings.

2.3.3. Formal Assessment and Informal Assessment

Formal assessment mostly follows systematic and planned procedures, which may take the form of standardized tests, achievement tests, placement tests, written tests, standard-based performance tests and so forth (Brown, 2003; Coombe, 2018). On the contrary, *informal assessment* is carried through incidental comments, informal observations, suggestions, or spontaneous feedback without the concern of recording learners' performance or making fixed judgements (Brown, 2003), in other words, formal evaluation of the learners (Coombe, 2018). The main goal of formal assessment is to intentionally elicit student achievement through planned testing techniques, while informal assessment is already embedded in classroom activities. Brown (2003) considers all types of tests as formal assessment practices; however, he underlines that not every formal assessment is necessarily testing. For instance, a systematic observation of learners' oral production is formal but cannot be regarded as a test.

2.3.4. Norm-Referenced Assessment and Criterion-Referenced Assessment

Norm-referenced and criterion-referenced assessment differ from each other in terms of interpreting the test results/scores according to a frame of reference. In *norm-referenced assessment*, this frame is based on the relative decisions about learners' performance (Lynch & Davidson, 1994) in that learners' scores are relatively compared to the other learners who have taken the same test (Bachman, 2004). In other words, the main purpose of norm-referenced assessment is to discriminate between learners (Cheng & Fox, 2017). Henning (1987) notes that the standards to discriminate learners' achievement in percentile ranks or rank orders can be determined after the test is designed and administered, for the interpretation of the scores involves some mathematical procedures of calculating the mean, median and standard deviation of learners' scores (Brown, 2003). On the other hand, in criterion-referenced assessment, the standards are planned even before the instruction much less administering the test.

When an assessor is concerned with determining how students perform or achieve in a particular domain based on preset standards rather than comparing it to the others, it becomes *criterion-referenced assessment* (Ornstein, 1993). Hughes (2003) defines the main purpose of criterion-referenced assessment as classifying learners according to their performance on a task or set of tasks whether it is acceptable or not; therefore, the amount of the individuals

who can be categorized as successful is not limited, unlike norm-referenced assessment, where learners' success depends on other learners' performance. Moreover, he proposes that by providing meaningful criteria in terms of what learners are expected to fulfill, it motivates learners to meet those criteria. Similarly, Cheng and Fox (2017, p. 105) asserts that the aim of CRA is to "describe what students know and can do" according to predetermined criteria such as CEFR or as Black (1993) and Brown (2003) mention, a curriculum.

In addition, Black (1998) proposes that norm and criterion-referenced assessment are not two distinct types of tests, but they contain properties that may overlap in some tests. For instance, a test that measures learner achievement in a particular domain can also provide normative data on the number of the students who can achieve that level, whether there are enough students or not.

2.3.5. Objective Assessment and Subjective Assessment

Both objective and subjective assessment are about nothing but scoring (Heaton, 1990; Hughes, 2003). The basic difference between objective and subjective assessment is whether a judgement by the scorer is required or not (Hughes, 2003). In *objective assessment*, the students will get the similar results since there is just one correct answer or a small number of correct options in some cases (Heaton, 1990), and predetermined criteria is used in order that scorer's judgment can be eliminated from the results (Bachman, 1990). With *subjective assessment*, a judgement is needed while interpreting the correctness of the learners' response (Hughes, 2003). Bachman (1990) states that objective assessment techniques such as multiple-choice tests and short answer responses may be preferred by testers for the reliability feature it adds on to the assessment; however, Hughes (2003) proposes that there are ways to increase the objectivity of subjective tests such as written compositions and oral interviews.

2.3.6. Direct Assessment and Indirect Assessment

Another dichotomy in the field of language testing and assessment, based on test construction, is direct and indirect assessment (Hughes, 2003). The main difference between them is that with *direct assessment*, learners are required to perform exactly the skills that are meant to be measured, whereas *indirect assessment* aims at discovering the underlying aspects which are intended to be assessed (Hughes, 2003). An example of direct assessment

could be oral presentations used to measure learners' oral abilities (Carr, 2011), while assessing learners' pronunciation skills with a pen and paper test is considered as indirect assessment since there is no actual oral production to assess pronunciation skills of the learners directly (Hughes, 2003). According to Hughes (2003), direct assessment can be advantageous since it is straightforward in terms of eliciting learners' behavior providing that the abilities to be assessed are clear, and when it comes to assessing and interpreting the learners' performance of productive skills. Besides, it may lead to positive washback in that practicing for the assessment entails practicing for the skills that are aimed to be enhanced. Indirect assessment also has some advantages in terms of providing more samples of the language rather than a finite number like direct assessment; therefore, it is more representative of the language.

Bachman (1990) claims that 'direct' tests are indirect, however, the problem here is not the direct tests or tasks themselves, but the label referring to them (Carr, 2011). The issue is learners' performance in direct assessment may not reflect their competence in real-life language use due to some reasons other than weak language ability, such as anxiety, unfamiliarity with the content (Bachman, 1990) even though performance is typically regarded as an indicator of the underlying competence (Carr, 2011). Despite the discussion about what they mean, the terms, direct and indirect assessment, are still commonly used.

2.3.7. Discrete Point Assessment and Integrative Assessment

Another classification of assessment types is between discrete-point and integrative assessment. Hughes (2003) explains that *discrete-point assessment* is the testing of language elements on an individual basis. It is assumed that language can be separated into its components and assessed successfully (Brown, 2003). Discrete-point tests focus on language elements like grammar, pronunciation, morphology, vocabulary, syntax and so on, and they mostly do not offer any purpose or situation (Douglas, 2010). Therefore, it may confuse learners due to the unrelated tasks and decontextualization of the language (Brown, 2003). Hidri (2018) depicts three techniques that discrete-point assessment relies on: phonemic discrimination, paraphrase and recognition, and response evaluation, generally appearing in the form of multiple choice, yes/no or true/false questions. Although this approach is criticized for isolating the language elements and not promoting authentic language use, Carr (2011) proposes there are various advantages of discrete-point assessment. When a specific component of language is assessed, it directly demonstrates learners' lack of ability in that

area, not the other aspects that might interfere with learners' performance. Moreover, it is a powerful tool for large-scale standardized tests for its accurate assumptions on analyzing the tests and items, and statistical methodology. Nevertheless, it still lacks to demonstrate what learners can actually do with the language as a result of being highly structured and unrepresentative of real-life language use (Carr, 2011). Thus, an integrative approach to language assessment is required.

Integrative assessment requires learners to combine more than one language element while completing a task (Hughes, 2003). An integrative task entails learners to process several language aspects simultaneously (Douglas, 2010), such as cloze-tests and dictation, which are considered as a form of integrative assessment since multiple skills are blended (Brown, 2003; Hidri, 2018; Hughes, 2003). Even though, as Hidri (2018) suggests, with integrative assessment, it is more possible to provide a larger context compared to discrete-point assessment, some problems may occur. Carr (2011) identifies the problematic areas as scoring and interpretation of the results. Integrative tests are more difficult when it comes to scoring, and the interpretation of these scores. To illustrate, when a student gets poor results from an integrative listening and speaking task, the problematic skill cannot be identified clearly. It can be due to their weaknesses in either listening or speaking skill or even both. Another issue is the scope of the content area that is covered. As with integrative assessment, several skills are combined, hence leading to be too broad to assess what is intended or assessing 'fuzzy constructs' as Hidri (2018, p. 3) denotes.

2.3.8. Traditional Assessment and Alternative Assessment

Alternative assessment is considered as an alternative to standardized testing and defined in the frame of its discrimination from traditional assessment. Therefore, there is not a universal definition of alternative assessment (Huerta-Macías, 2002). According to Alderson and Banerjee (2001), it involves procedures that are less formal, continuous, formative, and low-stakes compared to traditional assessment, which is formal, summative in function, high-stakes, and one-shot. Huerta-Macías (2002) points out that the problems and shortcomings of traditional methods in assessment have led the emergence of finding the alternatives in the first place or as Brown and Hudson (1998 p. 657) refers 'alternatives in assessment'. Rather than focusing on learners' ability to recall and reproduce, alternative assessment demonstrates what learners can actually do with the language and focuses on how they integrate and perform on real-life tasks (Huerta-Macías, 2002). The characteristics of

alternatives in assessment differ from traditional methods in that it a) is not implemented at a separate time frame, but integrated into daily classroom activities, b) mirrors the curriculum since the activities are already embedded in the curriculum, c) provides individualized evidence about the weaknesses and strengths of the learners, d) provides multiple sources of information to assess learners progress as various methods are employed, e) suits various populations in terms of multicultural sensitivity (Huerta-Macías, 2002). Gipps and Stobart (2003) believe that alternative assessment has particular weaknesses as well as strengths. Even though it can be efficient in relation to improving learning in classroom environment, considering large-scale contexts, some problems may occur. Alternative assessment is mostly discussed in the matter of reliability and validity (Fox, 2017; Gipps & Stobart, 2003). The most common alternatives in assessment are portfolio assessment, performance-based assessment, self- and peer-assessment, and teacher observations, journals and conferences are some of the tools for alternative assessment.

2.3.8.1. Performance-Based Assessment

Performance-based assessment necessitates learners to “accomplish approximations of real-life, authentic tasks, usually using the productive skills of speaking or writing, but also using reading or writing or combining skills” (Brown & Hudson, 1998, p. 662). The aim is to offer learners the opportunity to simulate real-life via more contextualized tasks (Brown, 2003). Thus, it can estimate learners’ actual language abilities and indicate future performances with more authentic tasks compared to traditional standardized assessment methods (Brown & Hudson, 1998). However, the problem with performance assessment is that producing and administering a task can be difficult and time consuming. Besides, reliability is a major concern on the account of subjective or inconsistent ratings (Brown, 2003).

2.3.8.2. Portfolio Assessment

Among the several methods of alternative assessment, portfolio assessment has been the most remarkable one (Fox, 2017). Portfolio assessment is “purposeful collections of any aspects of students’ work that tell the story of their achievement, skills, efforts, abilities, and contributions to a particular class” (Brown & Hudson, 1998, p. 664). Materials such as written works, artwork, personal reflections, reports and audio or video recordings can be included in portfolios (Brown, 2003). Even though it may be time consuming and

impractical, the characteristics of portfolio assessment present a foundation in order to enhance learning (Gipps & Stobart, 2003). Brown and Hudson (1998) enounce the advantages as: a) strengthening students' learning process in terms of motivating them to engage in meaningful activities and employ critical thinking skills through self-evaluation, and promoting collaboration of the students with their teacher and peers, b) enhancing teachers' role as a facilitator and demonstrating a vivid picture of each learners' process through individualized learning, c) improving assessment process by providing a wide variety of data about learners with several authentic tasks. However, there are some concerns that needs to be considered for while planning portfolio assessment (Brown, 2003). For a successfully developed portfolio, objectives, and time frame, in line with the existing curriculum, should be clearly specified with authenticity and washback in mind, tasks should be determined beforehand and negotiated with the learners, works of the students should be reviewed systematically.

2.3.8.3. Self-Assessment and Peer-Assessment

Self-assessment, a learner centered form of alternative assessment, is defined by Coombe (2018, p. 37) as “the ability of language learners to assess their own performance” so as to diagnose their strengths and weaknesses during the learning process. Thus, it is commonly associated with autonomous learning in that it assists learners to monitor and assess their own progress and concentrate on their own learning (Harris, 1997). Moreover, students are accountable for their own learning and assessment in self-assessment (Oscarson, 1989), thereby, promoting active learning even in formal settings, where learners hold a relatively passive approach (Harris, 1997). However, Oscarson (1989) believes that self-assessment can be beneficial as a complementary resource along with other established techniques in language assessment. Learning journals/diaries, questionnaires, self-evaluation checklists can be employed as tools for learners to evaluate their progress (Oscarson, 1989).

Peer assessment is based on the principles of cooperative learning (Brown, 2003). It is “an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners” (Topping, 2009, p. 20). Like self-assessment, the possible outcomes of peer-assessment are active engagement of the learners, enhanced motivation and learner autonomy (Brown & Hudson, 1998; Douglas, 2010). Learners can provide feedback or/and grades to their peers individually, in pairs or groups through several tasks from portfolios to quizzes (Hansen Edwards, 2014). However, the subjectivity and

accuracy of the assessment by students' self or peers needs to be considered since the results may be influenced by learners' linguistic abilities or their perceptions may not reflect reality (Brown & Hudson, 1998). To this respect, Douglas (2010) offers that learners need training to be able to assess themselves or their peers and provided with examples of the expected product/performance. Besides, a criterion can be established by the teacher or learners before assessment takes place.

2.3.9. Dynamic Assessment

Dynamic assessment (DA) is a recent approach to assessment that has caught the attention of the researchers, practitioners, and other stakeholders (Poehner, 2014). DA is not a method or a tool, but a principle for conceptualization for learner development through forms of unified teaching and assessment activities (Lantolf & Poehner, 2010). The theoretical framework of DA is based on Zone of Proximal Development, a sociocultural theory elaborated by Vygotsky, which aims learners to extend their performance beyond their current capabilities through appropriate guidance and support (Poehner, Davin, & Lantolf, 2017). In this respect, teachers work as a mediator, who assists learners and intervenes with their difficulties in the process for the purpose of continuous development, which is also referred to as mediation.

2.4. Principles of Language Assessment

All kinds of language tests or assessments should follow some certain principles (Harris, 1969). These principles are also named as qualities, criteria, key concepts or as Bachman and Palmer (1996) refer to as 'usefulness'. For a test to be good or effective, these principles need to be taken into consideration (Brown, 2003) while either designing or selecting an instrument. An effective instrument is expected to provide dependable data, measure intended objectives, be practical, be in line with real-life and have positive impact. In other words, the principles of a good test consist of reliability, validity, practicality, authenticity, and washback. However, it may not be possible for an instrument to involve all these properties at the same time; therefore, Bachman and Palmer (1996) inform that a proper balance should be determined depending on the specific testing situation.

2.4.1. Reliability

Reliability is directly associated with the consistency of an instrument or the test scores in varying testing situations (Bachman & Palmer, 1996; Brown, 2003; Chapelle, 2012; Harris, 1969; Popham, 2018). In other words, participants are expected to have similar results when the same test is implemented again after a certain period without any intervening factor in between, which is also called as test/re-test reliability, and if the results of the same participants differ to some extent, the test would be unreliable (Brown, 2003; Heaton, 1990). Chapelle (2012) explains that the maximum value of a reliable test is expressed with '1', and '0' is the minimum value, which indicates a completely unreliable test. For a test to be considered reliable, the value should be higher than .70, and reliabilities over .90 is considered highly reliable. However, if the value is below .70, then, it can be inferred that the test is problematic in terms of reliability. A test would provide inaccurate information on the ability that is intended to be measured unless it is stable and consistent. (Bachman & Palmer, 1996); thus, the interpretations of the test results or inferences about what those results indicate would also be inaccurate (Popham, 2018). For the sake of obtaining dependable data with a reliable instrument, factors that may interfere with the results, also called errors, need to be eliminated from the process (Jones, 2012). Types of errors and possible methods to minimize these errors to maximize the reliability have been illuminated by many scholars in the literature (Bachman, 1990; Brown, 2003; Harris, 1969; Hughes, 2003). The main sources of errors related to assessment are mostly affected by individual, environmental or instrumental factors.

2.4.1.1. Student-Related Reliability

Test-takers' physical (i.e., health problems, fatigue) and psychological (i.e., anxiety, motivation) conditions affecting the results are considered as potential sources of errors related to students (Brown, 2003). In addition, test-taking strategies or 'test-wiseness' may also prevent students from demonstrating their actual performance (Bachman, 1990; Brown, 2003).

2.4.1.2. Rater-Reliability

It has been said that the consistency of students' scores in different occasions indicates the reliability of a test. However, other individuals such as raters or markers are involved in the

process as well. Therefore, consistent scoring of the raters is looked for higher reliability (Brown, 2003). Subjectivity, bias, lack of training, experience and attention is a threat to rater-reliability. When two or more raters demonstrate agreement while assessing students' performance on the same task, it is referred to as *inter-rater reliability*. Conversely, *intra-rater reliability* is about the internal consistency of the same rater's assessment of the same performance (Coombe, 2018). It is considered as a risk for intra-rater reliability when a rater is expected to score an excessive amount of test or performance, leading to tiredness, or when scoring criteria is not defined clearly (Brown, 2003). Hughes (2003) discusses that some test formats or tasks are liable to subjectivity since a level of judgement may be required. However, there are ways to obtain objective scoring from these subjective tests, such as using clearly specified scoring criteria (Brown, 2003; Hughes, 2003).

2.4.1.3. Test Administration Reliability

Environmental factors in which the test is administered might impact the reliability adversely. Physical conditions of the desks, temperature and lighting of the room, quality of the photocopies and noise from the street may cause test takers to perform poorly on a test, which causes unreliable scores (Brown, 2003). For example, when assessing learners listening skills, the noise coming from the street may constrain learners to understand the recording properly. Thus, when a reliable test is administered in different testing conditions, the results are expected to be similar, and if not, it will yield to lower levels of reliability regarding test administration (Harris, 1969).

2.4.1.4. Test Reliability

A test, in its nature, may harm reliability. If a test is poorly constructed involving ambiguous items or multiple correct answers, it can cause measurement to be erroneous (Brown, 2003). Another factor that influences test reliability adversely is the duration of the assessment. For instance, fatigue and exhaustion can be observed among test takers when administration takes too long, and test takers may experience anxiety over time restrictions when they are expected to complete the task within a fixed period of time. Additionally, Harris (1969) indicates a positive correlation between the test reliability and the sufficiency of sampling. The more samples of the learners' performance are gathered, the more reliable the assessment will get.

Even though measuring language abilities may not be as precise as measuring someone's physical attributes such as height or weight since language ability is not considered to be an absolute concept (Green, 2014). Consequently, in educational measurement, there are only so many forms of measurement that can be qualified as “‘perfectly’ reliable” (Harris, 1969, p. 17.) However, there are numerous approaches to calculate the reliability of a test in order to estimate the degree to which measurement signifies test takers' true scores (Fulcher & Davidson, 2007; Heaton, 1990; Henning, 1987). The test/re-test, parallel forms, split-half and K-21 are the four common methods of estimating reliability.

2.4.1.5. Test/Re-test Method

Test/re-test method is administration of the same test twice after a period (Chapelle, 2012). The correlation between the scores of the test takers on each administration is analyzed in order to calculate reliability. This method assumes that there is no learning taking place in-between the time interval and the memory effect of the first administration on the second is none (Fulcher & Davidson, 2007). Hughes (2003) discusses the drawbacks of readministering the same test. The time in between the administrations should be considered cautiously since if the duration between test and retest is too short, participants might remember the items and their responses to them. On the contrary, when the duration between the administrations is too long, the testing conditions would not be the same since some learning or forgetting might take place in between. In order to prevent the changes in testing situations, thereby, revealing learners' true scores, Henning (1987) suggests interval limitations should not be more than two weeks.

2.4.1.6. Parallel Forms Method

Parallel forms method deals with the consistency between two forms of a test that are ‘supposedly’ equivalent to see whether they are parallel (Popham, 2018, p. 79). It is also named as equivalent forms, alternate forms along with parallel forms (Henning, 1987). The procedure of administering a parallel form of a test and measuring its reliability is similar to test/re-test method in that two test forms are applied to the same sample of individuals except for with a little or no waiting period in between (Popham, 2018). Therefore, one of the advantages of this method is believed to be decreasing the practice effect (Henning, 1987), albeit not completely (Harris, 1969). Harris (1969) puts forth that parallel forms of a test

should fulfill the requirements of a rigid criteria to be equivalent in its difficulty, length, test format and duration. Nonetheless, it is not likely to find two tests that are the same or equal (Hughes, 2003). Thus, split-half method is employed in most cases.

2.4.1.7. The Split-Half Method

Another method of measuring test reliability is the split-half method. When it is unlikely to administer the same test twice or no alternate form of a test exists, split-half method could be employed. In addition, the shortcomings of previous methods, such as not being able to administer the exact same test under the exact same circumstances, can be eliminated from the process with split-half method (Chapelle, 2012). In this method, an accurate instrument is divided into two equal parts and applied to the same participants as a single test (Heaton, 1990). The participants' scores obtained from one half are correlated to the other half whether to see if there is a significant difference between them, which would indicate the reliability of the test as a whole (Fulcher & Davidson, 2007). However, while breaking the test into two parts, it is essential that the halves are equal in terms of item difficulty (Heaton, 1990).

2.4.1.8. Kuder-Richardson Formulas

KR-20 and KR-21 are the formulas developed by Kuder and Richardson (1937) to measure the reliability estimate of a test by dividing a single test into its potential halves. Same as split-half method, the procedure is that a single test is administered to a group of individuals only once. It is concerned with the homogeneity and internal consistency of the items consisting of two dimensions, either right or wrong. However, they suggest that KR-21 formula should be employed where there is a large sample of items. Otherwise, KR-20 would be a better choice.

2.4.2. Validity

Validity is among the most primary concerns regarding testing and assessment. Early views on validity were built upon the notion that a test measures what it is supposed to measure (Henning, 1987; Hughes, 2003; Lado, 1961). In other words, validity of a test is determined by its appropriateness or accuracy of measurement depending on its purpose. However, this

traditional approach to validity has changed drastically to interpretations and use of the test scores rather than being just a characteristic of a test (Chapelle & Voss, 2013). In Messick's (1989, p. 5) description, validity is "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment". Therefore, what is validated is the justification of test interpretations and its use, not the test itself or other assessment devices. Additionally, he disagrees with the discrimination of validity into categories and embraces a unified view of validity against the traditional 'types of validity' and regards them as 'forms of evidence'. Traditionally, there are four types of validity, which are criterion-related validity (predictive validity and concurrent validity), content validity and construct validity (Cronbach & Meehl, 1955), and face validity.

2.4.2.1. Criterion-Related Validity

With *criterion-related validity*, the tester aims to predict a criterion with a particular test (Cronbach & Meehl, 1955; Fulcher & Davidson, 2007). It follows a statistical procedure, where a test is compared with another test of the same aspect of language that is already established to see the correlation between them according to the scores of the individuals (Akbari, 2012). If the scores of the individuals correspond with the scores of the criterion test against it, it can be claimed the predictions are effective (Farhady, 2012). The scheduling of the criterion measure identifies the two types of criterion-related validity. *Concurrent validity* is the administration of a newly prepared test and a criterion measure simultaneously and the correlation between both test scores are calculated (Brown, 2012), whereas *predictive validity* is estimated when a gap exists between the new test and criterion measure (Akbari, 2012). Thorndike (2014) further explains that the evidence from the test is utilized to forecast the future performance of the individuals in predictive validity, such as tests for selective purposes, which are the predictors of the individuals' performance in a particular program that they been selected for. However, Green (2014) and Messick (1989) argue that criterion-related validity has critical limitations in terms of the appropriateness of the criterion measure to the test, the likelihood of the contaminated test scores by extraneous variables, such as health, characteristics of the individuals, rater bias and alike, or namely, reliability (Popham, 2018).

2.4.2.2. Content Validity

Unlike criterion-related validity, *content validity* is not based on empirical evidence, but logical or subjective judgements, it is, on the other hand, vital to the overall validity (Farhady, 2012; Henning, 1987). Content validity is concerned with “the adequacy of a test in terms of its content or the sample of behavior it sets out to measure” (Akbari, 2012, p. 31). Sireci (1998, p. 101) defines the aspects of content validity as “domain definition, domain relevance, domain representation, and appropriate test construction procedures”. A test can be relevant with the intended content; however, this may not warrant content validity. In order to ensure content validity, a test is required to employ a representative and comprehensive sample of the course or curriculum objectives (Fulcher & Davidson, 2007; Henning, 1987; Hughes, 2003). The selection of the content that is included in the measure needs careful investigation due to the risks of covering underrepresented or irrelevant content (Hughes, 2003; Messick, 1996). Therefore, asking for opinions of an expert on the content area to be included in/excluded from the test or designing a table of specifications can entail higher content validity (Fulcher & Davidson, 2007; Hughes, 2003).

Based on the current views of unified framework of validity as opposed to distinct properties of a test, Messick (1989) argues that the concerns of content validity are directly related to construct validity, thereby, an incorrect terminology. In addition, Sireci (1998) and Kane (1992) conclude that even though it may not be sufficient on its own for overall validity, content validity provides fundamental evidence about the assessment instrument in order to attain construct validity.

2.4.2.3. Construct Validity

Cronbach and Meehl (1955, p. 283) define the term ‘construct’ as “some postulated attribute of people, assumed to be reflected in test performance”. In particular to language education, a ‘construct’ is “any underlying ability (or trait) that is hypothesized in a theory of language ability” (Hughes, 2003, p. 31). The definition of construct is significant for the sake of understanding the concept of construct validity. The capability of a test in “measuring certain specific characteristics in accordance with a theory of language behavior and learning” is related to its construct validity (Heaton, 1990, p. 161). This early concept of construct validity was based on the assumption that certain constructs or theories exist in test takers and their test scores would reflect the extent of existence of these properties (Fulcher &

Davidson, 2007). With his approach to validity that is widely agreed upon, Messick (1989) defines construct validity as a unitary concept that subsumes other abovementioned aspects of validity with the purpose of collecting evidence to support the test interpretation and utilization. This approach is also observed in Bachman's (1990) and Bachman and Palmer's (1996) work. They adopt the similar approach to validity and introduce it as a unifying concept pertaining to accurate and meaningful interpretations of test scores, in which justifications of the interpretations are subject to validation, not the test itself. In conclusion, construct validity is considered as a central and overarching concept in validation (Chapelle, 1999; Cumming, 1996).

2.4.2.4. Face Validity

Face validity is concerned if the “physical appearance of the test corresponds to what it is supposed to measure” (Farhady, 2012, p. 38). It is simply about the way an instrument looks to the administrators, test takers, test designers and so on (Harris, 1969). Therefore, a test should be perceived to be relevant to the language component or skill that are wished to be measured in order to have face validity (Akbari, 2012). A test that aims to measure test takers' pronunciation ability, yet not requiring them to speak can be said to have lower face validity (Hughes, 2003). Harris (1969) and Heaton (1990) discuss the importance of face validity in terms of its motivational impact on test takers. They believe that a test lacking face validity will adversely affect test takers' motivation concerning that they may not adopt the test and perform at their best. Thus, as Harris (1969) exemplifies, a test that is initially designed to measure native young learners' reading comprehension lacks face validity when administered on adult foreign language learners. Nevertheless, Popham (2018) urges to consider that appearance of a test may be deceiving in some cases; thus, it would be wiser to look for further evidence on different aspects of validity in order not to jeopardize overall validity. That is to say, a test may lack face validity and still can be proven to be highly valid according to other criteria, or vice versa (Farhady, 2012).

2.4.3. Practicality

Practicality is rather related to the test implementation than the interpretation of test scores (Fulcher & Davidson, 2007); therefore, its nature differs from other qualities of assessment (Bachman & Palmer, 1996). There are some considerations that should be taken into account

while designing or adapting, administering and scoring an instrument for the sake of practicality. (Farhady, 2012; Harris, 1969). Bachman and Palmer (1996) relate the practicality to the relation of the required resources and existing resources, which are specified as human, material and time besides the cost of each resource. The degree of this relationship between how much is needed to implement a test and how much exists determines the practicality of an instrument. The concerns of test practicality is further elaborated by Rogier (2014, p. 5) as “the cost of test development and maintenance, time needed to administer and mark the test, ease of marking, availability of suitably trained markers, and administration logistics”. Davies (1990) highlights the importance of test practicality by related to its effects on validity and reliability. A test that exceeds the limits of available resources would be impractical, thereby, may not be used in some cases. Thus, he suggests that the reliability and validity of the tool cannot be known as the test is not implemented.

2.4.4. Authenticity

Bachman and Palmer (1996) believe that to demonstrate individuals’ performance on tests as an indicator of their language use in particular domains exclusive of the test itself, a language test is desired to reflect real-life use of target language, in other words, authentic. Authenticity is about “the degree of correspondence of the characteristics of a given language test task to the features of a TLU [target language use] task” (Bachman & Palmer, 1996, p. 23). Authentic tasks reflecting real-life contexts help learners be motivated to perform well (Rogier, 2014). Moreover, lack of authenticity in a test will generate questions about the generalizability of the scores (Spolsky, 1985) in that authenticity entails a meaningful investigation on the degree of generalizability of test performance to the language use in target language contexts (Bachman & Palmer, 1996). However, Douglas (2010) identifies the problems with authenticity is that it can be difficult to achieve since the assessment task does not take place in the real world, but in learning context, where real-life can only be simulated. Similarly, the issue with authenticity is put forth by Spolsky (1985) as “any language test is by its very nature inauthentic” (p. 39). Even though the extent of authenticity is limited in language testing and assessment, Douglas (2010) urges that it is significant to provide contexts that resembles real-life so as to ensure valid interpretations of individuals’ performances.

2.4.5. Washback

Another fundamental concern in language testing and assessment regarding the influence of testing on various stakeholders in relation to such testing is washback (or backwash), impact and consequences (Tsagari & Cheng, 2017). *Impact* is defined as the effect of a test on “learners and teachers, on educational system in general, and on society at large”, whereas *washback*, considered as a part of the impact, is the “effect that tests have on learning and teaching” (Hughes, 2003, p. 53). For Messick (1996, p. 241) washback is “the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning”. Therefore, washback is believed to take two directions: positive or negative (Wall & Alderson, 1993). The possible positive and negative effects of testing that were discussed in several studies in the literature were questioned by Alderson and Wall (1993). They hypothesized that testing would influence content, method, rate and sequence, degree and depth, and attitudes to the content and method etc. of teaching and learning. In discussion of these hypotheses, they argued that both positive and negative washback may occur regardless of the quality of the test. A poor test may result in beneficial washback or vice versa, not due to the test necessarily, but as a result of societal, educational factors and factors related to school setting, which are, then, out of the scope of washback. Thus, the influence of a test in a wider perspective should be held under the notion of impact (Hamp-Lyons, 1997). Whereas Watanabe (2004, p. 22) denotes that being a complex phenomenon, washback and the context of the test should be considered as a whole. By virtue of this rationale, he presents five possible factors generating washback positive or negative, which are; *test factors* (i.e., purpose, method, content of a test, decisions based on the test etc.), *prestige factors* (i.e., importance of the test for learners and in educational systems etc.), *personal factors* (i.e., educational background of the teachers and their methodological beliefs etc.); *micro-context factors* (i.e., school settings etc.); and finally, *macro-context factors* (i.e., tests within society etc.). Based on his earlier conception of a unified validity, centered around construct validation (Messick, 1989), Messick (1996) discusses washback from the viewpoint of consequential aspect of validity and states that it forms the consequential basis for testing within its social context and should be associated with overall validity since it will support the interpretations derived from the test scores. He highlights that authenticity can be linked to washback. That is to say, a test is likely to have beneficial washback in the event of

including samples of communicative behaviors of the target language skills that are authentic and direct.

As it can be seen from the aforementioned theoretical considerations, washback is a complex notion that needs due attention in language assessment concerning its influence on teaching and learning, or broadly, education and society.

2.5. Assessment of Language Skills

In language education, it is aimed to develop learners' competence in the aspects of language. Language construct mainly underlies reading, listening, speaking, and writing skills. Each skill has its own unique concerns that require in depth discussions; that is, it will be handled individually in the following sections. When assessing these skills, whether separately or integrated, the construct definition is the first thing that should be done in order to determine what is intended to be assessed. Besides the construct definition, the aim of the assessment, test specifications, task types, techniques need to be established and considered thoroughly with background knowledge, proficiency level, authenticity and the principles of assessment that were discussed in the previous sections in mind.

2.5.1. Assessing Reading

In reading, a receptive skill, the aim is to figure out how well learners utilize their linguistic abilities to comprehend the related input, spoken or written (Green, 2014). Unlike productive skills of language, receptive skills, reading and listening, can be challenging to observe directly since the process and a particular product of reading can only be inferred not observed (Brown, 2003). The inferences of what learners have received from the reading text can be demonstrated by adopting other activities like drawing, writing, speaking, or acting (Green, 2014). It is believed that learners reading skills in their own language can be reflected on their reading skills in a foreign language. Therefore, learners' reading skills in their first language needs to be established before developing their reading skills in another language (Alderson & Banerjee, 2002) although mastery of first language reading skills may not guarantee that these skills will be transferred into reading a foreign language (Heaton, 1990). Brown (2003) posits that assessing reading should not be degraded to comprehension only; reading strategies of the test-takers should also be considered as a factor while assessing learners' comprehension. For instance, learners may be unfamiliar with a text type,

which may lead to misunderstanding, hence, lower comprehension. Thus, in order to become proficient in reading, learners are expected to acquire some strategies; *bottom-up* and *top-down*. With bottom-up approach, learners focus on smaller units of language in a particular sequence such as letters, sounds, words, and phrases in order to decode meaning, whereas the comprehension in top-down approach is influenced by the conceptual or background knowledge and predictions of the readers (Alderson, 2000). However, both bottom-up and top-down approaches are found to be inadequate to identify reading process. As a result, *interactive model* that combines both approaches are established. In this model, readers combine their background knowledge at global level and linguistic knowledge at local level in order to demonstrate their understanding of the reading text (Hubley, 2018). Strategies such as skimming for the main idea and scanning for details can be employed in interactive models.

While assessing reading abilities of the learners, the first thing to do is to define the construct of reading that is aimed to be assessed based on a particular setting or your purpose (Alderson, 2000). After that, test specifications of reading can be defined in correspondence with the objectives of a curriculum (Hubley, 2018). Test specifications involves the text types and skills that are required to be efficient readers. Understanding the gist of the entire text or at paragraph level, identifying the purpose, genre, type and audience of the text, authors' stance, and intentions, discriminating a fact from an opinion or a theory, making inferences based on the implications in the text, guessing the unknown words, reordering the items are among the most common skills that are assessed (Hubley, 2018; Hughes, 2003). However, selecting the reading text to be included in the test needs to be carefully considered regarding its length, subject or topic, style and genre and the difficulty of the language (Harris, 1969; Hubley, 2018). The length of a reading text should be properly balanced; that is, it should not be too long to read and comprehend yet contain sufficient content to be meaningful. In addition, the characteristics of a text regarding their type, genre or formality may have an impact on comprehension. Besides, the text in which syntactical structures and lexical components are extremely complex or too simple for the level of the learners would not be appropriate to assess their reading abilities. As for the topic or subject of the text, texts that learners are excessively familiar or unfamiliar with shouldn't be involved since it will affect reading comprehension. Alderson (2000) puts forth that texts requiring some sort of topical or background knowledge is not usually incorporated in the construct that is assessed though can be controlled or utilized as a facilitator of performance. Heaton (1990) addresses

another concern about reading texts. He suggests that reading texts should be diversified in their types such as poems, advertisements, articles from newspapers, guidelines of a tool rather than just using literary texts in prose version. Including various types of texts will enhance learners' motivation as they are exposed to the target language mirroring real-life use in addition to supporting more reliable and realistic assessment. In other words, authentic reading texts are significant through their contributions to assessing reading.

Finally, even though there is no best method (Alderson, 2000), most common techniques in assessing reading are short answers, multiple choice questions, cloze tests and gap filling activities, information transfer activities, matching, ordering/sequencing, true/false questions, yes/no questions, editing, summarizing, skimming and scanning, open ended questions (Alderson, 2000; Brown & Abeywickrama, 2010; Heaton, 1990; Hughes, 2003).

2.5.2. Assessing Listening

Assessment of listening skills is not believed to receive the necessary attention relative to other language skills (Brindley, 1998). The reason, as Brown (2003) indicates, could be that listening is considered as a component of speaking in spite of the instances where no oral production is required such as listening to a lectures, announcements or radio (Hughes, 2003). Another reason that he provides is the bias towards speaking rather than listening in that speaking is seen as the first and foremost indication of knowing a language. Besides, like reading, listening is a receptive skill that incorporates an internal process inside the minds of individuals, which means that the output of listening is not empirically observed or measured (Wagner, 2013). Thus, learners' listening abilities needs to be inferred from learners' responses to some type of tasks.

Buck (2001, p. 247) describes listening as “a complex process in which the listener takes the incoming data, an acoustic signal, and interprets it based on a wide variety of linguistic and non-linguistic knowledge”. Linguistic knowledge contains syntax, semantics, phonology, lexical structures, and discourse markers, while non-linguistic knowledge is about topical, contextual and world knowledge. Understanding an incoming sound entails many types of knowledge and the way that they are processed doesn't have to be in an explicit order necessarily but can occur simultaneously or in any sequence. Therefore, listening comprehension is thought to be an interactive process. Compared to reading, assessing listening abilities can be challenging since the spoken language is more complex than written

language in terms of ‘redundancy’ (Heaton, 1990). Examples of redundancy in speaking can be pictured as, when someone is speaking, the language generally contains hesitations, repetitions, fillers, repairing false usages, correcting mistakes in lexical and structural forms, gestures, intonation and stress, and speed, all of which helps to understand the message in the context (Buck, 2001; Heaton, 1990). Besides, written language comprises of more formal properties, whereas spoken language follows more informal and colloquial constitutes and non-standard properties of language such as slang and dialect (Buck, 2001). As a result of these complexities of the spoken language, listening could be ignored in language learning contexts regardless of its essentialness. In addition, defining the construct of listening abilities can be challenging since there are many factors influencing listening comprehension (Abeywickrama, 2018a). Along with the unobservable nature of listening, the texts adopted for testing and assessment should be carefully evaluated in terms of its type, form, length, language characteristics and speed rate as a consequence of their impact on comprehension. A possible solution for this problem is composing listening tasks according to the context where learners are expected to perform in non-test situations or as Bachman and Palmer (1996) present target language use domain. Another concern in the assessment of listening abilities is the correspondence of the listening texts and tasks with real-life (Abeywickrama, 2018b). To illustrate, assessing listening separately, though leads to clear construct definition and enhanced reliability since other skills will not interfere with it, may not reflect real-life situations in most cases as listening will entail other skills to be involved. Furthermore, listening texts are generally planned and practiced; thus, it lacks the properties of spoken language such as hesitations, repetitions, dialects, non-verbal cues and so forth that may occur in a real-life context. In conclusion, authenticity is one of the fundamental issues in assessing listening.

Buck (2001) specifies the approaches to assessing listening into three categories: discrete-point, integrative, and communicative approach. It is believed that language constitutes of elements that can be isolated and tested independently in discrete-point approach. Measuring the most crucial bits of language can represent the overall knowledge of language. Integrative approach, on the other hand, stresses the significance of practicing language as whole rather than knowing bits of it. Finally, communicative approach aims to mirror real-life listening context provided with authentic tasks. According to Brown and Abeywickrama (2010), there are four types of listening that learners are expected to perform, which are intensive, responsive, selective, and extensive listening. The first performance is related to

recognizing and discriminating linguistic components such as phonemes, words, and intonation. With the second type of performance, responsive listening, learners are required to provide some sort of limited response to the input by answering questions, following instructions, etc. The aim is not to obtain a general idea of the text provided, but to scan the text in order to detect a specific detail such as listening for a name, number or factual information, which are in the third type of performance, the selective listening. Unlike selective listening, extensive listening aims to acquire a global understanding of what is being heard. Understanding the main idea or the gist of the text and making inferences take place in extensive listening.

To conclude, Buck (2001) suggests that the physical condition of the setting, participants, timing, instructions of the tasks, scoring method and the characteristic of the spoken input needs careful consideration before designing or selecting listening tasks. Most common formats and tasks are short answers, multiple choice questions, true/false items, gap-filling activities, note-taking, summarizing, retelling, dictation, listening cloze, information transfer, editing (Brindley, 1998; Brown, 2003; Buck, 2001; Heaton, 1990; Hughes, 2003).

2.5.3. Assessing Speaking

Assessing speaking skills wasn't always the subject of interest in the past. The importance given to assessing speaking has changed with the theoretical changes in language education (Isaacs, 2016). The complexity of speaking requires "the simultaneous use of a number of different abilities which often develop at different rates" (Harris, 1969, p. 81). It is typically evaluated in the framework of these five aspects, which are grammar, vocabulary, pronunciation, comprehension, and fluency and as Heaton (1990) further elaborates; accuracy, range, flexibility, and size, appropriacy, amount and quality of communication, and effort. The comprehension element refers to the interdependent nature of speaking and listening as in most cases speaking generally doesn't occur in isolation except for monologues, reading aloud, telling a story, etc. (Brown, 2003). However, the impact of listening on test-takers speaking performance may lead to validity and reliability problems of the assessment. Consequently, the variables affecting learners' oral production needs to be paid due attention. In any type of assessment, the main objective is to measure the intended construct only. However, there are some factors that can influence the assessment scores or learners' performance, thereby need to be controlled. For assessing speaking skills, the factors can be characteristics of the test-taker and interlocutor, setting and timing, task

topic and format, rater performance and rating scale (O'Sullivan, 2013; Turner, 1998). Due to its nature, assessing speaking can lead to subjectivity of scoring. Therefore, an appropriate rating scale or scoring criteria should be established and employed based on the determined framework of speaking that is aimed to be achieved (Fulcher, 2003).

Luoma (2004) notes that the way an individual speaks informs the listener about the speaker's personality, attitude, and cultural background and so on. However, this poses the problem of judging the other person's pronunciation based on the assumption that it should be native-like even though determining the standard for pronunciation of a globally used language with variations in different regions is challenging. Therefore, the goal should be comprehensibility rather than accuracy of the pronunciation in real-life communicative situations. Moreover, some people prefer not to adopt a native-like pronunciation in that they might want to present their identity, not to sound pretentious, to be appreciated for learning another language, or to receive toleration in case of making mistakes that are culturally impolite as a result of their background. In addition to the concerns about pronunciation, it should be acknowledged that the use of grammar in spoken language is different from written language or even in itself whether the speech is planned or unplanned. Besides, speakers can make mistakes while speaking in a normal communication. These slips and errors in pronunciation, vocabulary choice, sentence structure may be due to not paying attention or in the case of learning another language, it may indicate lack of knowledge. Thus, assessors need to be careful with differentiating them.

Upon thoroughly considering the abovementioned issues of speaking, how assessment of speaking takes place needs to be understood. Luoma (2004) expresses that one-to-one format, paired tasks with two test-takers and group testing are the forms of assessing speaking. In addition to these, monologues can also be accounted as another form (O'Sullivan, 2012). Moreover, Brown and Abeywickrama (2010) illuminate the five types of speaking performance tasks as imitative, intensive, responsive, interactive, and extensive. Imitative speaking refers to the imitation of a word or a sentence, with which pronunciation is emphasized. An example for this type can be repetition tasks. In intensive tasks, learners are expected to produce short samples of language, which demonstrates their linguistic abilities at a certain level. They are mostly directed to the way that specified production of test-takers would occur. Thus, they are also referred to as limited-response or mechanical tasks. Reading aloud, completing the dialogue or sentence, directed response tasks, controlled picture cued tasks and sentence level translation tasks can be put under intensive

tasks. The third task type, responsive assessment tasks, differs from the previous ones in the freedom it provides to test-takers. Even though it entails some sort of interaction between the speakers, it is still limited and guided regarding length and content. To illustrate, responsive tasks can involve tasks that are; Q&As, paraphrasing, giving instructions and directions, short conversations, greetings, requests, apologies and so forth. The task types that are mentioned above are referred to as structured speaking tasks by Luoma (2004) since they focus on a limited and narrow aspect of spoken language, thereby, lacking creativity. However, scoring can be more objective and consistent as what is to be produced is predictable. The next speaking task is interactive, in which the interaction between interlocutors is more complex and longer in time compared to responsive tasks. Interaction can take place so as to exchange specific information (transactional) and maintain social relationships (interpersonal). Interviews, role-plays, games, discussions, and conversations are the common examples of interactive tasks of speaking performance. Similarly, the final task type of speaking, extensive, also requires complex and lengthy stretches of spoken language. The main difference between extensive and interactive tasks is the degree of interaction between the interlocutors, which is usually minimal in extensive tasks. Thus, oral presentations, storytelling with picture cues, retelling news or a story, extended translations fall into extensive task type incorporating a variety of monologues on the part of the speaker. As Luoma (2004) interprets, the last two task types enable test-takers to have more freedom to demonstrate their speaking skills in the given content and names them under the category of open-ended speaking tasks such as description, narration, comparison, instruction explanation, prediction, etc.

2.5.4. Assessing Writing

“While writing in a first language is a challenging, complex task, it is more so in a second language” (p. 38) is how Weigle (2002) expresses the sophisticated nature of writing. That is to say, writing incorporates a number of unique and complex skills and conventions beyond language ability that even in their own language, individuals may not be able to develop (Brown, 2003; Hamp-Lyons, 2018; Harris, 1969). According to Harris (1969), writing generally involves these components; *content* (how ideas are expressed), *form*, (how content is organized), *grammar* (how grammatical and syntactic elements are employed), *style* (how linguistic elements are utilized based on writers’ stance) and *mechanics* (how spelling and punctuation are used). Similarly, Heaton (1990) lists those components as;

treatment of content, stylistic skills, language use, judgement skills and mechanical skills, which may be different in their name, but refers to the same meaning of what writing skills consist of. Weigle (2012) proposes that learners' writing skills in their own language is not the primary concern of foreign language education. However, there are some factors that influence learners' abilities in their first language and cannot be distinguished from second/foreign language writing, and these issues will determine whether to involve conventions and strategies of writing in the assessment or not. For instance, writing with young learners, or older learners in some cases, will be limited to basic literacy skills since their literacy skills may not be developed in their first language yet. Therefore, the age of the learners is a vital factor. Another factor is the learners' proficiency level in that in the earlier process of foreign language education, the focus of writing may be on the linguistic structures such as grammar and vocabulary. As the learners get more proficient, the focus of writing can be broadened to other components such as style, organization, and content. Moreover, learning context whether second or foreign language is taking place has also an impact on whether language or writing, or both will compose assessment. In second language learning where learners are exposed to target language outside the class authentically, they need to be proficient in writing conventions and genres since they are required to write for work or school. In short, age, learning context and level of proficiency, among other concerns, needs to be considered for teaching and assessing writing.

Weigle (2002) urges that some issues that needs to be regarded while designing a writing task, which are subject matter and choice of tasks, stimulus material (prompt), genre, timing, instructions. First of all, the subject or content of the writing should be accessible to all learners; that is, it should enable learners to write with a general knowledge. Subject areas that require a background knowledge on a specific domain would raise a problem regarding validity. These aspects are not in the construct definition since the main objective of assessing writing is to assess learners' writing abilities, not their knowledge or creativity for that matter (Hughes, 2003). Similar to this concern, whether to provide learners a chance to select among various options is another issue in designing writing tasks. Even though it may enhance validity by eliminating the background impact as learners would choose according to their knowledge and interests, the equality of the tasks in their difficulty and the time that learners spend to decide on the task may pose a problem for reliability. Furthermore, the prompt of the writing tasks is expected to evoke some sort of background knowledge and support learners with ideas to write about. However, as the written prompt necessitates

reading skills, an irrelevant construct, it can be problematic for poor readers; consequently, validity is at risk. The prompt directs the learners to write on a particular genre that is intended. There are various genres of writing, which are academic writing (compositions, essays, subject reports, etc.), job-related writing (letters/e-mails, schedules, project reports, announcement, etc.) and personal writings (messages, invitations, diaries, etc.) (Brown & Abeywickrama, 2010). In addition to genre, timing should be arranged properly so that learners can have adequate time to write. In some cases, when writing takes too much time in a class environment, writing tasks can sometimes be assigned as homework. Moreover, the instructions should be clearly explaining the task in a simple way and what is expected from the learners. They should be neither too short to leave any important detail nor too long to take too much time of the administration. Additionally, Hughes (2003) adds that one task is not enough to collect sufficient evidence of learners writing skills; therefore, employing as many tasks as possible will support the reliability of writing assessment. To that end, portfolio assessment can be adopted to collect varying samples of the learners' writings periodically.

Along with the aspects above, another crucial aspect of writing assessment is the scoring. Scoring learners' performance can constitute a problem in terms of objectivity, thereby, validity and reliability. Thus, pre-determined scoring criteria and scoring rubrics can be adopted. There are two approaches two scoring of writing: holistic and analytic scoring. Holistic scoring or impressionistic scoring refers to assigning a single score on learners' written works based on an overall impression (Hughes, 2003). One or more than one rater can be involved in scoring for the sake of reliability (Heaton, 1990). In analytic scoring, various elements of writing, such as vocabulary, content, organization, etc., are scored separately, which can be beneficial for diagnosing the problematic areas of the learners (Hughes, 2003). Besides, Heaton (1990) asserts that analytic scoring could be adopted when scoring by multiple raters is not achievable. Weigle (2012) explains the advantages and disadvantages that both approaches and related rubrics have. Holistic scoring can be useful when the amount of writings needs to be scored in a limited time, whereas analytic scoring is efficient for providing useful evidence about learners' strengths and weaknesses.

Hughes (2003) disagrees with the tendency to assess writing skills indirectly as they will not accurately demonstrate learners' writing abilities. In order to understand someone's writing abilities, the most accurate way is to ask them to write something, namely, direct assessment. There are many types of tasks that can be utilized in the assessment procedures. Among the

four types writing performances (Brown & Abeywickrama, 2010), the first one is imitative, which is concerned about the basic literacy skills of spelling, punctuation and so on. The context and meaning are of secondary importance, whereas form is initially aimed. The second one is intensive or controlled writing. In this type, the main goal is to use lexical items properly in the appropriate context. The form is still the primary concern although meaning and context is to some degree. Responsive performance is the next one, in which written product is longer with a limited discourse level. Learners have more freedom to express their ideas. Therefore, content and meaning are strongly emphasized, while form is at the discourse level. The final type, extensive, refers to the mastery of all writing purposes with successful strategies employed. Learners aim to achieve a well-organized, logically developed written work with detailed ideas and various linguistic features.

2.5.5. Assessing Young Language Learners

Young language learners are described as pupils, generally aging between five and twelve, “who are learning a foreign or second language [...] during the first six or seven years of formal schooling” in primary or elementary school context (McKay, 2006, p. 1). They differ from adult learners in their attention spans, emotions, interactions, attitudes and behaviours, and reactions in addition to their linguistic and cognitive capabilities (Hasselgreen, 2012). Therefore, assessment of young learners needs to be considered separately from adolescents and adult learners in reflection of their developmental and curricular needs (Bailey, Heritage, & Butler, 2013). Young learners need language input and tasks regarding their age as well as short-term motivation, which is a crucially important factor when assessing younger learners (Hasselgreen, 2005). Besides, activities should be motivating especially for young learners, though it is still a necessity for all age groups, so as to diminish the negative effects of examination and reduce the anxiety. Another significant factor in the assessment of young learners is put forth as the context (Bailey, Heritage, & Butler, 2013). It is significant that various opportunities are provided to the learners to express their skills and knowledge in various contexts; otherwise, it would not be appropriately assessed within limited contexts. In addition, since their developmental process may vary according to the pupil and his/her unique needs, homogeneity of the test-takers may not be applicable to young learners. Thereby, individualized assessment procedures could be required.

McKay (2006) summarizes the characteristics that differentiate young learners from older learners as growth, literacy, and vulnerability. Their growth in terms of cognitive, social,

emotional, and physical is an important factor on the assessment procedures to be undertaken. For their cognitive characteristics it can be put forth that they can get distracted by their peers or other factors since their attention span is shorter than older learners. Besides, the difficulty level of the task may demotivate young learners to perform the task and cause them to quit. They, more frequently, learn from direct experiences as in visual aids and objects. However, most of them broaden their understanding of the abstract notions more as they develop. The content of the assessment should be carefully considered in that if the content extends beyond learners' knowledge, the results may not indicate their language ability since the learners' knowledge is limited to their experience of the world. Other than their cognitive growth, young learners develop socially and emotionally as well. The consideration of children's social and emotional development is crucial since it impacts on how they perform in a group or peer work and individual tasks, how they perceive their success, and how they are motivated, which are in relation with their personal characteristics whether introvert or extrovert, past experiences, familiarity with the content, and the type of feedback provided. In addition, physical characteristics are another factor to be kept in mind when assessing young learners. Since they are very active and enthusiastic about physical activities and playing games, physical tasks should involve language related response, such as pointing, colouring, drawing, and acting, in order to engage them in the task with the elements of fun and fantasy.

Another difference between younger and older learners, as McKay (2006) asserts, the literacy of young learners. Young language learners try to develop their literacy skills in their own language and target language simultaneously. Therefore, it is suggested to make judgements of the learners' performances according to their literacy development in their own language and target language because the former may have an influence on the latter. Finally, McKay (2006) propounds that young learners tend to be more vulnerable to failure and criticism and their experiences at school may influence their self-esteem. Thus, it is vital to provide a secure and supporting environment and approach, where their success is praised and appreciated so as to prevent loss of motivation and self-esteem, and as a result, to inhibit long-term effects.

Similarly, Bailey (2017) emphasizes the significance of developmental and contextual factors to be considered when assessing young language learners. She suggests that the test format and duration, test items and task types, task content change based on the characteristics of the learners.

Hasselgreen (2005, pp. 338-339) synthesizes a consensus on what procedures should be followed to assess young learners.

- Tasks should be intriguing for the particular age group, engaging and appealing with fun and game elements.
- Various types of assessment should be engaged in the process concerning the perspective the learners, parents, and teachers.
- Learners' strengths should be emphasized while designing the tasks and feedback.
- Learners should be assisted in performing the task.
- Teachers should be assisted so as to be aware of the language assessment methods and be familiar with basic assessment criteria.
- Assessment tasks should incorporate learning purposes.

The main goal of language teaching at early ages is to promote positive attitudes towards language learning (Hughes, 2003). However, in some cases when the effectiveness of the program and diagnosing young learners' success need to be ensured, testing may be employed even though the impact of inappropriate testing could damage learning or attitudes of the pupils. Therefore, if any testing is to take place for the sake of validity, reliability, and accountability purposes, it should be integrated with assessment and learning. Another recommendation that Hughes (2003) provides is that feedback should be immediate in order to maximize the outcomes, and also positive so that the negative impact of assessment is reduced. The last recommendation is about self-assessment in that allowing learners to monitor their own progress will develop their awareness about their achievements, and therefore, self- assessment should be embedded in educational programs.

In light of these considerations, assessment of young learners can involve incidental and planned observations, on-the-run assessment, crafts and projects, self and peer-assessment, classroom tests and tasks, record-keeping, portfolios as well as teachers' files to observe students' progress (McKay, 2006). In addition, there are a great deal of tasks for assessing young learners, which are storytelling, describing a picture, categorization tasks, presentations, question and answer, role-plays, interviews, information-gap tasks, pair works and group works, physical response tasks, true/false tasks, matching, finding mistakes/differences, dictation, gap-filling tasks, completing a story, re-forming a text, open-

response writing tasks (McKay, 2006). Overall, when assessing young language learners, developmental characteristics as well as contextual factors need careful consideration.

2.6. Assessment Literacy

Assessment is the process of gathering information about the learners' achievement so that instructional decisions can be made, and the efficacy of these decisions depends on the quality of the assessment data (Stiggins, 1991). That is, the information obtained from a poor assessment practice would be useless, hence, the decisions. Therefore, teachers need to be able to "understand, analyze, and apply information on student performance to improve instruction" (Falsgraf, 2005, p. 6) because "assessment, when used appropriately, is a powerful tool for learning" Popham (2011, p. 265). Assessment Literacy (AL), a term that is coined by Stiggins (1991), is defined as teachers' capability of understanding the purposes of assessment, using appropriate assessment methods and tools, administering, and making interpretations of assessment results (Inbar-Lourie, 2008) and making instructional or institutional adjustments based on the assessment information (Boyles, 2005). According to Popham (2011, p. 267), AL involves "an individual's *understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions*". Stiggins (1995) simply defines AL as the ability to understand the differentiation of 'sound and unsound assessments' (p. 240).

In a regular classroom, 50% of the instruction time is spent on the activities that are related to assessment (Stiggins, 1991). Similarly, Popham (2011, p. 267) sees assessment among teachers' 'day-to-day responsibilities' with the purpose of fulfilling these requirements appropriately, teachers should be assessment literate. Thus, the significance of assessment literacy cannot be underestimated. As Plake (1993) touches upon this by stating that teachers' competency in assessment-related procedures is vital since their decisions whether about teaching (instructional) or learning (evaluative) will depend on their level of competency. To put it differently, AL is mutually beneficial for both teachers and learners (Popham, 2009). In a similar vein, Coombe et al. (2012) underline the importance of developing assessment literacy since it enables teachers to improve their instruction by utilizing the assessment data as well as to decide on the proper assessment methods and techniques for a specific purpose such as achievement or placement tests. Therefore, in pre-service education, AL needs to be considered as important as instructional competence (Popham, 2011) because no matter how efficient the instructional methods are, learners'

academic achievement cannot be attained when assessment is not effective and proper (Stiggins, 1995). It can be inferred that teachers' higher level of AL will yield to better learning as a consequence (Plake, 1993). According to Popham (2011), there are two main reasons that why teachers need to be assessment literate. The first one is to improve instruction by modifying it in accordance with the assessment-driven adjustments and by diagnosing learners' strengths and weaknesses. The second one is for accountability purposes, which is the responsibility that teachers hold in order to inform others (parents, administrators, institutions, etc.) about learners' academic achievement. It is clear that assessment literacy is a must for teachers since assessment literate teachers will make more sound decisions based on sound assessment procedures (Popham, 2009).

Stiggins (1991) asserts that levels of assessment literacy are threefold based on who is concerned about the assessment. The *functional level*, which indicates the ability to interpret assessment data, consists of people who use the assessment data that others generate, such as students, parents, and public officials. The second level is *practical level*, concerning teachers and others education professionals who generate and use assessment data. They are expected to have higher understandings of all assessment related procedures at this level, such as the impact of assessment, possible issues that may occur and the ways to prevent them, designing and administering assessment tasks, giving feedback and so forth. Assessment literacy at *advanced level* is the final level that involves stakeholders who are concerned with the generating the data to be used by others. Educational specialists designing high-stakes exams or evaluating educational programs, who are expected to master assessment issues the most, can be considered at this level.

In the literature, there have been many studies to refer to what AL comprises of, and the skills that teachers need to have for that matter. To begin with, American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and National Education Association (NEA, 1990) aimed to prove a guideline for teachers, teacher educators and other educational specialists. The seven standards of teacher competencies regarding assessment are presented in their study of 'Standards for Teacher Competence in Educational Assessment of Students' (*the Standards*). The competencies that teachers should possess are: (a) choosing assessment methods appropriate for making instructional decisions; (b) developing assessment methods appropriate for instructional decisions; (c) administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods; (d) using assessment results when making decisions about

individual students, planning teaching, developing curriculum, and school improvement; (e) developing valid pupil grading procedure which use pupil assessments; (f) communicating assessment results to students, parents, other lay audiences, and other educators; (g) recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (pp. 31-32). However, there have been many changes in the educational field since *the Standards* were established. Thus, Brookhart (2011) believes that they need to be updated in order to subsume the current notions of formative and accountability assessment, which is believed to be lacking in *the Standards*. She lists 11 statements consisting of assessment knowledge and skills, some overlapping with the ones in *the Standards*, for teachers from the perspective of formative assessment as in providing feedback and involving learners in self-assessment procedures.

McMillan (2000, pp. 1-3) offers 11 principles of what teachers and administrators should be aware of so as to adopt good assessment practices whether classroom assessment or standardized tests: a) assessment is inherently a process of professional judgement; b) assessment is based on separate but related principles of measurement evidence and evaluation; c) assessment decision making is influenced by a series of tensions (formative vs. summative, standardized vs. classroom tests, criterion-reference vs. norm-referenced, traditional vs. alternative, etc.); d) assessment influences student motivation and learning; e) assessment contains error; f) good assessment enhances instruction; g) good assessment is valid; h) good assessment is fair and ethical; i) good assessments use multiple methods; j) good assessment is efficient and feasible; k) good assessment appropriately incorporates technology. In a similar way, Stiggins (1995) delivers a list of five standards for quality in assessment. For higher quality in assessment, he suggests that teachers should begin with clear purposes, address achievement targets clearly and properly, choose or design appropriate methods of assessment, broaden the sampling of student achievement and prevent bias and distortion.

Despite of these attempts and guidelines that are provided for teachers, they are believed to lack the qualifications of AL. Stiggins (1991, p. 535) makes a comment on this issue as “we are a nation of assessment illiterates”. Confirming this statement, Popham (2009, p. 5) utters that “test is a four-letter word, both literally and figuratively”, emphasizing the lack of knowledge in assessment. Plake (1993) links the reason for the inadequate knowledge in assessment of teachers to their insufficient pre-service training that they have received. However, it can be concluded that assessment illiteracy is not a risk teachers can take

(DeLuca & Klinger, 2010; Popham, 2009). Addressing the fact that AL is critical for teachers with respect to language education, Coombe et al. (2012, p. 25) assert that “in the view of the importance of assessment in today’s educational institutions around the world “assessment literacy” is a necessity for all language educators”.

2.7. Language Assessment Literacy

Language Assessment Literacy (LAL), derived from assessment literacy (Stiggins, 1991), is a term used in general education that is typically used to refer to the knowledge and skills on how to perform assessment-related activities. Inbar-Lourie (2008, p. 389) defines assessment literacy as; in order to be assessment literate, one should have “the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results”. LAL contains most of AL features yet needs to be differentiated since it requires additional competencies specific to language assessment such as designing, administering, interpreting, utilizing, and reporting language assessment information for different purposes. (Inbar-Lourie, 2013b).

LAL is relatively a novel concept and “research into assessment literacy is in its infancy” (Fulcher, 2012, p. 117). As a consequence, the construct and the theoretical framework of LAL in general, and LAL for teachers and other stakeholders in particular is still an ongoing issue due to the fact that assessment is to be shaped by the dynamics of unique contexts (Giraldo, 2021; Stabler-Havener, 2018). However, regardless of the type of stakeholders and context, LAL demands a mastery of essential components, which might comprise of knowledge about language, assessment, the context, and the ability to facilitate assessment related procedures such as designing, collecting, administering, or interpreting assessment information as a means of making sound ethical decisions (Inbar-Lourie, 2013a).

The level of assessment literacy may vary for each stakeholder with respect to the context and the extent of stakeholders’ involvement in assessment (Taylor, 2009). Jeong (2013) aimed to define if LAL differs according to two different stakeholder groups, language testers and non-language testers, based on the content of their trainings in testing and assessment. Even though the course structures and the content of the courses didn’t indicate a significant difference, language testers were concerned with theoretical and statistical aspects more while non-language testers valued the practical aspects of assessment more

such as classroom and alternative assessment. As can be seen, LAL needs may differ for stakeholders relative to their interest and involvement in testing and assessment though a basic understanding of LAL still needs to be held on the part of every stakeholder. Thereupon, Taylor (2009, p. 27) remarks upon this issue by stating that “training for assessment literacy entails an appropriate balance of technical know-how, practical skills, theoretical knowledge, and understanding of principles, but all firmly contextualized within a sound understanding of the role and function of assessment within education and society”.

2.7.1. Definitions of LAL

It has been acknowledged that the conceptualization of LAL is mostly drawn from AL literature while containing distinctive qualifications of “a knowledge base that incorporates unique aspects inherent in theorizing and assessing language-related performance” (Inbar-Lourie, 2017, p. 259). Even though the concept of LAL is rather new, there is a growing rate of research in this area. Therefore, there have been many attempts to define and conceptualize what LAL entails for language teachers and other stakeholders (Giraldo, 2021; Inbar-Lourie, 2017). In the literature, the focus of LAL concept is on the language teachers since they are considered as the core stakeholders of language assessment (Giraldo, 2018). The reason could be linked to the influence of teachers, the most crucial factor, on student learning; therefore, it is vital for language teachers to develop their LAL (Scarino, 2013). Consequently, the definitions of LAL have been shaped around the teachers’ assessment procedures from a classroom-based standpoint.

To begin with, Taylor (2009, p. 24) defines LAL as “the level of knowledge, skills, and understanding of assessment principles and practice that is increasingly required by other test stakeholder groups, depending on their needs and context”. Similarly, Malone (2013, p. 329) considers LAL as “language instructors’ familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language”. To put it simply, LAL is putting theory into practice.

Furthermore, an individual involved in language assessment should be familiar with the current perspectives in language teaching pedagogy including validity and reliability concerns, the ways and methods to properly conduct assessments and interpret assessment results ethically whether large-scale or classroom assessment along with the rationale and justification of the assessment procedures undertaken within a wider concern of societal

aspects (Inbar-Lourie, 2008). In its essence, Inbar-Lourie (2008, p. 390) considers LAL as “understanding the ‘what’ and performing the ‘how necessitates appreciation of the background and reasoning behind the actions taken, that is, the ‘why’”. Moreover, individuals who are considered as language assessment literate should be capable of applying the knowledge of theory and practice in sound assessment procedures in order to promote decision making process. In a similar vein, Vogt and Tsagari (2014, p. 377) describe LAL as “the ability to design, develop, and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade, and score assessments on the basis of theoretical knowledge”.

The increased attention given to assessment within the socio-cultural standpoint can be observed in the definitions of LAL. For O’Loughlin (2013), LAL incorporates “the acquisition of a range of skills related to test production, test score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within education and society” (p. 363). In the same respect, another definition of LAL put forth by Lam (2014, p. 172) is “teachers’ understanding and mastery of assessment concepts, measurement knowledge, test construction skills, principles about test impact, and assessment procedures which can influence significant educational decisions (evaluation of student learning) within a wider sociocultural context”.

Apart from familiarity and knowledge of assessment concepts, Scarino (2013) stresses that teachers should also acknowledge their role as teacher-assessor. According to Scarino (2013), LAL entails teachers to transfer their theoretical knowledge regarding language assessment into their classroom practice, thereby becoming more conscious of how they conceptualize assessment in addition to becoming “increasingly aware of their own dynamic framework of knowledge, understanding, practices and values, which shape their conceptualizations, interpretations, judgements and decisions in assessment and their students’ second language learning” (p. 311). However, it should be understood that what and how much is needed to be acquired by different stakeholders or whether this knowledge is related to theory or practice and to assessment literacy in general education as opposed to language assessment is still questioned (Inbar-Lourie, 2013a; Scarino, 2013).

Teachers are not the only stakeholders who are concerned with and involved in assessment related procedures. Thus, based on their level of involvement, the meaning and needs of LAL may vary for different stakeholders. Pill and Harding (2013), taking various stakeholders of language assessment into consideration, define LAL in a broader sense as

“indicating a repertoire of competences that enable an individual to understand, evaluate and, in some cases, create language tests and analyze test data” (p. 382).

Finally, Fulcher (2012) provides a wider definition of LAL based on the finding of his research project with EFL teachers. Adopting a procedural approach to LAL, he defines the components of LAL as the practices, principles, and contexts, and by extension, LAL is:

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. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (Fulcher, 2012, p. 125)

In conclusion, based on the definitions, LAL can be synthesized as an individual's familiarity, understanding, awareness, knowledge, skills, abilities as well as principles and practices about language assessment-related procedures in line with their specific contexts and the current theoretical concerns of foreign language education within broader social frameworks. Besides, as stressed by many scholars, the theoretical knowledge of assessment is to be reflected in teachers' practice.

2.7.2. LAL Models-Components of LAL

Influenced by AL in the field of general education, the conceptualization of LAL has been comprehensively studied in the literature. The contributions that the researchers have made in the language assessment field so as to set the ground for what constitutes LAL have been discussed by Stabler-Havener (2018) and four main LAL models from influential scholars have been illuminated; a) Brindley's (2001) five component model for professional development; b) the what, how and why (Inbar-Lourie, 2008) and skills, knowledge and principles model (Davies, 2008) based on the textbooks of language assessment; c) Fulcher's (2012) model of practices, principles and contexts; d) the models for LAL stakeholder profile (Pill & Harding, 2013; Taylor, 2013).

Aiming for the professional development of language teachers in assessment, Brindley (2001) has come up with a model that involves two core and three optional units. The first core unit deals with the social context of assessment in which accountability, ethics, and the society's perceptions of assessment within the social, educational, and political perspectives in accordance with the unique context of teaching/learning. The second core unit is related to how language proficiency is defined and described with the aim of offering a basis for

sound assessment procedures such as designing instruments and interpretation of the results. In the interest of implementing good assessment practices, teachers bear the responsibility of being aware of the existing approaches to and theoretical models of language construct. The other units that are referred to as optional address to the construction and evaluation of language tests, integration of assessment into the curriculum and curriculum-related assessment, and reflecting assessment theory and knowledge into practice, all of which should be pertained peculiar to the distinct needs and circumstances of the educational context.

Inbar-Lourie (2008) asserts that even though language assessment knowledge involves some components derived from assessment knowledge in the field of general education, it is a different concept in that it entails language specific competences. Her conceptualization is based on three questions regarding language assessment, which are *what*, *how*, and *why*. The ‘what’ is about the language skills, traits or components to be assessed, which leads to the differentiation between AL and LAL. In order to design proper measures according to ‘what’ language component to assess, teachers should be informed about the current theories and approaches to language teaching. The ‘how’ is related to knowledge about different methods and approaches to assessment and implementing them. As a result, teachers can choose and implement the appropriate methods and techniques based their knowledge of ‘how’. The final component of language assessment knowledge is the ‘why’, referring to the foundation or rationale that the assessment is based on. To sum up, LAL involves the knowledge of what to assess, how to assess and why to assess so that language teachers can design their assessments in line with proper methods that fit for purpose.

Aiming to define the current trends of the time and the changes in the textbooks about language testing and assessment, Davies (2008) discerns that there is a movement from *skills* and *knowledge* to *skills*, *knowledge* and *principles*. Covering the influential textbooks since Lado’s (1961) *Language Testing*, Davies (2008) relates the *skills* with the implementation of assessment in the matter of providing training in proper methodology and tools within psychometric issues such as statistical analysis, test development, item writing along with the programs that aid these procedures. The *knowledge* refers to establishing the framework for measurement pertaining to how language is described and what kind of models are employed for testing/assessment in line with language learning/teaching models whether communicatively, performance-based or within socio-cultural perspectives in addition to

considering the context. Finally, the *principles* are interested in the appropriate use of instruments and tools as well as ethical issues, fairness, and impact.

Based on the data of his research project concerning the LAL need of language teachers, Fulcher (2012) formulated a model of LAL that comprises of three components: *practices*, *principles*, and *contexts*. The diagram of these components indicates a hierarchical order, where *practices* are at the bottom as the fundamental dimension for LAL involving the knowledge, skills, and abilities in assessment procedures such as constructing and administering tests and interpreting the results. The following dimension is the *principles*, which is concerned with being conscious about principles, processes, and concepts guiding the practices. The final dimension in the diagram, the *contexts*, comprises of larger frameworks within historical, social, philosophical, and political perspectives to determine the origin, the justification, and the impact of assessment procedures. However, Fulcher (2012) notes that the level of the components may vary according to the type of the stakeholder except for language teachers since the data contributing to the model were derived from language teachers' LAL needs.

Although it has been acknowledged that LAL needs and competencies may differ for stakeholders (Fulcher, 2012; Taylor, 2009), what constitutes the needs and competencies for each stakeholder is not clearly established. Specific to this concern, Pill and Harding (2013) aimed to figure out the misconceptions in LAL and the level of expertise needed for non-practitioners. Applying the assessment literacy from other domains, such as mathematics and science, other foreign language education, their model of LAL indicates five levels of LAL proficiency in a continuum. The levels are *illiteracy* (lack of knowledge), *nominal literacy* (fundamental understanding with some misperceptions), *functional literacy* (sound understanding), *procedural and conceptual literacy* (knowledge in concepts and practice), and *multidimensional literacy* (a broader sense of knowledge in historical, social dimensions). Even though this model incorporates a progressive approach to LAL with various aspects, it is criticized for overemphasizing theoretical aspects while being superficial in practical concerns in addition to being deficient in specifying the level of proficiency in LAL that stakeholders should possess (Harding & Kremmel, 2016).

Concentrating on different stakeholder groups, Taylor (2013) proposed a model for stakeholders who require differing levels of LAL depending on their needs and context. In the model, there are three layers; *core* (test makers, researchers), *intermediary* (language teachers, course instructors) and *peripheral* (general public, policymakers). The roles and

the responsibilities of the stakeholders determine what kind of competencies they should possess since everybody doesn't need to know or do everything at the same level. For example, for stakeholders such as test makers, who are at the *core* level of assessment, extensive training in measurement theory, ethical principles and technical knowledge are expected whereas for language teachers, who are at the *intermediary* level, LAL can focus on practical aspects of assessment for designing tests while ethical principles and measurement theory stay at the surface level with a lighter emphasis. With reference to this perspective, Taylor (2013) propounds a figure that intends to represent the competencies each bearing a different level of interest for four stakeholders; a) test writers, b) classroom teachers, c) university administrators and d) professional language testers. LAL profiles in the figures subsume eight dimensions of *language pedagogy, knowledge of theory, principles and concepts, technical skills, sociocultural values, personal beliefs/attitudes, local practices, and scores and decision making* (p. 410). For instance, language teachers' profile necessitates competency in language pedagogy the most, and it is followed by technical skills, sociocultural values, local practices and their beliefs and attitudes. Even though Taylor's (2013) model provides a foundation for the competencies of different stakeholders, it has been debated over its deficiency in defining the language teacher profile (Harding & Kremmel, 2016) in that the LAL level could be more advanced in scores and decision making and local practices than she presents (Stabler-Havener, 2018).

Apart from the abovementioned frameworks of LAL, Kremmel and Harding (2020) aimed for a comprehensive model to define what LAL components are needed across different stakeholder groups. For this purpose, following a data-driven approach, they developed the Language Assessment Literacy Survey in order to evaluate Taylor's (2013) model for LAL stakeholder profile and to what extent LAL components diverge for various stakeholders including teachers, developers of test/assessment, researchers of testing/assessment, professional raters/examiners, test-takers, test score users, policymakers, and parents/legal guardians. However, the demonstration of the stakeholders' perceived needs only included the first three key stakeholders. The results indicated nine distinct LAL components, which are mostly in agreement with Taylor's model. However, it is noted that some components need to be distinguished and expanded to: *assessment in language pedagogy, developing and administering language assessments, statistical and research methods, assessment policy and local practices, assessment principles and interpretation, personal beliefs and*

attitudes, language structure, use and development, scoring and rating, and washback and preparation.

In conclusion, the early models of LAL are generally shaped around three dimensions while the more current models are more comprehensive in their conceptualization of LAL components. The components of LAL fundamentally comprise of knowledge, skills and abilities, principles, practices of language pedagogy, language teaching/learning approaches, assessment methods and concepts, designing and administering, statistical procedures, scoring, rating, giving feedback, interpreting the results, washback, and socio-cultural aspects and values, all of which are likely to be shaped by the teachers' personal beliefs and attitudes as well as their own educational contexts.

2.7.3. The Importance of LAL

As a consequence of the integral relationship between teaching and assessment, teachers are involved in both. They make decisions about the content of the course, teaching tasks and materials, choosing appropriate methods for the needs of their learners to facilitate learning as well as observing learners to make judgements about their progress and achievement (Rea-Dickins, 2004). Hence, teachers hold the dual role of teaching and assessing (Inbar-Lourie, 2008; Rea-Dickins, 2004). Therefore, this role requires teachers to be more knowledgeable of assessment related concerns, which helps them not only pursue more effective assessment procedures, but also evaluate their own instruction and construct appropriate assessments that motivate learners in their learning process (Rogier, 2014). High levels of LAL indicate a mutual benefit for both teachers and learners in that it provides feedback about both teaching and learning (Popham, 2009). In order to improve instructional quality and learner achievement in addition to determining appropriate methods and techniques of assessment for a particular purpose, teachers need to develop their LAL (Coombe et al., 2012).

In a similar vein, Herrera and Macías (2015) suggest that teachers should possess the capacity to correlate their assessment practices to approaches to language teaching, construct proper assessments, choose the best assessments that suit the purpose, and be aware of the impact of large-scale examination if they have high LAL competence. Moreover, as far as accountability is taken into consideration as in communicating the results with other related stakeholders more effectively, LAL will contribute to higher test validity and more

transparent procedures (Coombe et al., 2012). Even though the significance of LAL for teachers is established in the literature, they are not believed to be qualified in assessment related procedures (Plake, 1993; Popham, 2009; Stiggins, 1991). Thus, it is highly vital for teachers to receive trainings about testing and assessment procedures for improving their knowledge of assessment and developing their LAL (Popham, 2011). However, Herrera and Macías (2015) discuss that though pre-service trainings are beneficial, they are not enough since they may be highly dependent on textbooks. Therefore, they advise that continuous professional development programs for both pre- and in-service teachers should embrace alternative approaches to enhance LAL of teachers, such as workshops, conferences, study groups, and research projects that involve up-to-date developments in LAL which can be incorporated with classroom practice based on the needs and educational contexts of teachers.

2.7.4. Language Assessment Knowledge and Practice

It is clear in the literature that teachers need language assessment knowledge to be able to perform assessment-related procedures as a means to enhance the quality of language education (Coombe, Vafadar, & Mohebbi, 2020). In Xu & Brown's (2016) conceptual framework of TALiP (Teacher Assessment Literacy in Practice), the knowledge base is presented at the bottom of a pyramid figure as a base for other components in the framework that they have developed by scrutinizing a great deal of both empirical and non-empirical, on AL of teachers. In conclusion of their analyses and syntheses, TALiP is presented in a pyramid figure consisting of seven components; from bottom-up, they are knowledge base, interpretive and guiding framework, teacher conceptions of assessment, contexts (macro – socio cultural, micro – institutional), teacher assessment literacy in practice, teacher learning, and assessor identity (re)construction. Even though knowledge base is considered as an insufficient component for AL in practice, it is still believed to be a necessary facet that has a great role for AL to be reflected in assessment practices and teachers' role as an assessor. Therefore, the knowledge base covered in the trainings and programs for professional development needs to be able to broaden teachers' conceptualization and practices of assessment in light of the current approaches and policies with respect to different educational contexts (Coombe, Vafadar, & Mohebbi, 2020).

Similarly, though the importance of social and contextual perspectives that assessment has gained recently is stressed and believed to have an impact on assessment literacy and

practices (Yan, Zhang, & Fan, 2018), assessment epistemologically contains a universal knowledge that should not be degraded to the social standpoint of the individuals or groups (Fulcher, 2012). Thus, without forming a common understanding of the knowledge base of assessment, evaluation of the appropriacy of assessment practices would be pointless since the assessment criteria or standards are vague, which may eventually result in failure of the teaching/learning process (Fulcher, 2012).

In the definitions of assessment that have been mentioned above, assessment is simply reflecting the theoretical knowledge in assessment practices (Malone, 2013). Therefore, it is critical for teachers to possess not only a knowledge of assessment but also the skills to project that knowledge in their practices.

2.8. Studies on Language Assessment Literacy/Knowledge and Practices in the World

Language assessment is an indispensable part of language education. Thus, language teachers are required to possess certain knowledge and skills to be able to perform sound assessment practices, which is expected to lead to better learning. Since Fulcher (2012) expressed that research regarding LAL is at the very beginning, more researchers have been interested in LAL research. Even though there is a remarkable number of studies on LAL, it still needs to be comprehensively reviewed from different perspectives (Coombe, Vafadar, & Mohebbi, 2020). Research into LAL have been subjected around language assessment knowledge, training needs, factors contributing to LAL, perceptions, and practices.

Many researchers sought to identify the EFL teachers' needs regarding their LAL. First of all, focusing on LAK, Tavassoli and Farhady (2018) conducted a needs analysis targeting to expose the gaps in EFL teachers' language assessment knowledge based on their perceptions in consideration of their age, gender, experience, type of undergraduate degree, educational background, pre-service training, and type of their institution. It is shown that majority of the teachers needed to enhance their LAK though they were aware of the essential components of assessment and had an average degree of perceived self-competency. Besides, teachers with MA degree compared to BA teachers attributed more importance to assessment principles such as validity, which indicates that EFL teachers' perceptions and preferences are shaped by their educational background. In addition, teachers' perceptions of their assessment needs varied depending on their teaching context, whether state or private institutions. Based on their findings, Farhady and Tavassoli (2018) developed a knowledge

test based on the needs analysis that they conducted (Tavassoli & Farhady, 2018) in order to reveal EFL teachers' actual LAK level as opposed their perceptions about their knowledge of language assessment. The LAK test was administered to 164 Iranian EFL teachers with varying experience levels, educational background, type of undergraduate degree, and teaching context as in private or state institutions. The results demonstrated that LAK level of EFL teachers was insufficient despite their perceptions about their knowledge. Demographic factors of gender and type of their undergraduate degree whether English language teaching, English literature, or others had a possible impact on teachers' LAK level. However, the teaching context (state/private) and experience level influenced EFL teachers' LAK level in favor of the more experienced teachers and state institutions. The final study in their sequence of research on LAK, Farhady and Tavassoli (2021) aimed to reveal how EFL teachers of primary, secondary, and tertiary level perceived their LAK and utilize their knowledge in their instruction. 21 EFL teachers with high and low levels of LAK, which was determined by the LAK test they developed in their previous study (Farhady & Tavassoli, 2018), were interviewed so as to demonstrate their perceptions while the data on their utilization of LAK in classroom practice was obtained from the test samples that the participants sent. In the interviews, it was observed that EFL teachers indicating both high and low level perceived their LAK inadequate to take on the current notions in language assessment although they indicated a certain degree of knowledge in fundamental assessment concepts; therefore, there was no major differences in their perceptions of LAK whether they had high or low LAK level. When EFL teachers with high and low LAK level are compared in their utilization of assessment in their classroom practices, no major differences between them were observed except for minor variations. They both assessed all language skills with various assessment techniques and instruments besides interpreting assessment results to improve the problematic areas in learners' performances while ignoring reliability and validity. However, when the test samples that EFL teachers use were analyzed, low LAK group employed a limited variety of task types, such as gap-filling and completing the sentence whereas high LAK group adopted tasks such as editing, Q/As, true/false items, matching, gap-filling, and multiple choice. In conclusion, it was revealed that having a higher level of LAK had an effect on the way EFL teachers use and reflect their knowledge in their assessment, particularly in test development procedures. Overall, the study revealed that EFL teachers' LAK was not transferred into their assessment practices regardless of their knowledge level.

In his research project, Fulcher (2012) aimed to investigate the language teachers' needs in the context of testing and assessment and use the data to make a contribution to the concept of AL as well as delivering sources for language education programs for testing through constructing new materials. In line with these purposes, he developed a survey that was piloted with 24 international language teachers. After the revision of the initial survey, 278 language teachers and participants who were interested in language assessment participated in the online survey consisting of four factors: large-scale standardized testing, test design and development, validity and reliability, and classroom testing and washback. It was demonstrated that the teachers were highly aware of their needs which are the skills, knowledge, and principles based a procedural approach that balances the classroom and normative assessment within a wider social, historical, and ethical perspective. Additionally, the participants expressed that the textbooks should embrace a practical concern including a glossary and activities as well as additional resources for further exploration, such as links to websites. By virtue of this study, Fulcher (2012) offered an expanded definition of AL that embodies practices, principles, and contexts. Additionally, Muhammad and Bardakçı (2019) aimed to demonstrate Iraqi EFL teachers' self-perceived AL, their opinions about their pre-service education, and strengths and weaknesses in their knowledge regarding assessment. EFL teachers believed that their pre-service education trained them as good assessors and felt competent on the contrary to their assessment literacy, which was found notably low. The reason is linked to the quality pre-service education, lack of opportunities for professional development, and the importance placed on the national testing procedures in educational policies. Besides, assessment principles were the aspect that needed highest improvement. With a similar aim, Vogt and Tzagari (2014) identified the needs of EFL teachers in language testing and assessment based on how they perceive their LAL and pre- and in-service trainings. In this mixed-methods research, a questionnaire was conducted to 853 participants from seven different countries across Europe including Turkey, and interviews were conducted with 63 participants from three selected countries of Greece, Germany, and Cyprus. The teachers' LAL level was not satisfactory since no training or limited training was received in this field. Besides, teachers needed trainings mostly for the assessment purposes such as grading, placement of the students, and awarding certificates and did not feel competent to review the quality of the assessment tools according to the reliability and validity principles of assessment along with employing portfolio assessment, self- and peer-assessment in their practices. Besides, assessment practices can be influenced by former experiences of the teachers in that teachers were familiar with assessing language

skills as a result of their daily responsibilities proposed by the educational policies, which is believed to contribute to their LAL. However, they still expressed a need for further training. Based on the findings of their previous study, Tzagari and Vogt (2017) further investigated EFL teachers' utilization of assessment in the context of primary and secondary schools in Germany, Cyprus, and Greece. The data, obtained from the 63 EFL teachers, revealed that teachers were more inclined to use traditional methods of assessment as in pen-and-paper tests for receptive skills and writing as a result of the assessment culture of their context whereas speaking was ignored in the testing procedures. For feedback, teachers focused on learners' mistakes found in their test results and provided feedback mostly individually as well as utilizing feedback to adjust and revise the course content. Finally, portfolios, peer- and self-assessment methods as well as rubrics were not mainly implemented in EFL teachers' assessment practices even though they indicated a basic awareness. In conclusion, pre-service education was not sufficient for developing EFL teachers' understanding and utilization of language testing and assessment procedures; thus, more in-service trainings were required as part of professional development. The contextual and experiential aspects that contribute to the language assessment literacy have been emphasized by Yan, Zhang, and Fan (2018). Their analysis, based on the data obtained from EFL teachers in a middle school in China regarding their needs and practices, showed that the context of assessment, teachers' training experiences, assessment practices, and needs in knowledge and training contributed to EFL teachers' LAL development. Contextual factors had a role in teachers' assessment practices and training needs as a consequence of assessment policy and culture, expectations of the institutions, students, and parents, which resulted in a testing-oriented approach to assessment practices and needs. As a result of the mediating impact of contextual factors on assessment practices, EFL teachers developed their LAL in the technical aspects of assessment such as item-writing, item discrimination and analysis, reliability, and communicating the results with the stakeholders, which indicates that experiential factors contribute to the construction of assessment knowledge based on their experiences. In conclusion, type of the school, educational level of the school, and the location of the school can possibly form EFL teachers' LAL in their knowledge, practice, and training needs. In another study conducted in China, Xu and Brown (2017) investigated Chinese EFL teachers' assessment literacy and if their experience, trainings, educational background, and gender had an impact on their assessment literacy. It is seen that Chinese EFL teachers' assessment literacy was poor as a result of inadequate trainings in pre- and in-service education, not

bearing a grounded criterion for assessment literacy, and lack of quality standards for assessment practices. Besides, no impact was found on teachers' assessment literacy considering their age, gender, educational background, experience.

Research into LAL have also been shaped by the investigation on EFL teachers' perceptions/beliefs/attitudes practices. To begin with, Chan (2008) dealt with the beliefs and practices of EFL teachers regarding multiple assessment and the difficulties contributing to their classroom practices. The data were collected through a questionnaire containing the questions in the form of Likert-type, multiple choice, and open-ended from 520 EFL teachers working in the elementary schools in Northern Taiwan. In addition to these, the demographic factors such as age, field of graduation, educational background, age and experience, and training were also scrutinized. The results showed that EFL teachers had a great deal of knowledge regarding multiple assessments and utilized them in their practices to a great degree. They mostly preferred a combination of multiple assessment and traditional assessment methods with an emphasis on the former as the main assessment method to be used. However, it was also observed that pen-and-paper tests were still the most frequent method that was used followed by classroom observation. Due to time constraints, the use of portfolio was not used as frequent as the other ones even though EFL teachers believed that it could improve learners' self-assessment abilities. In addition to time constraints, heavy workload, large class sizes, and conception of subjectivity regarding multiple assessment were the main reasons for not adopting multiple assessment. Finally, teachers' years of experience had a significant impact on their beliefs and practices. Concordantly, how an EFL teacher in China grounded her assessment practices in terms of what, how, and why she assessed her students have been scrutinized by Gu (2014). In this single case study, the data were obtained through classroom observations and interviews with the teacher. The analysis of the data indicated that course book used in the class and high-stakes exams that students had to take were the preliminary sources for the content of assessment while not taking curriculum standards into consideration, whereas the teacher's teaching experiences and inferences based on their observations in the class as well as the high-stakes exam shaped how assessment was implemented. Therefore, assessment was not a component that was embedded in the teaching and learning process, but rather used as a form of evaluation. Besides, the teacher was not able to provide a rationale for her assessment practices, which were mostly summative in addition to employing a limited range of assessment and feedback techniques.

In a comprehensive study, Jannati (2015) aimed for revealing the perceptions and practices of EFL teachers about their use of language assessment in addition to the experience factor. In this qualitative study, the data, which are obtained from interviews with 18 EFL teachers in the higher education context in Iran, demonstrated that EFL teachers had an adequate level of knowledge on the fundamental concepts of language assessment. Besides, EFL teachers viewed informing students about the assessment process and involving them in the process, placing importance on assessment principles of validity and fairness though not adopting them in their practices, and considering washback were the significant aspects that should be taken into consideration while assessing learners. For their practices, EFL teachers asserted that both summative and formative assessment should be implemented with a focus on the latter one. Besides, writing, speaking, reading, and listening were the skills that they mostly assessed as well as grammar, vocabulary, and pronunciation. Teachers utilized assessment for diagnosing students' weaknesses and strengths for giving feedback, communicating the results with their students as well as for evaluating their own instruction and educational program so as to make appropriate adjustments in their assessment methods. Finally, experience did not have a significant influence on teachers' perception. Overall, this study signified that even though EFL teachers were assessment literate to a certain degree, the reflection of their literacy was not observed in their assessment practices. Contradicting with Jannati's (2015) research, Hakim (2015), aiming to reveal EFL teachers' ideologies of language assessment and the extent of their awareness of language assessment, expressed that more experienced teachers, integrated assessment, principles of assessment, peer- and self-assessment, and planning assessment in accord with the course objectives with a list of test specifications were more appreciated even though the results indicated that teachers demonstrated a satisfactory level of understanding regardless of their years of experience. As for the EFL teachers' ideologies of language assessment, Hudaya (2017) was interested in assessment principles and finding out if EFL teachers were prepared to assess their students and how they applied assessment principles. A questionnaire was administered to 43 EFL teachers in Indonesia in the context of primary, secondary, and higher education. It was concluded that many EFL teachers were assessment literate and felt prepared for assessing their students. Besides, even though some of the teachers' form of feedback was limited to grades or numbers, reliability, validity, practicality, authenticity, and washback were the assessment principles that EFL teachers were considerate of in their assessment practices.

Additionally, Berry, Sheehan, and Munro (2019) examined EFL teachers' attitudes and practices of language assessment. The data were collected by interviews, classroom observations, follow-up interviews, and group discussions from 54 EFL teachers, mostly had a certificate to teach EFL, in France, Spain, and the United Kingdom with varying years and opportunities of experience. The interviews and observations revealed that assessment trainings of the EFL teachers were limited or none; thus, they developed their conceptions of assessment based on their past personal experiences, in other words, they assessed their students as they were assessed when they were a student. Besides, they had negative attitudes towards assessment due to the fact that they considered assessment as testing even though self- and peer-assessment methods in addition to a variety of assessment methods and techniques were integrated in their classroom practices as a part of teaching/learning process. Contrastingly, the participants in Giraldo's (2019) study EFL teachers did not rely only on language tests but adopted both traditional and alternative assessment in their practices for gathering evidence on learners' reading, speaking, listening, and writing skills as well as grammar and vocabulary in addition to other skills that are technical or performance related during students' oral performances such as designing presentations and presentation skills. Feedback was provided as a means to improve learning and teaching process as well as to evaluate assessment practices. Additionally, EFL teachers believed that assessment and learning had an integral relation and validity, reliability, washback, and students' opinions should be taken into consideration. Parallel to Berry, Sheehan, and Munro's (2019) findings, it is indicated that EFL teachers' previous experiences in language assessment were influential on how they shape their knowledge rather than theoretical information.

2.9. Studies on Language Assessment Literacy/Knowledge and Practices in Turkey

It is seen that research on LAL in Turkey have lately become a subject of interest since most of the studies have been conducted recently. In the studies, knowledge (Genç, Çalışkan & Yüksel, 2020; Ölmezer-Öztürk & Aydın, 2019a), perceptions and practices (Büyükkarcı, 2014; Işık, 2021; Kırkgöz, Babanoğlu & Ağçam, 2017; Önalın & Gürsoy, 2020; Öz & Atay, 2017; Tanyer & Susoy, 2018; Üçok-Atasoy, 2020) , training needs (Ballıdağ & İnan-Karagül, 2021; Mede & Atay, 2017), and problems (Büyükkarcı, 2014; Tuzcu-Eken, 2016) as well as the impact of demographic and contextual factors (Büyükkarcı, 2016; İnan-Karagül, Yüksel & Altay, 2017) on language assessment have been scrutinized. Higher

educational context was mainly preferred by the researchers in Turkey whereas the research regarding K-12 context is limited. Besides, the participants' LAL regarding knowledge base was scarce (Genç, Çalışkan & Yüksel, 2020; Ölmezer-Öztürk & Aydın, 2019a) since in most studies EFL teachers' LAL was based on their perceptions.

Ölmezer-Öztürk and Aydın (2019a) aimed to measure EFL instructors' general and skill-based (reading, writing, listening, and speaking) language assessment knowledge level (LAK), which is also the source of inspiration for this current study. The participants of the study consist of EFL teachers who work at the preparatory programs at the universities in Turkey. In order to investigate language assessment knowledge base of EFL teachers, they have used the Language Assessment Knowledge Scale (LAKS) that they have developed in another study (Ölmezer-Öztürk & Aydın, 2018), which is used as a quantitative data collection tool in this very study. After administering the scale to the 542 EFL teachers from both state and private universities in Turkey, they have also investigated the impact of the demographic factors such as experience, educational background, and being a testing office member on their LAK level. In conclusion, EFL teachers' LAK in general was found to be low given that they had answered 25 questions correctly over 60 questions. As for the skill-based LAK of the participants, the study showed that the participants were found knowledgeable in assessing reading while in the assessment of speaking, writing, and listening, they indicated lower levels of competence, listening being the lowest. It is also seen in the study that participants that are a member of a testing office at the universities showed significantly higher scores than the ones who were not. Therefore, the only demographic factor that influenced LAK level of EFL teachers were being a testing office member. Similarly, in another study that also utilized the Language Assessment Knowledge Scale by Ölmezer-Öztürk and Aydın (2018), Genç, Çalışkan and Yüksel (2020) found similar results. Aiming to analyze Language Assessment Knowledge (LAK) level of assessing speaking and writing of EFL teachers who work at the high schools in Turkey as well as the influence of demographic characteristics of the participants on their LAK level. 58 English language teachers from both state and private institutions participated in this quantitative study. The results demonstrated that the EFL teachers had significantly lower scores in assessment of writing skills whereas they had higher scores in assessment of speaking skills. Additionally, participants' experience level, type of their undergraduate degree, and taking trainings for testing and assessment was proven to have no impact on their LAK level.

In accordance with the results of the abovementioned studies, Büyükkarcı (2016) addressed the assessment literacy of EFL teachers at both primary, secondary, and higher education context. The data for the study were collected with an inventory consisting of several assessment scenarios from 32 EFL teachers in Turkey. The overall assessment literacy of the EFL teachers was found to be rather low. The researcher also inspected the impact of educational background and experience on the assessment literacy of EFL teachers. the effect of experience as well as educational background whether having a BA, MA, or a PhD degree was not observed reflecting the similar findings as Ölmezer-Öztürk and Aydın (2019a) and Genç, Çalışkan, and Yüksel (2020). It is concluded that both novice and experienced teachers need more trainings as a part of their professional development. Concordantly, Mede and Atay (2017) focused more on the EFL instructors' needs of further training concerning language assessment and conducted a study to demonstrate Turkish EFL instructors' language assessment literacy by analyzing their pre- and in-service trainings and their attitudes. Accordingly, they administered a questionnaire to 350 EFL instructors followed by a focus-group interview process with 34 participants. The results indicated low levels of LAL on the part of Turkish EFL instructors who were working at the preparatory programs of the state and foundation universities. That is to say, as a consequence of insufficient training particularly in assessing language skills except for vocabulary and grammar, EFL instructors did not feel competent enough to design a test and were in need of further training mostly in classroom-based assessment methods such as self- and peer-assessment and other informal assessment methods. Besides, EFL instructors felt relatively more knowledgeable in assessment purposes while a comprehensive training was required for assessment principles of validity and reliability regardless of the trainings received on this concept.

In a recent study by Ballıdağ and İnan-Karagül (2021), the results were found to be corresponding to Mede and Atay's (2017) study. Concerned with the language assessment literacy of the EFL teachers in the K-12 context of Turkey, Ballıdağ and İnan-Karagül (2021) conducted a study to reveal EFL teachers' perceptions of their trainings and their perceived training needs. In this mixed-methods research, the researchers collected the data with a questionnaire and interviews from 101 EFL teachers working at the primary, secondary, and high schools. According to the results, the participants stated that they received sufficient trainings in providing feedback and self- and peer-assessment, defining the language constructs to be involved in assessment, and assessing grammar and vocabulary whereas

EFL teachers expressed a need for further training for portfolio assessment and assessment of cultural aspects even though portfolio assessment was employed by the teachers in their classroom practices, which leads to questioning the quality of their assessment practices. Besides, teachers did not adopt alternative assessment in their practices as a result of insufficient time and lack of trainings. Overall, EFL teachers did not find the trainings they reviewed during pre-service education sufficient for language assessment and needed more trainings in all aspects of language assessment particularly in portfolio assessment, self- and peer-assessment, placement purposes, assessing cultural aspects, and assessment principles. In parallel, Öz (2014) revealed that summative assessment methods were more dominant over formative ones in EFL teachers' classroom practices. That is, the assessment methods that are commonly preferred were gap-filling items, multiple choice questions, matching, true/false, and short-answer questions followed by oral exams, group work activities, projects and portfolios, assessment of performance tasks, essay-type exams, and oral presentations whereas drama, self- and peer-assessment, using rubric, etc. were favored the least by EFL teachers.

LAL demands a possession of certain components of knowledge, skills, and the ability to mirror assessment into classroom practice (Inbar-Lourie, 2013a). Even though pre-service education provides a fundamental knowledge on language assessment, it still needs to involve more practical concerns for higher LAL (Sevimel-Şahin & Subaşı, 2021). Consequently, many studies have investigated EFL teachers' perceptions and practices in various educational contexts. To begin with, Öz and Atay (2017) carried out phenomenological research with the purpose of discovering EFL instructors' perceptions on in-class language assessment and its reflection on their classroom practices. The researchers have also inquired the influence of experience on their perceptions and practices. 12 participants with various educational background and experience level participated in the semi-structured interviews of which the questions were developed by (Jannati, 2015). The results revealed that there is a discrepancy between teachers' perceptions and practices regarding in-classroom assessment regardless of the experience factor. Therefore, their assessment knowledge was not reflected in their classroom practices although they had indicated assessment literacy to a certain degree. Even though all teachers were aware of some fundamental elements of in-class assessment, novice teachers were more capable of using the related terminology. However, this is not related to experience by the researchers, but to being recently graduated and obtaining a theoretical knowledge that is recent. Besides,

it is demonstrated that teachers adopted formative assessment methods more and emphasized assessment of speaking skills in their practice despite using a limited variety of assessment instruments. In another study by Kırkgöz, Babanoğlu and Ağçam (2017), similar results were observed though different educational context was investigated. The researchers aimed for exploring the perceptions and practices of Turkish EFL teachers in the primary education context. In order to investigate the types of assessment and questions involved in the assessment by EFL teachers, and how they perceive the assessment of language skills in the primary schools, the researchers gathered the data via a questionnaire from 56 participants who were working as an EFL teacher in Turkish state primary schools. Based on the findings, the results show that Turkish EFL teachers believed that assessment was an essential part of foreign language education in terms of monitoring students' progress, meeting the curriculum objectives and adjusting it accordingly, making interpretations about learning and teaching. In addition, EFL teachers considered speaking, reading, and grammar as a significant skill to be assessed to a certain extent whereas assessing vocabulary received the highest emphasis. However, listening was seen as a skill that can be ignored in the assessment process while writing is considered somewhat important. Interestingly, grammar received less attention when compared to speaking and vocabulary skills. As for the types of assessment, EFL teachers utilized pen-and-paper exams, classroom performance of the students, projects and portfolios, oral exams, and presentations. In a similar vein, Arslan and Üçok-Atasoy (2020) observed that EFL teachers who work at the state middle schools in Turkey generally preferred conventional methods over alternative methods as well as ignoring listening and speaking skills completely while writing and reading skills were involved to a certain degree. Investigating the accordance between language assessment practices and policy, the researchers discovered that matching, true-false items, multiple-choice questions, and gap-filling activities were the mostly preferred techniques among traditional methods whereas role-plays and project assessment among alternative assessment techniques were employed. As for the assessment of language skills, the researchers' analyses of the exam papers and assessment documents that EFL teachers have used demonstrated that grammar and vocabulary were excessively prioritized in the assessment of young learners along with reading skills. Consequently, a discrepancy between practice and suggested assessment procedures by the policy makers is observed.

In another study, Işık (2021), adopting a mixed-methods approach, investigated ELT teachers' perceptions on trainings regarding language assessment and their perceptions and

practices of assessment. With these objectives in mind, he collected the data with a questionnaire and follow-up interviews developed by Vogt & Tsagari (2014) as well as classroom observations from 198 ELT teachers who work at primary, secondary, and higher educational levels of both state and private institutions. The results showed that even though ELT teachers' assessment literacy was low, they were confident with their current status. Additionally, pre- and in-service trainings they received were inadequate in providing the practical concerns of assessment, but rather focused on theoretical aspects. As a result, it is revealed that teachers needed further training since they perceived assessment as a formal procedure for grading purposes. Similarly, their perceptions and practices of assessment exhibit a conventional understanding and a limited variety of assessment methods that are generally inconsiderate of assessment principles, improving instruction, appropriacy of the assessment, and evaluation criteria across all educational levels. On the other hand, a dilemma between grading students based on their actual performances and meeting the parents' and school administrators' expectation of success was observed in ELT teachers' assessment practices working at private institutions. In conclusion, it is clear that contextual factors such as workplace have an impact on ELT teachers' assessment practices. Another study that scrutinized the perceptions and practices of EFL teachers of young learners is carried out by Tanyer and Susoy (2018). In line with Işık's (2021) findings, it is demonstrated that both pre- and in-service teachers' knowledge and skills in assessing the language skills of young learners were insufficient. On the contrary to Işık's (2021) study, the participants in this study considered it as a challenging procedure and did not feel competent to sustain as a result of lack of training dedicated to this particular learner group though they believed that assessment of young learners should be prioritized. Besides, in-service teachers' assessment practices are mostly reliant on non-systematic classroom observations based on their intuitions. In addition to employing assessment methods such as giving homework, competitions, giving awards, classroom participation, and student behaviour, in-service teachers believed that assessment should be encouraging and motivating considering the students' age. As a conclusion, even though in-service EFL teachers adopted mostly formative assessment methods for assessing young learners, their perceptions and practices were largely based on intuitive inferences rather than scientific methods and approaches or a sort of assessment criteria.

LAL incorporates many components such as language pedagogy, principles, and technical skills, yet how these components are practiced may be influenced by teachers' personal

beliefs and attitudes or local context (Taylor, 2013). Thus, considering the impact of demographic and contextual factors, Önalın and Gürsoy (2020) conducted a study to reveal EFL teachers' perceptions on assessment and its use in primary and secondary schools in Turkey. For this objective, the researchers questioned how teachers use assessment and assessment results, how they involve students in the assessment process, how they perceive their knowledge on assessing language skills, and which assessment methods they prefer. Moreover, they also aimed to identify whether the participants' gender, level of experience, and the type of school they work at made any difference in their perceptions. The scale was administered online to 348 English language teachers who were working at primary and secondary schools at both state and private institutions. The results demonstrated that EFL teachers used assessment and assessment data more for giving feedback to students, parents, and school administrators, discovering whether students have failed or succeeded as well as for utilizing it as a reflection of their own practices and designing syllabus. However, using assessment results for lesson planning and involving students in the assessment process except when used with the object of self-assessment were demonstrated at a moderate degree aside from the teachers with higher experience, who used assessment and its results more in line with assessment purposes and believed that students should be involved in assessment process more. Besides, the teachers believed that they had adequate knowledge on assessing language skills and adopted formative and summative assessment methods and utilized a rubric for assessing four skills and students' performances. However, for less experienced teachers and teachers who work at the state institutions, rubrics were not preferred as much. In another study, Yastıbaşı and Takkaç (2018) aimed to reveal EFL instructors' decision-making process while planning and designing their assessment. Based on the data obtained from think-aloud protocols, the results displayed that the instructors' decision-making process of designing an assessment tool were highly centered on the course books they adopt in their program. Due to using course books as a base for assessment preparation, validity of the assessment tool, which is a fundamental assessment principle, was improved. Another finding of the study is that the exams were in accord with classroom objectives and instructional activities in addition to being course book-centered, instructors also took their students' opinions under consideration, which lead to positive washback.

In addition, Büyükkarcı (2014) conducted to find out about EFL teachers' beliefs and practices and within the scope of formative assessment in Turkish primary schools. For this particular objective, he administered a questionnaire to 69 EFL teachers and carried out

semi-structured interviews with 10 EFL teachers. The data revealed that even though EFL teachers had a positive attitude towards formative assessment in that adopting self- and peer-assessment, involving learners in the process, and giving feedback were considered beneficial, their beliefs cannot be transferred into their actual classroom practices as a result of crowded class sizes reaching up to 60 students and their busy schedule of class hours, which was on average 25 lesson hours. Therefore, a mismatch appears between the assessment procedures suggested by the curriculum and actual classroom practices. Likewise, İnan-Karagül, Yüksel and Altay (2017) demonstrated that lack of course hours for English and overcrowded classrooms is also reason for EFL teachers in the context of primary, secondary, and higher educational context to prefer mostly traditional pen-and-paper tests over alternative assessment techniques. 25 teachers from state schools were interviewed for their assessment practices and grading procedures. Even though they had a positive attitude towards the benefits of alternative assessment, their beliefs did not reflect on their classroom practices. In the same vein, Tuzcu-Eken (2016) aimed for addressing the occurring issues regarding testing methods EFL teachers use in the vocational schools at tertiary level education. Semi-structured interviews with five instructors and focus-group interviews with student groups consisting of 15 participants were conducted. The study revealed Turkish EFL instructors' and students' views on testing and instructors' preferred testing techniques in vocational schools in Turkish higher education context. In conclusion, EFL teachers believed that testing is a necessary procedure that informs them about the evaluation of both students and their instruction. The instructors also stated that they tried to involve as many testing techniques as they can whether benefitting from ready-made tests or adapting them, such as Q/As, gap-filling, translation, reordering, and multiple-choice questions despite the barriers of large class sizes, inadequacy of the technical equipment, and students' lack of motivation towards learning English.



CHAPTER III

METHODOLOGY

3.1. Research Design

In this study, mixed-methods approach is employed with the aim of engaging both quantitative and qualitative research methods. Creswell (2012) informs that mixed-methods research is used when “both quantitative and qualitative methods, in combination, provide a better understanding of the research problem and question than either method by itself” (p. 535). The reason behind adopting mixed-methods research is to achieve a more comprehensive understanding of a complex phenomenon by involving different perspectives or different sources of data (Fraenkel, Wallen, Hyun, 2011). Therefore, using quantitative methods for collecting the data on language assessment knowledge and qualitative methods for collecting the data on language assessment practices would provide deeper insights about EFL teachers’ Language Assessment Literacy (LAL).

Dörnyei (2007) delivers nine types of mixed-methods research designs that are based on the two principles of sequence or dominance of the methods being used. Among these research designs, this study adopts the QUAN → qual combination, in which the upper- and lower-case letters indicate the dominance while the arrow indicates the sequence of the method. In other words, a quantitative phase bearing the higher dominance in the study is followed by a qualitative phase. This research design is also referred to as explanatory sequential design analysis, in which quantitative data is collected first and qualitative data is collected later in the process by selecting the participants, such as extreme cases, based on the results obtained from the quantitative data (Creswell, 2012). Thus, in this current study, first the quantitative data on EFL teachers’ language assessment knowledge is gathered with a scale that measures the participants’ language assessment knowledge of reading, writing, listening, and speaking skills, and then, the qualitative data on EFL teachers’ language assessment practices is

collected through semi-structured interviews after the analysis of the quantitative data. In sequential research designs, the second form of data is collected after the first form of data is obtained and analyzed (Ary, et al., 2010). The rationale behind this sequence is that the results of the first stage may inform the researcher about the operations to be conducted in the following stage. That is, the participants for the semi-structured interviews that constitute the second phase of the study, is determined according to the results and analyses obtained from the scale that constitutes the first phase of the study. As for the final stage of the study, after the separate discussions on the quantitative and qualitative data analyses are presented, interpretations based on quantitative and qualitative data combined are drawn upon so as to provide more profound understanding of the complex phenomenon of LAL.

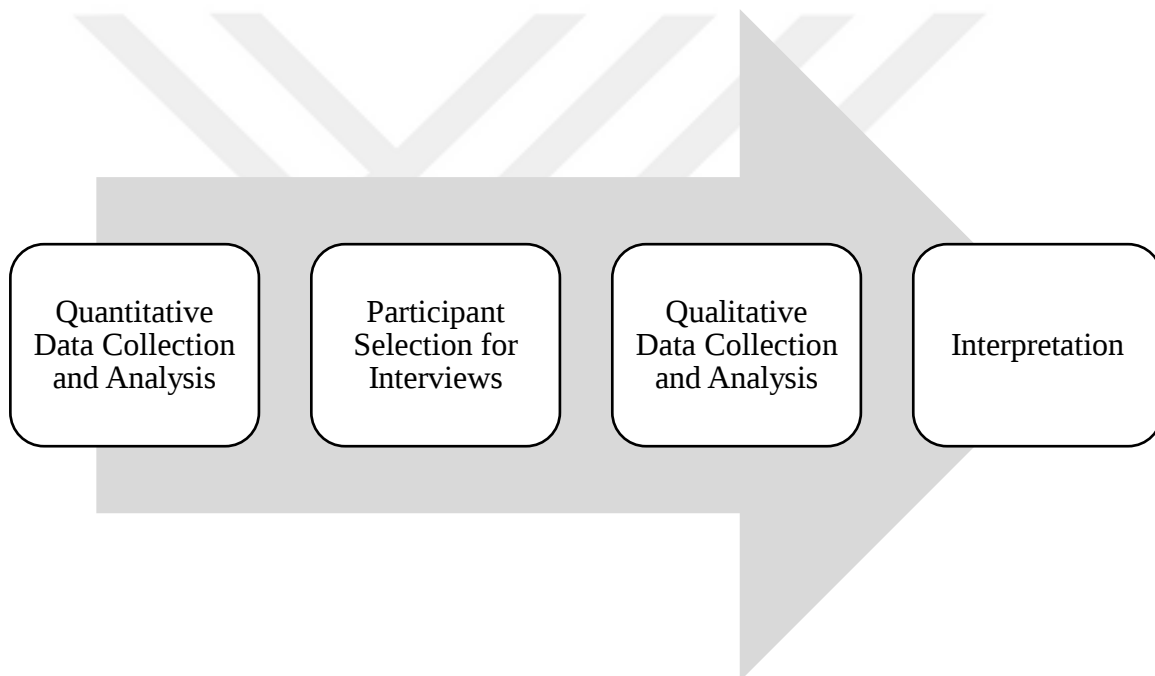


Figure 1. Research design of the study (Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.). Boston, MA: Pearson, p. 541)

3.2. Research Context

English is taught as a foreign language in Turkey, where foreign language education is involved in all the educational levels of primary, secondary, and tertiary level. Primary (1st, 2nd, 3rd, 4th grades), lower-secondary (5th, 6th, 7th, 8th grades), and upper-secondary (9th, 10th, 11th, 12th, grades) educational levels, namely the K-12 context, constitutes the context for this current study including both state and private schools. In the K-12 context of Turkey, foreign language education in state schools starts in the 2nd grade and continues until the end

of the secondary education. In primary schools, 2nd, 3rd, and 4th grades take 2 hours of English per week while in lower-secondary level consisting of middle schools, 5th and 6th grades take three hours and 7th and 8th grades take four hours of English. In addition to this, students can additionally take two hours in elective courses in middle schools. In upper-secondary level, consisting of high schools, students take approximately four hours of English although this varies depending on the type of the high school. However, in some high schools (Fine Arts High Schools, Sports High Schools), students take two hours of English whereas in high schools that involves an additional year of compulsory preparatory year at the very beginning of their education, students can take up to 20 hours of English. In private schools the amount of the lesson hours differs according to each school based on their educational program. As for the testing and assessment procedures in the K-12 context, which is the main concern of this study, they are under the responsibility of English teachers who work at these institutions even though students at lower- and upper-secondary levels must take a standardized test administered by the Ministry of National Education at the end of each educational level.

In order to practice as an English language teacher in Turkey, a four-year Bachelor of Arts (BA) degree in the department of English language teaching is required. Additionally, undergraduates of English language and literature, American culture and literature, translation and interpreting, and English linguistics can also work as an English language teacher if they have a certificate of pedagogical formation. After a four-year university education, based on their test and interview results that MoNE administers, they can practice as an English language teacher in the state schools in Turkey.

3.3. Participants

The aim of this study is to investigate EFL teachers' language assessment knowledge and their classroom practices in the K-12 schools in Turkey. Thus, the teachers who work at the primary, middle, and high schools in Turkey, whether private or state, in the 2021-2022 educational year constitute the universe of this research.

For the quantitative part of the study, after obtaining the necessary permissions from the MoNE (Appendix B) to implement the LAKS to the teachers, the scale is sent to EFL teachers all over the country via MoNE's electronic document management system. Among the whole population of EFL teachers, 202 participants took part in the scale on a voluntary basis. Demographic information of the participants is presented in the table below.

Table 1

Demographic Features of the Participants in the LAKS

Demographic Features		Number	Percentage
Gender	Male	48	23.8
	Female	154	76.2
Years of experience	1-5 years	91	45.0
	6-10 years	34	16.8
	11-15 years	23	11.4
	16-20 years	27	13.4
	More than 20 years	27	13.4
BA program	ELT	181	89.6
	Non-ELT	21	10.4
Educational background	BA degree	161	79.7
	MA degree	37	18.3
	PhD Degree	4	20
Educational level of the institution	Primary school	35	17.3
	Secondary school	103	51.0
	High school	64	31.7
Type of the institution	State/Public	187	92.6
	Private	15	7.4
Geographical region	Aegean Region	16	7.9
	Black Sea Region	29	14.4
	Central Anatolia Region	53	26.2
	Eastern Anatolia Region	22	10.9
	Marmara Region	24	11.9
	Mediterranean Region	11	5.4
	Southeastern Anatolia Region	47	23.3
Having testing/assessment course in pre-service	Yes	113	55.9
	No	89	44.1
Attending professional development programs/courses on testing/assessment	Yes	105	52.0
	No	97	48.0

As it is seen from the table, 75% of the participants consist of female teachers. When their experience level is considered, teachers with 1-5 years of experience constitute the half of the participants while the others have an approximately equal distribution. 90% percent of the participants have graduated from the department of English Language Teaching. Among 202 participants, only 4 of them had a PhD degree, while the number of the participants with MA degree is 37, and with BA degree is 161, which is 80% of the whole participants in the study. Another variable on the scale was about the participants' workplace. It is observed that while almost all the teachers work at a state school, nearly half of the teachers work at the middle schools followed by high school with 30% and primary schools with 17%. In this

study, the geographical region is also aimed to be analyzed and most of the participants took part in the Central Anatolia and Southeastern Anatolia Region whereas the least participation is from Mediterranean Region. Finally, teachers' pre- and in-service trainings regarding language assessment is analyzed. It is seen that more than half of the teachers have received a sort of an assessment course in pre-service education. However, it can be deducted from the table that half of the participants did not attend to any in-service training for their professional development.

For the qualitative part of the study, purposive sampling is adopted, or as Creswell (2012) puts forth, purposeful sampling, which is participant selection is intentional so as to reach an informed data about the phenomenon. Therefore, the participants who took part in the scale were contacted via e-mail based on their demographic features, such as gender, workplace, experience, etc., with the aim of forming a representative data of the population. After sampling procedure, participants who volunteered are interviewed. In total, 17 EFL teachers took part in the semi-structured interviews. The demographic information about the participants in the interview are presented in the table below.

Table 2

Demographic Features of the Participants in the Interview

Participants	Gender	Experience	Type/Educational level of the institution	Average size/Total number of the students	Teaching hours	Educational background/BA program	Pre-service/In-service training	Geographical region
Ahmet	M	11	State/Secondary	33/330	40	BA/ELT	Yes/Yes	Central Anatolia
Burcu	F	3	State/Secondary	32/200	30	BA/ELT	Yes/No	Eastern Anatolia
Canan	F	6	State/Primary	15/100	18	MA (non-ELT) /ELT	Yes/No	Black Sea
Damla	F	5	State/Primary	27/260	18	BA/ELT	Yes/No	Marmara
Elif	F	6	State/Secondary	10/70	20	BA/ELT	Yes/No	Southeastern Anatolia
Ferhat	M	4	State/Secondary	25/150	20	BA/ELT	Yes/No	Central Anatolia
Giray	M	22	State/High	30/120	16	BA/ELT	Yes/Yes	Black Sea
Hale	F	14	State/Primary	20/180	20	BA/ELT	Yes/Yes	Aegean
Işın	F	5	Private/Secondary	16/110	30	BA/non-ELT	No/No	Mediterranean
Kaan	M	16	State/Secondary	25/120	29	BA/ELT	Yes/Yes	Marmara
Leyla	F	13	State/Primary	30/390	26	BA/ELT	Yes/No	Mediterranean
Merve	F	10	State/High	20/45	23	BA/non-ELT	Yes/No	Aegean
Nazım	M	9	State/High	18/170	25	BA/ELT	Yes/Yes	Aegean
Oya	F	16	State/Secondary	17/120	23	BA/non-ELT	Yes/No	Central Anatolia
Pırlı	F	5	State/Secondary	25/250	40	MA (ELT) /ELT	Yes/Yes	Central Anatolia
Rana	F	21	State/Primary	38/300	18	BA/ELT	Yes/No	Black Sea
Sevgi	F	14	State/High	35/170	21	BA/ELT	Yes/No	Aegean

As the table presents, there are 12 (70%) female, 5 (30%) male participants. As for the teachers' years of experience, 5 (30%) teachers have 1-5 years of experience, 4 (20%) teachers with 6-10 years, 4 (20%) teachers with 11-15 years, 2 (10%) teachers with 16-20 years, and 2 (10%) teachers have more than 20 years of experience. All the participants except for one are working at a state school. The number of the teachers who are working at primary schools is 5 (29%), the number of the teachers who are working at a middle school is 8 (47%), and the number of the teachers who are working at a high school is 4 (24%). In the semi-structured interviews, teachers are also asked about their teaching hours and

average class sizes at their current workplace. The number of the classes with up to 20 students is 5 (30%), the number of the classes with students between 20-30 is 6 (35%), and the number of the classes with students of 30 and above is 6 (35%). When it comes to the total number of the students that the teachers have, there are 2 (12%) teachers who have below 100 students, 9 (52%) teachers have between 100-200 students, 3 (18%) teachers have between 200-300, and 3 (18%) teachers have over 300 students. Another factor is teaching hours, among the participants, 4 (23%) teachers have less than 20 teaching hours, 9 (53%) teachers have teaching hours between 20-30, and 4 (23%) teachers have 30 and above teaching hours. It is clear that EFL teachers in the K-12 context in Turkey have a high workload in regard to the class sizes, the number of the students they are responsible for, and teaching hours. For the participants' educational background and professional development, it is seen that 15 (88%) teachers have a BA degree while 2 (12%) teachers, one of which is a non-ELT department, have an MA degree. 14 (82%) teachers have graduated from an ELT department whereas 3 (18%) teachers have graduated from non-ELT departments. All the participants except for one stated that they have received a training in pre-service education about testing and assessment while 6 (35%) of the participants have participated in an in-service training program.

3.4. Data Collection Tools

The data collection tools in this study consist of Language Assessment Knowledge Scale and semi-structured interviews.

3.4.1. Language Assessment Knowledge Scale (LAKS)

For the first phase, language assessment knowledge level of EFL teachers in the K-12 context in Turkey is aimed to be investigated. In order to collect the quantitative data of the study, Language Assessment Knowledge Scale (LAKS) (Appendix D), developed by Ölmezer-Öztürk and Aydın (2018), is used. The scale consists of two parts, and the first part involves the questions about the demographic features of the participants such as their gender, years of experience, BA program being graduated, educational background, type of the institution they work at, being a testing office member, and attendance to additional training in testing and assessment. However, since this study is concerned with the K-12 context, being a testing office member is omitted since it was not related to this particular

context. Instead, educational level of the workplace, geographical region they work in questions were added. In the second part, there are four factors that aim to measure EFL teachers' language assessment knowledge level in assessing reading, listening, writing, and speaking. In the scale, consisting of 15 items in each factor and 60 items in total, the participants are asked to read the statements and select the most appropriate option for each statement among True/False/Don't Know options. The scale has been administered online. Before the administration, the participants are informed about the aims and the scope of the study, that the participation is on a voluntary basis and they can stop participating any time they want, that their personal information and answers are not to be shared with third parties and is only to be used for scientific purposes as well as the information of the researcher that they can reach for further questions.

For the validity of the scale, Ölmezer-Öztürk and Aydın, (2018) sought for expert opinion. That is, the scale has been scrutinized by 11 experts in this field and 18 practitioners, six of which have MA or PhD degree in English language assessment, and took its final form, which is developed in the form of "true", "false", and "don't know". The participants' answer to each statement is scored "1" point if their answer is correct, and "0" point if their answer is wrong or "don't know" option is selected. Accordingly, the highest score that can be obtained from each factor of the scale is "15", in total "60" points, and the lowest score is "0". The scale has been developed after a thorough literature review process and consists of elements required for a fundamental knowledge on language assessment (Ölmezer-Öztürk and Aydın, 2018). The reliability analysis of the scale has been completed by the researchers who developed the scale, and the reliability of the scale has been measured as .91 for Cronbach Alpha coefficient.

3.4.2. Semi-Structured Interviews

Interview is a communication tool in our social context as well as being a research tool to obtain information about the subject matter (Dörnyei, 2007). In Nunan's (1992, p.231) depiction, it is "the elicitation of data by one person from another through person-to-person encounters". Among the many types of interviews, semi-structure interviews are utilized for the second phase of the study, in order to obtain more profound understandings about EFL teachers' language assessment practices in the K-12 context in Turkey. In semi-structured interviews, the researcher has a set of pre-prepared questions, which is combined with follow-up question to provide elaboration from the participants on the aspects of their

response (Griffiee, 2012). The semi-structured interview used in the study consists of two parts (Appendix F). The first part contains the questions about the participants' demographic information such as gender, age, years of experience, educational background, educational level of the schools they work in, professional development, and so forth. The second part contains 16 questions, 2 of which is specific to teachers of young learners, related to EFL teachers' language assessment practices, such as the purposes of assessment, language skills they assess, assessment methods they adopt, planning and decision-making process, student involvement in the assessment, feedback etc. The data obtained from the semi-structure interviews constitute the qualitative data for the study. With the aim of collecting more profound and in-depth data from the participants about their classroom practices, semi-structured interviews are preferred since they provide the interviewer with the freedom and flexibility to probe for clarification and elaboration (Mackey & Gass, 2005). The questions for the interview are developed after a thorough process of literature review and involves the fundamental concepts concerning language testing and assessment. In order to prevent any sort of loss in meaning and to make sure that participants can express themselves clearly, the interview questions are developed in Turkish language. After the initial development of the semi-structured interview questions, they are submitted to expert opinion so as to ensure the construct and content validity of the questions. That is, the interview questions are scrutinized by two professors, who are experts in the language assessment field, and a research assistant who has a PhD degree in this field. Necessary changes and adaptation are made in the questions regarding its content and form based on the expert opinion. Thus, it is considered as a valid instrument for collecting data on language assessment practices of EFL teachers in the K-12 context in Turkey.

3.5. Data Collection Process

For the data collection, firstly, ethics committee approval is received (Appendix A). Since the participants for the study are the EFL teachers that work at the schools held under MoNE, necessary permissions are taken from the organizations of MoNE for administering the scale and interviews.

For the quantitative data collection, LAKS is prepared in an online format through Google Forms. It is sent to every EFL teacher throughout the country via electronic document management system of MoNE. However, the number of the participants were not adequate to be used and analyzed. Therefore, schools from various regions and cities are reached to

remind the teachers to complete survey. Besides, it is also shared on online platforms where EFL teachers are the participants of professional development communities. At the end of the process, which took a year in total in the 2021-2022 academic year, 202 EFL teachers participated in the quantitative part of the study.

For the qualitative data collection, semi-structured interviews are designed and presented to expert opinion to ensure construct validity of the questions. After necessary adjustments are made, the interview questions took its final form to be used as a qualitative data collection instrument. The interviews are conducted in Turkish so that any possible loss in meaning due to language related issues can be eliminated and the participants can express themselves comfortably. The participants for the interviews are selected with purposive sampling techniques among the EFL teachers who have completed the LAKS in the first phase of the study. The EFL teachers are contacted via e-mail for their participation, and 17 EFL teachers accepted to be interviewed. The interviews are conducted online via Zoom application and recorded by voice recorders in an audio form to be transcribed for data analysis. After completing the interviews, the audio recordings are transcribed into written form to be coded later.

3.6. Data Analysis

The analysis process of quantitative data obtained with the LAKS and qualitative data with semi-structured interviews are explained below.

3.6.1. Quantitative Data Analysis

The quantitative data of the study is collected with LAKS to analyze EFL teachers' language assessment knowledge, the impact of their demographic features on their LAK, and if their difference between their perceived self-competency and actual LAK level. Therefore, in the quantitative data analysis descriptive and inferential statistics are adopted. SPSS program is utilized to perform the data analysis. Firstly, the normality test is carried out, and the results indicated that the data is normally distributed. Therefore, parametric tests are conducted. To estimate reliability of the quantitative data, Cronbach alpha coefficient is determined as 0.58. Reliability rate between 0.50 and 0.70 indicates moderate reliability (Hinton, McMurray, & Brownlow, 2014). The participants' LAK level in general and skill-based is analyzed with one-sample t-tests. The impact of the demographic characteristics of the participants on their

LAK level is analyzed with independent samples t-tests and one-way ANOVA and presented with inferential statistics. For the third research question of the study, which aims to find out whether there is a difference between the participants' perceived- self-competency in assessing language skills and their actual LAK level, one-way ANOVA analyses are employed.

3.6.2. Qualitative Data Analysis

In order to collect the qualitative data of the study, semi-structured interviews are conducted in Turkish language with 17 EFL teachers in the K-12 context in Turkey. The interview questions are asked to the participants and their responses are recorded in the audio format. For the qualitative data analysis, content analysis approach is utilized. Creswell (2012) puts forth the process of content analysis as after collecting the data, it is transcribed to be read by the researcher to obtain a general understanding of the data, which is followed by the process of coding the data into chunks or segments in order to determine the emerging themes. Similar process is followed by the researcher for the qualitative analysis of the data. Participants' responses are divided into chunks, and codes are assigned. Emerging categories and themes are presented and discussed in the light of the literature. The participants are also assigned with a pseudonym. In order to ensure the reliability of the qualitative data analysis, inter-rater reliability technique is employed. Therefore, 10% of the data is coded by another researcher and inter-rater reliability is estimated as 0.82 with Miles and Huberman's (1994) formula for inter-rater reliability, which is $\text{consensus}/(\text{consensus} + \text{disagreement})$.

The statistical analysis procedures according to the focus of the research questions and data collection instruments are presented in the table below.

Table 3

Statistical Analysis Procedures

Focus of the research questions	Data collection instrument	Statistical methods used
Language assessment knowledge level in general and skill-based	LAKS	Descriptive statistics (mean, Std. Deviation) T-test (one sample)
Effects of demographic features	LAKS	Inferential statistics T-test (independent samples) ANOVA (one-way)
Relationship between perceived self-competency and actual LAK level	LAKS	ANOVA (one-way)
Assessment practices	Semi-structured interviews	Content analysis (codes, categories, themes)
Assessment-related problems that EFL teachers encounter with	Semi-structured interviews	Content analysis (codes, categories, themes)

CHAPTER IV

FINDINGS AND DISCUSSION

In this chapter, EFL teachers' LAK level in general and skill-based, the impact of the demographic features and perceived self-competency on their LAK level, classroom practices, and the problems that occurs regarding language assessment in the K-12 context is presented and discussed in the light of the literature. First three research questions are based on the quantitative data obtained from the LAKS whereas the other research questions are based on the qualitative data obtained from the semi-structured interviews.

RQ1: What are the Language Assessment Knowledge (LAK) level of EFL teachers in the K-12 context?

RQ2: Does EFL teachers' level of LAK change according to:

- a. Gender,
- b. Years of experience,
- c. The BA program being graduated,
- d. Educational background,
- e. Type of the institution they work at,
- f. Educational level of the workplace,
- g. Geographical region they work in,
- h. Taking a separate testing and assessment course in BA,
- i. Attendance to additional training in testing and assessment?

RQ3: Does EFL teachers' perceived self-competency in assessing each language skill have an effect on their LAK level?

RQ4: What are the language assessment practices of EFL teachers in the K-12 context in Turkey?

RQ5: Is EFL teachers' LAK level reflected on their classroom practices?

RQ6: What are the problems that EFL teachers encounter with in the K-12 context in Turkey regarding language assessment?

4.1. EFL Teachers' General and Skill-based Language Assessment Knowledge Level

The first research question focuses on the EFL teachers' LAK level in general and skill-based in Turkish K-12 context. Therefore, the findings are presented and discussed based on the participants' mean scores in each skill and in general, the significance level of their scores, and the correlation between the participants' scores in assessing language skills. Finally, the items on the scale are analyzed individually according to the participants' correct and incorrect answers.

The analyses of the EFL teachers' answers on the scale regarding their general LAK show that they had a mean score of 32.099 over 60, indicating that their general LAK level is higher than the half of the total score. The highest score that can be obtained from the scale is "60" while the lowest is "0". Therefore, to decide whether the participants have low or high LAK level, the point of reference for the study is determined as the half of the total score, which is "30". To figure out whether the participants have significantly high level of LAK, one sample t-test is conducted. The table below present the findings of one sample t-test.

Table 4

One Sample T-test Results for General LAK Level

	N	$\bar{X}$	Mean diff.	df	t	p
General LAK	202	32.099	2.099	201	5.485	.000*

*p<.05; Test value=30

When the point of reference (30) is compared to the participants' mean scores, a difference of 2.099 is observed between the participants' mean scores and point of reference. The results indicate that the teachers' general LAK level is significantly high in the K-12 context.

For the purpose of analyzing EFL teachers' mean scores in each language skill, one sample t-test is applied. In the scale, the total score that can be obtained from each language skill is "15" while the lowest is "0". Thus, the test value is determined as the half of the total score, which "7.5". The participants' mean scores in each skill are $\bar{X}$ =9.5693 (reading), $\bar{X}$ =6.8317 (listening), $\bar{X}$ =7.0594 (writing), and $\bar{X}$ =8.6386 (speaking). It is observed that in assessing reading and speaking skills, the participants' mean scores are higher than the reference point (7.5) whereas in assessing listening and writing skills, their mean score seems to be lower than 7.5. To determine whether the difference between the participants' mean scores and test value (7.5) is significantly higher or lower, one sample t-test results are shown in the table below.

Table 5
One Sample T-test Results for Skill-based LAK Level

	N	$\bar{X}$	Mean diff.	df.	t	p
Reading	202	9.5693	2.0693	201	14.821	.000*
Listening	202	6.8317	-.6683	201	-4.444	.000*
Writing	202	7.0594	-.4405	201	-2.900	.004*
Speaking	202	8.6386	1.1386	201	7.154	.000*

*p<.05; Test value=7.5

The results reveal that the participants' skill-based LAK level in assessing reading, which is the language skill that the participants displayed the highest knowledge, was significantly higher than the test value (7.5). Thus, it is evident that the most knowledgeable skill that the participants have is assessing reading. On the contrary, the participants demonstrated the lowest knowledge level in assessing listening, and the mean difference between the test value (7.5) and the participants' mean scores ($\bar{X}$ =6.8317) is found to be statistically significant. As a result, it is concluded that the participants' LAK in assessing listening is insufficient. Another skill that has received a low mean score is assessing writing. Even though there was not a drastic difference between the participants' mean scores ($\bar{X}$ =7.0594) and the test value (7.5), the one sample t-test results indicate that the participants' knowledge level in assessing writing is significantly low. The final skill, which is assessing speaking, has received a high knowledge level. Based on the table above, it is shown that the participants' mean scores in assessing speaking ($\bar{X}$ =8.6386) is significantly high when compared to the test value (7.5). All in all, the results illustrate that while the participants' knowledge level in assessing

reading and speaking is adequate, their knowledge level in assessing writing and listening is insufficient.

When the correlation between the language skills is scrutinized, the results indicate a both based on the Pearson correlation analysis.

Table 6

The Correlation Among the Skill-based LAK

	LAK	Reading	Listening	Writing	Speaking
LAK	1	.594**	.673**	.712**	.568**
Reading		1	.246**	.241**	.089
Listening			1	.369**	.106
Writing				1	.197**
Speaking					1

** Correlation is significant at the 0.01 level (2-tailed); N=202

The findings revealed that EFL teachers have high LAK in general, in assessing reading, and speaking whereas low LAK is indicated for assessing writing and listening. To be able to make an in-depth discussion of these findings, the participants' correct and incorrect answers to each item on the scale is investigated and discussed. The table below presents the descriptive statistics of LAK level of EFL teachers in the K-12 context in Turkey.

Table 7

Descriptive Statistics of Language Assessment Knowledge Level of EFL Teachers in the K-12 Context

Items	N	True	False	Don't Know	$\bar{X}$	S
Assessing Reading						
1. Asking learners to summarize the reading text is a way of assessing their reading skills.	202	170*	17	15	,8416	,36604
2. When asking several questions about a reading text, all the questions are independent of each other.	202	59*	136	7	,2921	,45585
3. Cloze test is used for assessing the main idea of the text.	202	77	106*	19	,5248	,50063
4. In a reading exam, using a text learners have encountered before is not a problem.	202	78	96*	28	,4752	,50063
5. One reading text is enough to be included in a reading exam.	202	70	113*	19	,5594	,49769
6. The language of the questions is simpler than the text itself.	202	140*	43	19	,6931	,46237
7. Errors of spelling are penalized while scoring.	202	78	97*	27	,4802	,50085
8. Taking vocabulary difficulty into consideration is necessary in assessing reading skills.	202	160*	36	6	,7921	,40683

9. Including not stated/doesn't say along with true/false items has advantages over true/false items.	202	144*	25	33	,7129	,45355
10. The more items a reading text is followed, the more reliable it becomes.	202	140*	41	21	,6931	,46237
11. Using the same words in the correct option as in the text is not a problem.	202	61	123*	18	,6089	,48921
12. Simplification of reading texts is avoided.	202	60	111*	31	,5495	,49878
13. Reading texts in a reading exam include various genres (essay, article, etc.).	202	176*	15	11	,8713	,33571
14. In top-down approach, assessment is on overall comprehension of the reading text.	202	151*	21	30	,7475	,43551
15. Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.	202	147*	40	15	,7277	,44624
Reading-Total	202				9,5693	1,98443
Assessing Listening						
16. Using reading texts for listening purposes poses a problem.	202	74*	102	26	,3663	,48300
17. Including redundancy (e.g., what I mean to say is that) in a listening text poses a problem.	202	56	129*	17	,6386	,48160
18. Any type of listening text is used for note-taking.	202	98	76*	28	,3762	,48564
19. Spelling errors are ignored in scoring the dictation.	202	95*	96	11	,4703	,50036
20. Errors of grammar or spelling are penalized while scoring.	202	98	83*	21	,4109	,49322
21. A listening cloze test is a way of selective listening.	202	159*	20	23	,7871	,41035
22. Phonemic discrimination tasks (e.g., minimal pairs such as sheep-ship) are examples of integrative testing.	202	120	39*	43	,1931	,39569
23. Scoring in note-taking is straightforward.	202	107	41*	54	,2030	,40321
24. In discrete-point testing, comprehension is at the literal/of local level.	202	100*	26	76	,4950	,50122
25. Using dictation diagnostically in assessing listening skills does not pose a problem.	202	90*	78	34	,4455	,49826
26. Giving learners a transcript of the listening text is a valid way of assessing listening skills.	202	75	115*	12	,5693	,49640
27. Dictation is a kind of discrete-point testing.	202	135	36*	31	,1782	,38365
28. Inference questions based on intelligence are avoided in listening tests.	202	111*	59	32	,5495	,49878
29. Asking learners to listen to names or numbers is called intensive listening.	202	112	54*	36	,2673	,44366
30. In selective listening, learners are expected to look for certain information.	202	178*	12	12	,8812	,32437
Listening-Total	202				6,8317	2,13742
Assessing Writing						
31. Giving two options to learners and asking them to write about one ensure reliable and valid scoring.	202	147	35*	20	,1733	,37942
32. Analytic scoring is used to see the strengths and weaknesses of learners.	202	162*	17	23	,8020	,39950

33. The parts of a scoring scale and the scores in each part do not change for different levels of learners.	202	61	123*	18	,6089	,48921
34. When there is a disagreement between the scores of the two raters, they score the written work again.	202	139	32*	31	,1584	,36604
35. Learners are required to write about at least two tasks in the exam rather than one task.	202	93*	89	20	,4604	,49967
36. Giving restrictive prompts/guidelines to learners for the writing task is avoided.	202	72	100*	30	,4950	,50122
37. Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.	202	150	34*	18	,1683	,37508
38. Using visuals which guide learners for writing poses a problem.	202	53	135*	14	,6683	,47199
39. Holistic scoring is used to see whether the learner is proficient or not at the end of the term.	202	117*	28	57	,5792	,49491
40. Analytic scoring leads to greater reliability than holistic scoring in writing.	202	119*	34	49	,5891	,49322
41. In controlled writing, learners have the chance to convey new information.	202	93	75*	34	,3713	,48435
42. Classroom evaluation of learning in terms of writing is best served through analytic scoring rather than holistic scoring.	202	89*	49	64	,4406	,49769
43. Irrelevant ideas are ignored in the assessment of initial stages of a written work in process writing.	202	109	67*	26	,3317	,47199
44. Providing a reading text for writing is a way of assessing writing skills.	202	116*	60	26	,5743	,49568
45. Mechanical errors (e.g., spelling and punctuation) are dealt with in the assessment of later stages of a written work.	202	129*	54	19	,6386	,48160
Writing-Total	202				7,0594	2,15943
Assessing Speaking						
46. When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.	202	111	76*	15	,3762	,48564
47. Giving learners one task is enough to assess speaking skills.	202	49	145*	8	,7178	,45118
48. Interlocutors' showing interest by verbal and non-verbal signals poses a problem.	202	66	111*	25	,5495	,49878
49. When it becomes apparent that the learner cannot reach the criterion level, the task is ended.	202	78*	83	41	,3861	,48807
50. Using holistic and analytic scales at the same time poses a problem.	202	67	74*	61	,3663	,48300
51. Reading aloud is a technique used to assess speaking skills.	202	96*	94	12	,4752	,50063
52. In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.	202	163*	19	20	,8069	,39569
53. In interactive tasks, more than two learners pose a problem.	202	72*	113	17	,3564	,48014
54. The interlocutor gives the score when the learner is in the exam room.	202	87	89*	26	,4406	,49769
55. In a speaking exam, production and comprehension are assessed together.	202	166*	26	10	,8218	,38365

56. Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills.	202	95*	90	17	,4703	,50036
57. Discussion among learners is a way of assessing speaking skills.	202	179*	13	10	,8861	,31843
58. A checklist is a means of scoring oral presentations in in-class assessment.	202	170*	20	12	,8416	,36604
59. When the focus is to assess discourse, role plays are used.	202	166*	17	19	,8218	,38365
60. In peer interaction, random matching is avoided.	202	65*	109	28	,3218	,46832
Speaking-Total	202				8,6386	2,26219
LAK-TOTAL	202				32,0990	5,43895

* Refers to the correct answers.

4.1.1. Assessing Reading

The first fifteen items in the scale aim to measure the participants' knowledge level in assessing reading skills. Based on the participants' mean scores in this part, it is seen that the highest score belongs to the reading assessment ($\bar{X}=9.5693$), which means that EFL teachers in Turkey are more knowledgeable in assessing reading skills than listening, writing, and speaking. The items in the assessing reading part according to the correct answers of the participants are revealed from highest to lowest.

Among the items between 1 to 15, the highest scored item is 13, "Reading texts in a reading exam include various genres (essay, article, etc.)". 87% percent of the participants, that is 176 EFL teachers, have answered this item correctly. Heaton (1990) addresses that including various types of text would increase learner motivation as they are exposed to target language as well as enhancing the reliability of the reading exam. Besides, the authenticity of the text needs to be taken into consideration for incorporating different genres such as articles, newspapers, letters, etc. that are extracted from real-life would result in positive washback. However, 36 teachers were not aware that different genres should be involved when assessing reading.

"Asking learners to summarize the reading text is a way of assessing their reading skills." item that had the second highest score in the scale. The answer to this statement should be true. Summarizing a reading text can be used in assessing reading among many other techniques (Alderson, 2000; Brown & Abeywickrama, 2010; Heaton, 1990; Hughes, 2003) to check learners' comprehension of the text by expressing what they have understood in written or oral form. Most of the teachers (84%) answered this item correctly, while 15% of the teachers in total either chose the false option or did not know the answer to this item. It

can be deducted that EFL teachers were knowledgeable in the assessment techniques that can be utilized for assessing reading. The reason why teachers did not answer correctly, maybe because summaries are mostly associated with writing skills, hence leading to a confusion.

The third highest scored item is “Taking vocabulary difficulty into consideration is necessary in assessing reading skills”. Vocabulary and reading skills are directly related to each other. Thus, while selecting a reading text, there are many factors to be considered, which are the length, content, genre, and the difficulty of the text (Harris, 1969; Hubley, 2018). The vocabulary items in the text should be appropriate for learners’ level rather than being too simple or too complex for comprehension (Harris, 1969). As a matter of fact, Hubley (2018) suggest that unknown vocabulary should be limited to 10% of the total vocabulary. EFL teachers were mostly aware of the importance of difficulty of the lexical items used in a text since 160 (79%) teachers have answered this item correctly while only 6 out of 202 the teachers didn’t have an idea about the issue, and 36 (17%) of the teachers answered incorrectly.

The next item on the scale also has a high mean score, which is “In top-down approach, assessment is on overall comprehension of the reading text”. Alderson (2000) asserts that in top-down approach learners are expected to obtain the gist of the text and compose a general prediction. This item has been answered correctly by 151 teachers (%75). The teachers who chose don’t know option (n=30) is more than the teachers who gave a wrong answer (n=21). Besides, the teachers who didn’t know the answer is one of the highest in assessing reading part, which may be due to being unfamiliar with the reading strategies of top-down or bottom-up. It is clear from the findings that EFL teachers were mostly knowledgeable about top-down approach to assessing reading.

Another item that has been answered correctly by most teachers is “Using ungrammatical distractors in multiple choice questions in a reading exam is a problem”. The purpose of assessing reading is to check learners’ comprehension of the text rather than their knowledge of vocabulary or grammar. Even though the grammatical correctness of the distractors may be overlooked by the teachers, teachers should not involve ungrammatical options in a multiple-choice question (Heaton, 1990). 72% of the teachers (n=147) stated that ungrammatical options are a problem when assessing reading. Among the rest 40 of the teachers gave an incorrect answer while 15 of the teachers did not bear the knowledge for

grammatical appropriateness of the distractors. It is understood that EFL teachers generally know that the correct grammatical structures should be used in multiple choice options.

When the item “Including not stated/doesn’t say along with true/false items has advantages over true/false items” is examined, it is seen that 144 teachers (71%) were aware of the chance factor of true/false items. Alderson (2002) states that adding a third option of not given/stated decreases the possibility of chance of guessing, which is 50%, on the part of the students. Another possible way to eliminate the disadvantages of true/false items is to ask the learners to elaborate the false option. The mean scores of the teachers on this item indicate that the teachers were informed about the disadvantages of true/false items while the rest of the participants did not remark that a third option is advantageous for assessing reading with true/false items.

“The more items a reading text is followed, the more reliable it becomes” is another item on the scale that has a high mean score. 69% (n=140) of the teachers have correctly answered this item. The number of the teachers choosing the don’t know option is 21 while there are 41 teachers who have given an incorrect answer. In order to increase the reliability of a reading assessment, teachers should provide more questions to the reading text (Green, 2014). Therefore, it is observed that more than half of the teachers knew how to make a reading assessment more reliable by including more items.

Another high mean score on the scale belongs to “The language of the questions is simpler than the text itself”. The objective behind assessing reading is to figure out whether students have understood the text. Therefore, when the questions of a reading text are beyond learners’ linguistic capabilities, it becomes a problem for comprehension (Hughes, 2003), which may lead to validity problems since the test did not measure what it supposed to measure, which is the learners’ ability to understand the text not the items (Harris, 1969). Thus, Hubley (2018) suggests that items should be lower than the level of the reading text. The number of the correct answers on this item is 140 (69%) teachers. 43 teachers among the rest answered incorrectly while 19 of the teachers did not have any idea about what the level of the questions in a reading assessment should be. The fact that more than half of the teachers answered this question correctly indicates that they are considerate of the level of the questions of a reading text.

The next item “Using the same words in the correct option as in the text is not a problem”, to which the answer should be false. The teachers who selected the false option for this item is 123 (60%). More than half of the teachers put forth that the words in a reading text should

not appear in the correct option of a multiple-choice question. To serve its purpose, which is to assess interpretation of the reading text rather than detecting the recurring vocabulary in a text, it is wise not to use the same words in the reading passage (Harris, 1969). However, 61 (30%) did not believe that it poses a problem for the purposes that a reading assessment should bear. The teachers who selected the don't know option is 18 (8%), which is quite low. As a result, the teachers who think that using the same word is wrong have doubled the ones who think it is not wrong.

The tenth item that has indicated a high mean score is "One reading text is enough to be included in a reading exam", to which the participants gave 113 (55%) correct answers. The correct answer for this item is false and the mean score of this item is slightly higher than the average. However, 70 (34%) of the participants, which is one third of the total number, did not project that more than one reading passage needs to be included in a reading exam. On the other hand, 9% (n=19) of the teachers had no idea about the number of the texts to be involved in a reading exam. The number of the items in an exam is directly related to reliability as the more sample of the learner performance is collected the more reliable the assessment becomes (Harris, 1969; Heaton, 1990). However, within a classroom assessment perspective the reliability may not be given the utmost importance by the teachers by means of determining the reliability coefficient of an exam or other statistical procedures, which may be due to the workload of the teachers (Popham, 2018). In addition, if a teacher is constantly assessing and collecting evidence of the learners' reading performance, the reliability would be enhanced. Besides, the context of this study is K-12 schools, which involves primary, middle, and high school a very diverse range of age groups including young learners. Consequently, all of which may have interfered with the teachers' thinking process while answering this item.

"Simplification of reading texts is avoided" is another item that is answered correctly by 111 teachers (54%). Selecting a reading text should depend on some careful considerations such as the length, genre, topic, vocabulary, difficulty, etc. (Hubley, 2018; Hughes, 2003). If the difficulty of the text is excessively beyond the current level of learners' proficiency, it may lead to negative washback on the part of the students by means of unfamiliar vocabulary items, unknown or complex grammatical structures, and length of the text being too long for learners to comprehend. However, 60 teachers (29%) did not believe that reading text should be adjusted according to the learners' proficiency level. The other teachers (n=31) did not possess a sort of knowledge about simplifying the linguistic components of a reading

passage, which is the second highest item that the participants did not present any knowledge in the assessing reading part. The reason behind this could be due to teachers' dependency on the reading materials in the coursebook they use in their classrooms, thus, not being able to modify it based on their learners' needs. However, it is shown that more than half of the teachers indicated their knowledge about the fact that if a text is beyond the learners' language level, it should be simplified.

The next item has received a knowledge level at a medium degree, which is "Cloze test is used for assessing the main idea of a text". The answer to this item should be false since the gaps in a cloze test is systematically determined by deleting the every -nth word of the text on the contrary to gap-filling tasks, in which the gaps are controlled by the teacher (Alderson, 2000). Since the teachers have no control over the gaps in a cloze test, the overall comprehension of the text is not needed when the gaps are a part of phrase or a grammatical structure that can be predicted by the learners. The half of the teachers (n=106) knew that cloze tests may not be related to understanding the gist of the passage, but to the smaller units in the passage. The other half of the teachers, 19 (9%) of which did not bear any idea on the issue, have given incorrect answers.

The first item that the teachers had low mean scores is "Errors of spelling are penalized while scoring". This item has received 97 correct answers (47%), which is slightly lower than the half of the participants. This item is also related to the purposes of reading assessment, which is to assess reading ability rather than grammar or spelling. Therefore, Hughes (2003) suggests that penalizing the grammar and spelling mistakes of the learners in a reading exam decreases the validity of the assessment since they are related to writing assessment rather than reading. The number of the teachers who believed that errors of spelling should have an impact on the scoring process (n=78) is close to the ones who doesn't believe as such (n=97). When the teachers who chose don't know option (n=27) is added, it is clear that teachers were not aware of the fact that spelling can be ignored while scoring a reading exam. Teachers' tendency to correct every mistake or error that their students make may have led this result.

"In a reading exam, using a text learners have encountered before is not a problem" is the second to lowest based on the mean scores of the participants. This item has been answered correctly by 96 teachers (47%). The number of the teachers with incorrect answers (n=78) combined with the ones who did not know if this item is true or false (n=28) is 106 (52%) in total, which is slightly higher than the correct answering teachers. It may be wise to use a

text that the students are familiar with the topic, but if it is a text that they have encountered before, it not only masks learners' actual performance but also reduces the reliability of assessment (Hughes, 2003). Therefore, it can be concluded that teachers' knowledge about the criteria for selecting a reading text is not high.

The final item on the reading assessment, which is the one that the teachers had the lowest mean score on the reading assessment, is "When asking several questions about a reading text, all questions are independent of each other". Only 59 teachers (29%), by selecting the true option, have answered this item correctly, which is quite low. When the number of the false (n=136) and don't know (n=7) options are compared, it is observed that nearly all of the teachers had an idea about the characteristics of the questions about a reading exam regarding their relevance to each other. It is observed that 70% of the teachers have low level of knowledge about preparing sound reading questions. In other words, they were unaware of the risks that it carries, which is, according to Alderson (2000), questions that are relevant to each other may influence the students answers to other questions.

Overall, the teachers were mostly knowledgeable in the areas of utilizing various types of genres, reading assessment techniques or strategies of summarization, top-down approach, true/false items, cloze tests, and the characteristics of multiple-choice questions as in grammatical correctness, not including the same word in the correct option, and the linguistic properties a text and items should bear such as simplifying the language of the text and question in addition to reliability related areas as in increasing the number of samples from learner performance by using more items and reading texts. On the other hand, they were not informed about the not including spelling in the scoring criteria, using a text that students have seen before. Teachers were the least knowledgeable about that a question should be unrelated with other questions of a reading. In conclusion, among 15 items related to assessing reading, the participants had higher mean scores than the half of the total score in 12 of them while in 3 of the items, participants had lower mean scores. To conclude, it is possible to say that EFL teachers in the K-12 schools in Turkey have high knowledge level in assessing reading skills. These findings coincide with the results of another study by Ölmezer-Öztürk and Aydın (2019a), in which the most knowledgeable are of assessing language skills was found to be reading in the context of tertiary education.

4.1.2. Assessing Listening

The items from 16 to 30 aims to measure EFL teachers' knowledge in assessing listening skills. From the table above it is seen that EFL teachers had the lowest mean score in listening ($\bar{X}=6.8317$), which illustrates that EFL teachers in the Turkish K-12 context are least knowledgeable in assessing listening compared to reading, writing, and speaking. The mean scores of each item are unveiled from the highest to lowest based on the participants' answers in this part.

The highest mean score that the teachers have demonstrated belongs to the item "In selective listening, learners are expected to look for certain information". When the aim is to scan the listening text for a particular detail such as a name, a number, or a place selective listening is expected to be performed (Brown & Abeywickrama, 2010). Unlike extensive listening, acquiring a global understanding of the text is not the main concern. In total, 178 teachers (88%) provided a correct answer to this item. When the numbers of the participants with incorrect answers ($n=12$) and the participants who didn't have an idea about the item ($n=12$) are examined, it is seen that they are the same. The mean scores of the teachers indicate that they were quite informed about the purposes of selective listening tasks.

The second highest mean score is measured in the "A listening cloze test is a way of selective listening" item, to which the answer is true. More than half of the teachers have answered this item correctly, which is 159 (78%). Besides, 20 (10%) teachers have stated that for selective listening, a listening cloze test cannot be used while 23 (12%) did not possess any information regarding the ways of selective listening or the purpose of a listening cloze test. Buck (2001) defines a listening cloze test as similar to a regular cloze test for assessing reading, learners are given a written form of a speech in which some words are omitted systematically or randomly. However, teachers should be aware of the possible problems with these task types since the similarity between the regular and listening cloze test may lead learners by completing the text by reading it rather than listening. In conclusion, the teachers were mostly aware of the fact that cloze tests can be utilized to assess listening selectively.

"Including redundancy (e.g., what I mean to say is that...) in a listening text poses a problem" is the third item that the teachers had high mean scores. Redundancy can be pictured as gestures, intonation, speed, hesitations, repetitions, fillers, etc. that occurs naturally in a speech (Buck, 2001; Heaton, 1990). 129 of the teachers (63%) have presented their knowledge about redundancy by choosing the false option for the item. In a speech that

takes place in a real-life context, these expressions are naturally uttered. Therefore, redundancy reflects the properties of spoken language corresponding with real-life situations and needs to be involved in terms of authenticity (Abeywickrama, 2018b). Even though two-thirds of the teachers believed that a listening text should mirror spoken language with redundancy expressions, approximately one-third of the teachers (27%) believed that redundancy is a problem. Finally, a minority of the teachers (8%) were not knowledgeable enough regarding redundancy.

The next item for assessing listening is “Giving learners a transcript of the listening text is a valid way of assessing listening skills” to which the answer should be false. This item has been answered correctly by 115 teachers (57%). More than half of the teachers were aware that giving learners the transcript of the listening text hampers the validity of assessment. Accordingly, it is not a sound way of assessing listening as a result of the fact that it becomes more of a reading assessment than listening (Brown, 2003; Heaton, 1990). The incorrect responses (n=75) were significantly higher than the ones who did not present their knowledge (n=12) about the validity of the assessment when learners are given the transcript. One explanation that can be made upon why 43% of the teachers have given an incorrect answer is that the teachers may be using songs and giving the lyrics with gaps for the learners to fill in, thus, they may be deducing from their classroom practices.

“Inference questions based on intelligence are avoided in listening tests” is the last item that also has been answered correctly by more than half of the teachers. The correct answer rate for this item is 55% (n=111) while 59 (29%) of the teachers were incorrect, and don’t know option was selected by 32 (16%) teachers. Buck (2001) explains that the extent of background knowledge influences the inferences made from a listening text. Therefore, eliminating the factor of background knowledge on the part of the learner is desired to make the comprehension dependent on their listening abilities. It is not wrong to say that more than half of the teachers were informed about that intelligence or background knowledge are not among the construct that is aimed to be assessed in listening, thus, it should not interfere with the listening performances of the learners.

The following item is answered correctly by the half of the participant teachers in the scale: “In discrete-point testing, comprehension is at the literal/of local level”. Among the approaches to listening assessment, discrete point is independently measuring the crucial bits and language elements in isolation based on the idea that it will represent the overall listening abilities (Buck, 2001). In total, 50% of the teachers were aware that discrete point assessment

is not interested in the global understanding of the listening text but deals with the comprehension at local level. 50% of the teachers (n=100) were correct to think that the answer to this item is true. What stands out in the findings is that the teachers who did not know the answer to this item (37%), which is the highest in assessing listening part, have nearly tripled the number of those who have given a wrong answer (13%). One conclusion that can be drawn upon this is that the teachers may not be familiar with the terminology occurring in the item either discrete point testing or literal/local level regarding listening.

The first item that have indicated a lower mean score in the scale is “Spelling errors are ignored in scoring the dictation”. The same number of teachers gave a correct (n=95) and incorrect answer (n=96) to this item. In addition, 11 teachers, which is the lowest number in the assessing listening part, did not know whether spelling errors should be ignored or not while scoring the dictation task. As a result, it can be deduced that most of the teachers had an idea about the scoring criteria for dictation regarding spelling errors. Dictation is a task that can be incorporated in assessing both listening and writing, yet spelling is not the main concern of listening assessment (Brown, 2003). Thus, dictation task should not be considered as a spelling test, hence can be ignored (Buck, 2001). Based on the table, the total number of the incorrect answers (53%) is higher than the correct ones (47%). It is observed that teachers’ knowledge in the scoring procedures in dictation regarding spelling errors tasks is slightly low. This finding is similar to the findings on an item related to penalizing of spelling errors in the assessing reading part, where more than half of the teachers believed that spelling is penalized.

In the item “Using dictation diagnostically in assessing listening skills does not pose a problem”, the teachers have also demonstrated a low mean score. The item has been answered correctly by 90 teachers (45%). Similar to the item above, this item is also related to dictation, in which integrative approach is adopted to assess learners’ listening and writing skills (Brown, 2003). Therefore, it can be utilized to detect learners’ problematic areas in writing, listening, grammar, and spelling (Buck, 2001). The findings present that in total, 112 teachers (55%), consisting of 78 incorrect and 34 don’t know answers, didn’t know that dictation can be used for diagnostic purposes even though it is a commonly preferred technique by many teachers (Brown, 2003).

“Errors of grammar or spelling are penalized while scoring” is another item that is answered correctly by less than the half of the teachers. The answer to this item is false since the aim behind assessing listening is to assess learners’ listening comprehension rather than their

knowledge in spelling and grammatical structures even though listening is a complex skill that involves many components of language (Abeywickrama, 2018a). The number of the teachers who selected the false option, knowing that spelling and grammar related mistakes should be ignored while scoring, is 83 (41%). When the participants selecting the don't know (n=21) and true option (n=98) are investigated, it can be reported that 60% (n=119) of the teachers were not knowledgeable in the scoring process of assessing listening regarding the learners' syntax errors. Listening is challenging skills, compared to other skills, to define since it involves many other components occurring in real life conversations (Abeywickrama, 2018a). This could be the reason why teachers were not sure about which listening constructs needs to be involved in the assessment. In addition to this item, 7th (assessing reading) and 19th (assessing listening) items in the scale were related to the penalization of spelling errors, and the mean scores of these items indicated that teachers believed that spelling should be included in the scoring.

The next item is "Any type of listening text is used for note-taking". 76 teachers (37%) have answered this item correctly by stating that it is false. It is a striking finding that the number of the participants with incorrect answers (63%), including the ones choosing true (n=98) and don't know options (n=28), has outnumbered the participants who knew the correct answer. Note-taking is a process that learners follow while listening to a text to write their notes about the questions provided or the text overall (Heaton, 1990). Therefore, taking notes, which wouldn't suit for short texts, would be useful when the listening text is longer (Hughes, 2003). As a result, it can be concluded that more than half of the teachers were unaware that notetaking may not be a proper technique for any type of listening text in terms of its length.

Another item that has been answered incorrectly by more than half of the participants is "Using reading texts for listening purposes poses a problem". 64% of the teachers (n=128) believed that reading texts can be used for listening purposes. However, spoken and written forms of language have many different components such as formality, redundancy, non-verbal cues, and intonation etc. (Buck, 2001). Besides, authentic listening text should be used, therefore, using a reading text would not mirror learners' listening performances in real-life situations as it eliminates the characteristics of spoken language such as stress, hesitations, correction, intonation and so forth (Brown, 2003). 36% of the teachers (n=74) were aware of the fact that reading texts are not suitable to be used for listening purposes. A possible explanation for this finding can be that teachers' own practices of listening

assessment depending on the educational level of their schools. Dialogues used in the coursebooks for lower grades are likely to be presented with visuals and speech bubbles, in which the dialogues are written.

“Asking learners to listen to names or numbers is called intensive listening” is sixth item that demonstrates low mean scores of the participants. The number of the teachers choosing the false option, which is the correct answer, is 54 (26%) while 112 teachers (55%) thought that in intensive listening, specific details are looked for, and 36 teachers (18%) did not know the aim of intensive listening. Words, intonation, and phonemes are the examples of what to look for in intensive listening whereas selective listening deals with specific details of factual information, names, numbers, etc. (Brown & Abeywickrama, 2010). Interestingly, 21st and 30th items were items that the teachers demonstrated the highest mean scores in assessing listening part. Thus, it is deduced that the teachers are likely to know that selective listening is interested in specific information in the text yet may not be informed about the purposes of intensive listening.

Another item that is related to scoring is “Scoring in note-taking is straightforward”. This item has received considerably low mean score. Only 41 of the teachers (20%), which is one-fifth of the total participants, have answered this item correctly. The remaining 80% (n=161) of the teachers were not knowledgeable in the scoring procedures of note-taking. The teachers who selected the don’t know option (n=54) for the item is one of the highest in assessing listening part. Brown (2003) expresses that scoring is a complicated and time-consuming process in note taking in addition to being subjective. It is shown that teachers were not knowledgeable about how scoring takes place in note taking. The reason behind this could be that even if note taking is used in listening assessment, it may not be included in the scoring process.

The second to last item in assessing listening part is “Phonemic discrimination tasks (e.g., minimal pairs such as sheep-ship) are examples of integrative testing”. Integrative approach stresses the importance of using language as a whole rather than the knowledge of language elements in isolation, which is directly the concern of discrete point approach (Buck, 2001). As a result, phonemic discrimination tasks are examples of discrete point testing. However, most of the teachers (81%) have answered this item incorrectly while the rest of the participants (19%) gave a correct answer. It is clearly shown that teachers are not knowledgeable about integrative assessment tasks. The reason behind could be the teachers’ unfamiliarity with the terminology used in the item regarding integrative assessment.

The last item, which also have the lowest mean score among the 15 items in this part, “Dictation is a kind of discrete-point testing”. The number of the participants who gave a correct answer for this item is 36 (18%), while 31 (15%) did not have an idea about whether the item is true or false. Besides, 135 of the teachers (67%) have chosen the true option, giving a wrong answer. As mentioned previously, dictation involves both listening and writing skills, hence integrative testing (Brown, 2003). It is clearly observable that teachers have significantly low level of knowledge in discrete-point testing techniques regarding dictation.

From the discussion of the abovementioned items, EFL teachers in the K-12 context were mostly knowledgeable in the selective listening, using listening cloze-tests for selective listening, characteristics of the listening text in terms of background knowledge, redundancy and authenticity involved. On the other hand, the teachers’ knowledge level in assessing listening remained low regarding discrete-point testing, scoring and purposes of discrete-point testing, intensive and integrative assessment approaches, scoring criteria whether to include spelling and grammar, and selection of listening text. In conclusion, among the 15 items in assessing listening part, 5 of the items have presented high mean scores while 10 of the items remained below the average. Supporting the evidence from other studies (Ölmezer-Öztürk & Aydın, 2019a), this research found that EFL teachers’ LAK in assessing listening is notably low.

4.1.3. Assessing Writing

The items from 31 to 45 aim to measure EFL teachers’ knowledge level in assessing writing. The participants’ mean scores ($\bar{X}=7,0594$) in this part demonstrates that EFL teachers’ knowledge level is lower than the half of the total score, which puts assessing writing in the third place. The mean scores obtained from the items, based on the participants’ answers, are individually revealed, and discussed in this part from the highest to lowest.

The item “Analytic scoring is used to see the strengths and weaknesses of learners” have the highest mean score in assessing writing part of the scale. Scoring is one of the most significant aspects of writing assessment. One approach to scoring written works of the learners is analytic scoring, in which different component of writing such as content, organization, vocabulary, etc., is separately scored (Hughes, 2003). Most of the teachers (80%) have answered this item correctly while a minority of the teachers (11%) didn’t know

what analytical scoring is used for, which is slightly higher than the number of the teachers (9%) who didn't think that strengths and weaknesses of the learners can be scrutinized with analytic scoring. However, utilizing analytic scoring rubric can be efficient for providing diagnostic evidence on learners writing performances (Weigle, 2012). The teachers were mostly knowledgeable about the fact that analytic scoring is useful for diagnosing learners' strengths and weaknesses.

The second highest mean score that the participants obtained in this part belongs to "Using visuals which guide learners for writing poses a problem". 135 of the teachers (67%) knew that this item was false, therefore, provided a correct answer. On the other hand, the rest of the participants (33%) were not aware of the fact that visuals are a helpful resource to stimulate learners' creativity (Hughes, 2003; Weigle, 2012). It is clear that the teachers with correct answers have doubled the teachers with incorrect answers.

The following item, which is "Mechanical errors (e.g., spelling and punctuation) are dealt with in the assessment of later stages of a written work", possesses the third highest mean score. 129 teachers (64%) were correct to believe that mechanical errors should be considered later in the writing stages. In order not to interfere with the flow of ideas that learners put on their written work, syntax errors can be ignored in the initial stages of writing (Weigle, 2002). Among the rest of the teachers, 19 of them (9%) didn't know when to deal with mechanical errors while 54 teachers (27%) were incorrect. However, Brown (2003) suggests that the main concern in the first stages of writing should be on the learners' ideas, organization of thought, irrelevant subjects rather than minor lexical or grammatical errors. Overall, it is possible to say that teachers' knowledge in error correction in the first stages of writing is higher than the average.

The highest fourth item regarding the participants' mean scores is "The parts of a scoring scale and the scores in each part do not change for different level of learners". It is known that every learner has their own educational needs which can be based on their educational program or curriculum. Therefore, teachers have the chance to select or adjust the instrument for assessing their learners' performances. 61% of the teachers (n=123) knew that a scoring scale can be modified according to the level of the learners whereas 39% (n=79) did not bear the knowledge of adjustments can be made on the criteria on a scoring scale on the contrary to what Brown (2003) proposes as scoring scales can be tailored to the curricular or educational needs of the learners based on the emphasis given to writing traits. Still, the

teachers are considered to be knowledgeable about alterations in the scales are required for varying levels of learners.

“Analytic scoring leads to greater reliability than holistic scoring in writing” is another item that the participants had high mean scores. The answer to this item is true as scoring each unit of writing determined in the scale separately would make the results more reliable (Hughes, 2003). The number of the correct answers is 119 (59%), slightly higher than the half of the participants. The participants who selected the don’t know option (n=49) is higher than the ones answering incorrectly (n=34). The reason could be that teachers were not aware of the terminology regarding the types of the scoring scales. In conclusion, more than half of the teachers were knowledgeable about the reliability factor of different scoring scales.

The next item is also related to scoring scales, which is “Holistic scoring is used to see whether the learner is proficient or not at the end of the term”. Analytic scoring deals with learners’ problematic areas in detail as mentioned above. Therefore, it can be utilized to provide evidence on the learners’ strengths and weaknesses for adjusting the classroom instruction accordingly (Brown, 2003). On the contrary, holistic scoring can provide information on learners’ overall writing proficiency (Hughes, 2003). 117 teachers (58%) have stated that the answer is true and gave a correct answer. Surprisingly, the number of the don’t know option (n=57) has doubled the incorrect answers (n=28). Similar to the item above, it can be uttered that the teachers who didn’t answer correctly are not familiar with the terms used in the item in terms of holistic scoring. However, it wouldn’t be wrong to say that were informed about when to use holistic scoring in line with its aims.

More than half of the participants have given a correct answer to the item “Providing a reading text for writing is a way of assessing writing skills”. 116 (57%) is the number of the correct answers while 60 (30%) is for incorrect, and 26 (13%) is the number of the don’t know option chosen for the item. 43% of the teachers didn’t think or know that a reading text can be used as a prompt for a writing task. However, in order to activate learners’ thinking process, a prompt in various forms, such as visuals, reading texts, a quotation, etc., should be provided to learners as a basis that all learners can draw upon (Weigle, 2002). One thing that needs careful consideration is that giving a reading text as a prompt may lead the learners’ reading abilities to interfere with their writing in the case that they don’t understand the reading text or excessively use the same vocabulary items occurring in the text. All in all, it is seen that the teachers’ knowledge in giving a reading text as a prompt for writing tasks is higher than the average.

“Giving restrictive prompts/guidelines to learners for the writing task is avoided” the answer to this item should be false since if the prompts or guidelines are not limited to a certain point in terms of length, genre, audience, etc., scoring of the writing performances would vary excessively (Weigle, 2002), which may negatively affect the reliability of the scoring procedure (Heaton, 1990). The item has answered correctly and incorrectly by nearly the same number of participants. While 100 teachers (49%) gave a correct answer, the other half of the teachers (51%) were unaware of the fact that prompts should be restricted by means of particular writing components. A conclusion that can be drawn upon this is that teachers’ knowledge about providing guidelines for a writing task is at a medium level.

The items below have remained below the half of the total score that can be obtained in assessing writing part of the scale. The first item receiving a low mean score is “Learners are required to write about at least two tasks in the exam rather than one task”. This item can be correlated with reliability issues in assessing writing, since by giving at least two tasks will provide more samples of the learners’ performances, hence higher reliability (Harris, 1969; Heaton, 1990). However, less than half of the teachers (n=93) did select the correct option for the item. In total 109 teachers (54%), 89 being incorrect and 20 being don’t know, were unaware of the fact that one task is not enough to assess learners written performances. Thus, the teacher’s knowledge about the amount of the tasks to be involved in writing exams is found slightly low.

“Classroom evaluation of learning in terms of writing is best served through analytic scoring rather than holistic scoring” is the item that many teachers didn’t have an idea about, that is, 64 of the teachers (32%), which is the highest in assessing writing part, have selected the don’t know option. 89 teachers (44%) chose the correct and 49 (24%) teachers chose the incorrect answer for the item. The characteristics of analytic and holistic scoring have been mentioned previously in terms of their foci. While the former deals with the scoring of the writing components individually, therefore, can be used for diagnostic purposes (Brown, 2003), the latter deals with the overall scoring of writing performances, therefore lack enough feedback on the problematic areas in the written works of the learners (Heaton, 1990). Consequently, when the concern is classroom evaluation, as it is emphasized in the item, analytic scoring is the most appropriate choice (Brown, 2003). Based on the answers of the participants, it is seen that they didn’t have enough knowledge regarding analytical scoring within the scope of classroom evaluation, which is interesting since the highest scoring item in assessing writing part of the scale was related to analytical scoring, and other

items including analytical scoring and holistic scoring were also answered correctly by many teachers. A conclusion that can be drawn upon this can be that the teachers may not be familiar with scoring techniques that would serve better in terms of classroom evaluation.

Another item presenting a low mean score by the participants is “In controlled writing, learners have the chance to convey new information”. In controlled writing tasks, which have an emphasis on the form, do not provide the learners a chance to convey meaning or information from one to another. (Brown, 2003). Thus, it becomes more of a demonstration of learners’ writing abilities than an information transfer in an authentic context. The number of the correct answers is 75 (37%) and 93 of the teachers (46%), which is higher than the correct responses, were incorrect. In addition, 34 teachers (17%) were not sure whether the item is true or false. It is clear that most of the teachers were unaware of the content that may occur in a controlled writing task. The reason behind can be related to the term controlled writing and teachers’ unfamiliarity with it.

The next item, which is “Irrelevant ideas are ignored in the assessment of initial stages of a written work in process writing”, has a low mean score. The exact number of the participants with correct response is 67 (33%) which is quite below than the average. When the number of the incorrect answers (n=109) and don’t know options is combined (n=26), the number of the teachers who were not knowledgeable (66%) have doubled the ones who knew the correct answer, which should be false. According to Brown (2003), in different stages of writing, the focus should also differ in that the initial stages for dealing with learners’ ideas, organization, and general meaning while minor lexical or grammatical errors can be dealt with in the later stages. Consequently, in the first stages of a written work irrelevant ideas need to be the main concern. Teachers’ low level of knowledge about what is the main concern in the stages of process writing can be observed based on the findings. In contrast to another finding of the study, in which the participants showed higher knowledge level about dealing with the mechanical errors in the later stages.

“Giving two options to learners and asking them to write about one ensure reliable and valid scoring” is another item that got one of the lowest mean scores. Validity of an assessment is briefly its ability to measure what it is supposed to measure (Heaton, 1990). Besides, it would make the scoring process more unreliable since comparison of the writings is not possible on an equal basis since the difficulty of these options may not be the same (Weigle, 2002). Only 35 teachers (17%) have given a correct answer, which is highly low than the average. While 20 teachers (10%) did not indicate a form of knowledge by selecting don’t know

option, 147 teachers (73%) have given a wrong answer. It is evident that teachers were not aware that providing options to learners to select for writing would negatively impact reliability and validity of the assessment.

Another item, showing one of the lowest mean scores, is “Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills”. The appropriacy of the prompt given to learners should carry some features, such as clear, specific, not requiring a world knowledge, yet enough to stimulate their background knowledge (Harris, 1969; Weigle, 2002). Thus, giving learners an opinion that can be vague or ambiguous is insufficient to direct learners to write about the desired subject (Harris, 1969). A highly small number of the teachers (n=34) believed that validity of the assessment cannot be ensured by giving an opinion as a prompt. The other 150 teachers (74%), on the other hand, did not see a problem with providing an opinion in a writing task as a prompt, which is the highest incorrect answer in the assessing writing part. Finally, 18 teachers (9%) did not have sufficient knowledge to answer this item. Thus, the overall knowledge of the teachers regarding the quality of the prompt for writing tasks is not enough.

The lowest mean score in this part is on the item “When there is a disagreement between the scores of the two raters, they score the written work again”. The answer of this item is false. There are some measures that can be taken to ensure the reliability of the scoring procedures. Raters can score the writings anonymously to eliminate bias, or they can score the writings twice for intra-rater reliability. Another option is involving a second rater in the procedure, which is called inter-rater reliability. However, if there is a disagreement between the raters, a third rater needs to score the writings of the learners (Weigle, 2002). The participants’ answers to this item were interesting in that the number of the teachers who gave a correct answer (n=32) is the same with the ones who don’t know whether this item is true or false (n=31). Besides, 139 teachers (69%) believed that the raters should score the writings. It is highly clear that teachers are not knowledgeable in the scoring procedures of writing assessment when a discrepancy occurs.

All in all, The EFL teachers in the K-12 context were mostly aware of the analytic and holistic scoring in terms of their advantages and reliability, prompts that can be used such as visuals or a reading text, when to focus on mechanical errors, whether to adjust the scales based on the learners’ needs or not. However, they were not knowledgeable enough in controlled writing, when to deal with irrelevant ideas, the amount of writing tasks involved, using opinion-based prompts, giving options to learners, and scoring procedures when two

raters are involved. Among 15 items related to writing assessment, the teachers had high mean scores in 7 of them while they had lower mean scores in 8 of the items. Similar to the findings of the other studies that have employed the same scale by Ölmezer-Öztürk and Aydın (2019a) and Genç, Çalışkan, and Yüksel (2020), the participants' knowledge in assessing writing is below average.

4.1.4. Assessing Speaking

The items from 46 to 60 aims to measure EFL teachers' knowledge in assessing speaking skills. Based on the table above, it is observed that compared to reading, listening, and writing, speaking is the second highest regarding the participants' mean scores ($\bar{X}=8.6386$), which is higher than the half of the total score that can be obtained in this part. Below, the participants' mean scores in each item are displayed from the highest to the lowest.

The highest mean score in assessing speaking part belongs to "Discussion among learners is a way of assessing speaking skills". Among the task types that can be used to assess learners' speaking skills, discussion is one of the ways (Fulcher, 2003; Luoma, 2004). Additionally, learners can be actively involved in discussions as they offer a certain degree of authenticity and meaning (Brown, 2003; Heaton, 1990). 89% of the teachers (n=179) knew that discussions can be utilized in speaking assessments. A very small number the participants, on the other hand, have given an incorrect answer (n=13) or didn't know to assess learners' speaking skills by means of discussions. It is apparent that most of the teachers were familiar with the use of discussion as a speaking task for assessment.

The next item, which the second highest, is "A checklist is a means of scoring oral presentations in in-class assessment". Checklists involve some predetermined criteria on the aspects of speaking to explore whether the learners have demonstrated or not (Luoma; 2004). As a consequence, learners can make use of the results of their performances by virtue of checklists' descriptive and diagnostic nature. However, 32 (16%) teachers in total, containing 20 incorrect and 12 don't know option, were not familiar with the use of checklist. The number of the teachers who selected the true option, the correct option, is 170. In other words, 84% of the teachers were familiar with the utilization of checklists for scoring purposes in classroom assessment, therefore, mostly knew why checklists are used.

The third item with the highest mean score is "When the focus is to assess discourse, role plays are used". Role-plays are intended to mirror reality by means of learners' adopting

another character that is fictitious so that they are prepared for actual conversations outside the classroom context (Brown, 2003; Fulcher, 2003; Heaton, 1990; Luoma, 2004). Given a persona, such as a cashier at the grocery store or a tourist at the airport, etc., learners are expected to cope with the situation with the help of their linguistic abilities. This item has been correctly answered by 166 teachers (82%) while the rest of the teachers (18%) either gave a correct answer or did not know the answer to this item. Thus, when the emphasis is on discourse, using role-play is an issue that the teachers are mostly knowledgeable about.

“In a speaking exam, production and comprehension are assessed together” is the fourth item teachers presented high mean scores. Speaking, in a natural conversation, is an integrated skill requiring different abilities to be performed simultaneously (Harris, 1969). The interdependent nature of speaking and listening necessitates both comprehension and production on the part of the speaker, thus needs to be assessed together (Brown, 2003). A minority of the teachers (n=10) did not know that whether production and comprehension should be included in the assessment criteria while 26 teachers did not believe that they should be assessed together. Conversely, 82% of the teachers, nearly four teachers in five, indicated their familiarity with the nature of speaking considering that speaking and listening in real life relies on each other, thus, needs to be incorporated in the assessment of learners’ speaking performances.

The item next in line is “In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked”. Interview is a typical task that is highly used in speaking exams whether a prompt or question is given to learners or in the form of a discussion (Luoma, 2004). Whether the teachers can change the questions was unknown to 20 teachers (10%) while 19 teachers (9%) didn’t think that they can adopt the questions. Nevertheless, Douglas (2010) asserts that adjustments can be made on the prompt questions so as to match the level of learners’ linguistic abilities. Likewise, 81% of the teachers (n=163) thought that adapting the questions during an interview is not a problem, thus can be considered knowledgeable.

Another item receiving a high mean score by the participants is “Giving learners one task is enough to assess speaking skills”. It is apparent in the literature that more samples should be gathered from the learners’ performances so that a representative judgement can be made upon the data (Harris, 1969; Heaton, 1990; Hughes, 2003). Even though a quite small number of teachers (n=8) did not know the answer, which is the lowest in this part, and had a wrong idea about the number of the tasks to give students to perform (n=49), there was a

high number of teachers (n=145) who gave a correct answer. In total, 72% of the teachers were aware of the inadequacy of giving only one task for speaking assessment. Interestingly, when compared to another item in the assessing writing part, which is about writing two tasks rather than one for writing assessment, the results contradict. While teachers think that at least two tasks should be involved in a speaking assessment, they also think that one task is enough when it comes to writing.

“Interlocutors’ showing interest by verbal and non-verbal signals poses a problem” is the item that has a slightly high mean score than the half of the total score, which is 55%. In a conversation that would take place authentically, speakers may show interest or use non-verbal signals instead of standing with a straight face. Thus, the main concern being a simulation of real-life conversation, in a speaking exam, interlocutors need to be encouraging towards the learners (Heaton, 1990; Hughes, 2003). In total, 111 teachers chose the correct option for this item, which is false as mentioned. 66 teachers (33%) thought that showing interest would be a problem while 25 teachers (12%) had no idea whether teachers should verbally or non-verbally show interest during a speaking exam. In conclusion, teachers were somewhat knowledgeable regarding showing interest in their interviews with the learners.

The first item that has indicated a lower mean score than the average is “Reading aloud is a technique used to assess speaking skills”. According to Brown and Abeywickrama (2010), there five types of speaking tasks, one of which is intensive tasks. In this type of task, speaking is limited to a certain degree in that a specific form of production is expected from the learners. Still, reading aloud, though highly mechanical, is a technique can be used in speaking assessment. The number of those who were correct to think that reading aloud can be utilized as a technique to assess speaking skills (n=96) and those who do not (n=94) are approximately the same whereas only 12 teachers didn’t have an idea whether it can be used or not.

Similar to the item above, “Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills” is another item that low mean score is reported. Repetitions are imitative speaking tasks, in which learners are expected to produce the given word or a short phrase orally (Brown & Abeywickrama, 2010). The emphasis behind this task type is pronunciation, thus can be utilized to diagnose learners’ problems in pronouncing particular vocabulary items (Hughes, 2003). Even though the correct answers (n=95) were above the incorrect ones (n=90), when the number of the don’t know option (n=17) is added, the

teachers who were familiar with using repetition technique as a way of speaking assessment is lower than the half of the participants.

Another item is “The interlocutor gives the score when the learner is in the exam room”. The correct responses to this item (n=89) are lower than the average. While there are 87 incorrect answers (43%), the number of the don’t know option is 26 (13%). Consequently, 44% of the teachers thought that scoring should be done after the learner leaves the exam room. Likewise, it is suggested that the interlocutor focus on the learners’ performance while they are trying to answer a question as well as showing sympathy (Heaton, 1990; Luoma, 2004). The reason behind this low mean score can be that learners’ speaking skills may not be assessed formally in an exam room due to the logistics in terms of lack of time, human, or physical conditions.

“When it comes apparent that the learner cannot reach the criterion level, the task is ended” is an item whose mean score is low. Interlocutors have the chance to alter the questions to match the learners’ level to some degree (Douglas, 2010). If the learners cannot respond to the questions or maintain the discussion based on the criteria, then, there is no meaning in keeping the exam going since it may lead to anxiety and negative washback (Brown, 2003; Harris, 1969). In total 124 (61%) participants, selecting 83 incorrect and 41 don’t know option, didn’t think or know that the task should be concluded if the learners can’t achieve the predetermined criteria. In addition, 39% (n=78) of the teachers responded that the task should be ended, consequently, the teachers’ knowledge on this issue is considered to be low.

“When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem” is the next that the answer to which should be false. The teacher can make a contribution to the discussion without any interruption of the learners’ performances or domination (Fulcher, 2003). The number of the teachers who didn’t have knowledge (n=126) has a dominance over the ones who knew that the interlocutors could express when they don’t understand the learners (n=76). Therefore, it can be reported that teachers did not have the knowledge about whether they should or shouldn’t reflect when they don’t understand the learner.

The item “Using holistic and analytic scales at the same time poses a problem” has also demonstrated a low mean score. Holistic and analytic scales have both advantages and disadvantages. Holistic scales provide an overall impression of the learners’ proficiency while analytic scales offer a detailed information on learners’ strengths and weaknesses

(Brown, 2003; Hughes, 2003). Thus, it is believed that using these scales together may be beneficial in eliminating each other's deficiencies (Luoma, 2004). Nevertheless, 67 teachers (33%) have given an incorrect answer while 61 teachers (30%) did not know whether these scales can be used together or not, which is the highest in assessing speaking part. In other items in the scale related to the use of analytic and holistic scales, the participants exhibited a high level of knowledge. An inference that can be drawn from this is that the teachers do not have enough knowledge about the use of these scales together given that only 37% of the teachers have answered correctly.

One of the lowest mean scores belongs to the "In interactive tasks, more than two learners pose a problem" item. In an authentic context, more than two speakers can be involved in the conversation, yet in educational contexts where learners speaking skills are being assessed, more than two learners may be problematic in terms of the learners' dominance in the interaction (Brown, 2003; Heaton, 1990). This item has 72 correct responses (36%) whereas the rest of the responses are 113 incorrect (56%) and 17 don't know (8%). It is exhibited that more than half of the teachers thought that involving more than two learners in interactive tasks wouldn't be a problem, thus, their knowledge can be considered low.

The last item, which also got the lowest mean score, is "In peer interaction, random matching is avoided". As mentioned in the item above, the interaction between learners needs careful consideration not only in terms of the number of learners involved but also in terms of the characteristics and language level of the learners (Brown, 2003; Fulcher, 2003; Luoma, 2004). Thus, in order to provide an opportunity for learners to interact equally, they should not be assigned to their partners randomly (Heaton, 1990). Teachers' correct answers to this item is 32% (n=65). While 28 teachers (14%) did not know whether peers should be randomly matched or not, remaining 109 teachers (54%) have given an incorrect answer. In conclusion, teachers' knowledge about not matching learners in peers randomly is considered to be insufficient.

In summary, EFL teachers in the K-12 context in Turkey are generally have sufficient knowledge in using role-plays to assess discourse and discussion tasks for speaking, gathering more samples of learners' performances by using more than one task, use of checklist as a scoring scale, speaking constructs in terms of assessing comprehension and production together, and the role of the interlocutor concerning adapting the interview questions and showing interest during the interview. On the other hand, their knowledge level remained below the average on these issues: using repetition and reading aloud as a

speaking task, the role of the interlocutor in terms of when to score, when to end the interview, and whether to express if they don't understand the learner or not, how many learners to involve and how to match them in interactive tasks, and finally, the advantages of using holistic and analytic scales together. Even though teachers have low mean scores in 8 of the items, and in 7 items they have high mean scores, their overall mean score in assessing speaking part is still high. The results reflect those of Genç, Çalışkan, and Yüksel (2020), who also found that the teachers' knowledge in assessing speaking skills are higher than the average.

4.2. Effects of Demographic Features on EFL Teachers' LAK Level

The second research question of the study aimed to figure out whether the participants' demographic characteristics have an impact on their LAK level. In this vein, independent samples t-tests and one-way ANOVA analysis are performed based on the participants' gender, years of experience, the BA program being graduated, educational background, the type of the institution they work at, the educational level of their workplace, the geographical region they work in, taking a separate testing course in pre-service education, and attendance to additional trainings in language testing and assessment. The findings of each variable are demonstrated in the tables below.

Table 8

Independent Samples T-test Results According to Gender and LAK

Gender	N	$\bar{X}$	S	Sd	t	p
Female	154	32.06	5.01	200	-.159	.874
Male	48	32.20	6.66			

The table above shows the findings of the effects of gender on the participants' LAK level. It is seen that the number of the female participants (N=154) are considerably higher than the male participants (N=48). To identify whether there is a difference between the LAK level of female and male participants, independent samples t-test is employed. The results show that the participants' LAK level did not differ according to their gender.

Table 9

One-way ANOVA Results According to the Years of Experience and LAK

Years of experience	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
1-5 years	91	32.50	6.14	Between Groups	138.673	4	34.668	1.176	.323
6-10 years	34	31.52	5.24	Within Groups	5807.347	197	29.479		
11-15 years	23	30.30	4.97	Total	5946.020	201			
16-20 years	27	31.81	3.92						
More than 20 years	27	33.25	4.62						

Another variable in the study is the participants' level of experience. It is revealed in the table that nearly half of the participants consist of novice teachers who have 1-5 years of experience. The participants' level of experience is analyzed in five categories consisting of 1-5 years (N=91), 6-10 years (N=34), 11-15 years (N=23), 16-20 years (N=27), and more than 20 years (N=27). To find out whether there is a significant difference between the groups mean scores, one-way ANOVA is conducted. Based on the table, it is clear that no significant difference is found between the participant groups based on their experience levels. Therefore, it is concluded that the level of experience is not a influential factor in the teachers' LAK level.

Table 10

Independent Samples T-test Results According to BA Program and LAK

BA Program	N	$\bar{X}$	S	Sd	t	p
ELT	181	32.07	5.60	200	-.208	.835
Non-ELT	21	32.33	3.77			

When it comes to the impact of BA program being graduated, the mean scores of the participants who have graduated from an ELT department (N=181) and non-ELT department (N=21) is analyzed with independent samples t-test. The results display that the participants' LAK level did not change according to their department in BA whether ELT or non-ELT since there was no statistically significant difference between the groups.

Table 11

One-way ANOVA Results According to Educational Background and LAK

Educational Background	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
BA degree	161	32.15	5.56	Between Groups	8.152	2	4.076	.137	.872
MA degree	37	32.00	5.16	Within Groups	5937.868	199	29.839		
PhD degree	4	30.75	2.21	Total	5946.020	201			

The fourth variable, which is the educational background, is also investigated to identify the effect of it on the participants' LAK level. Most of the participants in this very study has graduated from a BA program (N=161) while 37 teachers have a degree in an MA program. However, there are only 4 teachers who held a PhD degree in the study. Based on the results of one-way ANOVA, to exhibit whether the participants' educational background has a significant role on their LAK level, it is evident that no significant difference is found. However, since the number of the teachers with a PhD degree is very low, it may have interfered with the results. Thus, it is likely that LAK level of the participants is not influenced by their educational background.

Table 12

Independent Samples T-test Results According to the Type of the Institution and LAK

Type of the Institution	N	$\bar{X}$	S	Sd	t	p
Public/State	187	32.23	5.35	200	1.259	.209
Private	15	30.40	6.29			

Another variable that is scrutinized to reveal the impact of the type of the institution that the participants work whether private or state. Thus, independent samples t-test is utilized in this regard. In the table above, it is shown that nearly all of the participants (N=187) work at a state school while only 15 teachers work at a private institution. The results exhibit that the difference between working at a state and private institution is not significant based on the table.

Table 13

One-way ANOVA Results According to the Educational Level of the Workplace and LAK

Educational Level of the Workplace	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
Primary school	35	31.45	4.17	Between Groups	23.971	2	11.986	.403	.669
Secondary school	103	32.07	5.45	Within Groups	5922.049	199	29.759		
High school	64	32.48	6.03	Total	5946.02	201			

The sixth demographic factor, which is the educational level of the schools that teachers work at, is investigated with the aim of discovering whether the participants' mean scores changed accordingly. In this vein, one-way ANOVA is implemented among the groups of teachers working at a primary school (N=35), secondary school or middle school (N=103), and high school (N=64). The results of the analysis prove that the teachers' LAK level did not differ according to the educational level of their schools.

Table 14

One-way ANOVA Results According to the Geographical Region and LAK

Geographical Region	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
Aegean	16	31.81	5.79	Between Groups	148.581	6	24.763	.833	.546
Black Sea	29	32.31	3.74	Within Groups	5797.439	195	29.730		
Central Anatolia	53	31.26	6.22	Total	5946.020	201			
Eastern Anatolia	22	33.13	6.91						
Marmara	24	32.33	5.52						
Mediterranean	11	29.90	4.13						
Southeastern Anatolia	47	32.91	4.68						

The geographical region that the teachers work in is another focus that the study is concerned with regard to its impact the participants' LAK level. In this vein, the participants are asked about the city they work in and grouped under the seven geographical regions of Turkey, which are Aegean (N=16), Black Sea (N=29), Central Anatolia (N=53), Eastern Anatolia (N=22), Marmara (N= 24), Mediterranean (N=11), and Southeastern Anatolia Region (N=47). The results of one-way ANOVA show that the geographical regions have no statistical impact on the teacher" LAK level.

Table 15

Independent Samples T-test Results According to Taking Separate Testing Course in BA

Testing Course in BA	N	$\bar{X}$	S	Sd	t	p
Yes	113	32.41	4.90	200	.933	.352
No	89	31.69	6.05			

Whether the teachers took a separate testing and assessment course in pre-service education is another focus of this study since taking a course in this subject might be influential in the participants' knowledge level. In the table, it is shown that more than half of the teachers (N=113) took a separate course in BA while the rest of the participants (N=89) did not state taking a course in pre-service education regarding assessment. Thus, to see whether taking a course in BA has an impact on the participants' LAK level is analyzed with independent samples t-test. The result reveal that no statistically significant difference is displayed with respect to the assessment courses in pre-service education.

Table 16

Independent Samples T-test Results According to Attending Additional Trainings in Language Testing and Assessment

Additional Trainings	N	$\bar{X}$	S	Sd	t	p
Yes	105	32.77	5.73	200	1.839	.067
No	97	31.37	5.02			

The last variable that is examined is the effect of taking additional trainings in language assessment for professional development. The table indicates that the participants who continued their professional development by attending trainings (N=105) and who did not (N=97) are close to each other even though the former group is slightly higher than the latter. In this respect, to demonstrate the effect of professional development on LAK level of the teachers is scrutinized by comparing the mean scores of the participant groups by independent samples t-test. Based on the table, it is shown that attending trainings in terms of language assessment did not have a significantly positive or negative impact on the teachers' LAK level.

All in all, the results of the analyses indicate that the impact of the abovementioned demographic features of the teachers is not statistically significant. Therefore, it can be concluded that the contextual factors such as the institution type, the educational level of the workplace, the geographical region they work in, and experiential factors such as experience

level the department of graduation, the educational background in addition to pre- and in-service education is not influential in EFL teachers' LAK level in the K-12 context of Turkey.

4.3. EFL Teachers' Perceived Self-competency and Actual LAK Level

The third research question of the study aimed to reveal whether the teachers' LAK level changed according to their perceived self-competency. The teachers are asked to select their level of competency in the scale for each language skills consisting of “very competent”, “competent”, “not very competent”, and “not competent” options. Therefore, teachers' perceptions of their knowledge in assessing each skill have been investigated with one-way ANOVA analysis. The results are presented in the tables below.

Table 17

One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Reading

Assessing Reading	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
Very competent	110	9.74	1.88	Between Groups	12.639	2	6.320	1.615	.202
Competent	84	9.28	2.03	Within Groups	778.891	199	3.914		
Not very competent	8	10.12	2.64	Total	791.530	201			

The table shows the level of competency perceived by the teachers about their knowledge level in assessing reading. It is seen that more than half of the teachers (N=110) believed that they are very competent while a small number of the teachers (N=8) didn't think that they are very competent in reading assessment. The teachers who stated that they are competent (N=84) in their knowledge is also high in number. However, it is seen in the table that the “not competent” option is not demonstrated in the table since there was no participant that selected that option. Therefore, it is possible to infer that teachers are competent in their reading assessment knowledge even though the teachers who are not very competent in their knowledge in assessing reading performed higher than the others. When the participants' mean scores in assessing reading ($\bar{X}$ =9.5693) is considered, which is significantly high, it is possible to say that their perceptions and actual LAK level are in line with each other due to the fact that assessing reading has received the highest mean score among all language skills. Besides, there is no significant difference between the competency levels of the teachers

regarding their mean scores. Thus, it can be suggested that the teachers are generally aware of their knowledge level in assessing reading.

Table 18

One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Listening

Assessing Listening	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
Very competent	55	7.00	2.49	Between Groups	5.748	3	1.914	.416	.742
Competent	112	6.75	1.79	Within Groups	912.529	198	4.609		
Not very competent	34	6.88	2.57	Total	918.277	201			
Not competent	1	5.00							

The table shows the level of competency perceived by the teachers about their knowledge level in assessing listening. It is seen that most of the teachers felt competent (N=112) and very competent (N=55) regarding their knowledge in assessing listening, who also have the highest mean scores compared to the teachers who selected the options not very competent (N=34) and not competent (N=1). Nevertheless, the mean scores that the participants have in assessing listening part in the scale ($\bar{X}$ =6.8317) was significantly lower than the half of the total score (7.5) in addition to being the lowest among other skills. Additionally, the difference between the participants' mean scores based on their level of competency is not statistically significant. Therefore, it can be inferred that the participants' perceptions about their knowledge may not coincide with their actual knowledge level in assessing listening.

Table 19

One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Writing

Assessing Writing	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
Very competent	70	7.40	2.10	Between Groups	21.211	3	7.070	1.528	.208
Competent	108	6.99	2.15	Within Groups	916.076	198	4.627		
Not very competent	21	6.28	2.32	Total	937.287	201			
Not competent	3	7.00	1.00						

The table shows the level of competency perceived by the teachers about their knowledge level in assessing writing. It is demonstrated that the teachers have high level of self-competency since 88% of the teachers stated that they are either competent (N=108) or very competent (N=70). Contrastingly, the mean scores of the participants in assessing writing

($\bar{X}=7,0594$) is found to be significantly low. Additionally, the difference between the participants' mean scores based on their perceived self-competency is not significant. Even though the participants who are very competent received the highest mean score in assessing writing, it might be suggested that their perceived self-competency is not reflected on their actual LAK in writing assessment.

Table 20

One-way ANOVA Results of Perceived Self-competency and LAK in Assessing Speaking

Assessing Speaking	N	$\bar{X}$	S	Source of Variance	Sum of Squares	df	Mean Square	F	p
Very competent	49	9.40	2.32	Between Groups	73.453	3	24.484	5.075	0.002
Competent	122	8.54	2.09	Within Groups	955.166	198	4.824		
Not very competent	29	7.58	2.41	Total	1028.618	201			
Not competent	2	11.00	0						

Finally, the table shows the level of competency perceived by the teachers about their knowledge level in assessing speaking. The teachers' competency level is found to be considerably high as the number of the participants who selected the option very competent (N=49) and competent (N=122) constitutes the 84% of the total participants. Likewise, when their mean scores in assessing speaking in the scale ($\bar{X}=8.6386$) is taken into account, it could be the case that their perceptions about their knowledge level is in accordance with their actual LAK in assessing speaking due to the fact that they have significantly high knowledge level. However, it is seen that the participants who are not competent have the highest mean score compared to other groups indicating a significant difference between the groups in favor of the low competency group according to the one-way ANOVA results.

4.4. Classroom Practices of EFL Teachers

The fourth research question of the study aims to find out about the classroom practices of teachers in the K-12 context regarding language assessment. In line with this aim, semi-structured interviews have been conducted with 17 EFL teachers working at primary, secondary, and high schools throughout the country regarding their classroom practices in terms of language assessment. The results are presented based on the findings of the content analysis derived from the qualitative data by means of assigning codes and detecting the emerging categories and themes. Based on the data obtained from 16 questions, two of which

are specific to teachers of young learners, 6 themes have emerged in total: characteristics of language assessment, assessment of language skills, assessment results, planning assessment, professional development, and problems. The themes are reported in the tables according to the frequency of the codes and discussed in the light of the literature.

4.4.1. Characteristics of Language Assessment

The first theme, the characteristics of language assessment, has emerged from the teachers' responses to how they define assessment besides why and when they assess their learners in addition to the characteristics of assessing young learners. Therefore, the codes that are assigned to teachers' responses are grouped under the categories of purpose, time and young learners and presented in the table below.

Table 21

Teachers' Views on the Characteristics of Language Assessment

Theme	Category	Code	Frequency
Characteristics of Language Assessment	Purpose	To determine the degree of learning	12
		To diagnose weaknesses	8
		To plan/adjust classroom practices	8
		To monitor learning process	4
		To give feedback	2
		To determine if objectives are met	2
		To enhance motivation	1
	Time	During instruction	17
		After instruction	9
		Before instruction	2
	Young Learners	Motivation	6
		Focus on strengths	6
		Positive attitudes	5
		Attention span	5
		Visualization	4

The table indicates that teachers use assessment for a variety of purposes. It is seen that most of the teachers used assessment in order to determine the level of learning by means of seeing where learners at regarding learning outcomes, what is learned and what is not, to what extent achievement took place, and how much is received by the learners. In addition, teachers also used assessment to diagnose learners' weaknesses and use the assessment data to strengthen these weaknesses in addition to monitoring the learning process in terms of the

progress that learners show. Teachers' views regarding their purposes of assessment are as follows:

"Of course, we use it to assess whether the kids have acquired the outcomes or not, to what extent it is acquired, and to assess their knowledge and skills. To determine the degree of progress. I use assessment for these purposes." (Canan)

"We use assessment to identify learning gaps and address them" (Giray)

"I use assessment to determine whether they understand the given subject well enough and to determine the reasons why not" (Burcu)

"I assess certain qualities of the education we have given, that is, to see how much we have achieved the outcomes." (Nazım)

Additionally, teachers' responses demonstrated that half of the teachers utilized assessment to evaluate their teaching methods and techniques so that they can plan and adjust their classroom practices. Teachers reported their opinions as:

"I can say that I use assessment to detect learning gaps, if there are any, and to adjust or change and plan my own teaching methods and techniques." (Sevgi)

"I use it to determine the level of the learners and to decide what to do next. I mean, if assessment is done in the process, I determine the next step based on the missing aspects in the result. What should I emphasize more? Where should I give better instruction or slower?" (Işın)

Among the purposes of assessment that teachers have stated, only a few teachers used assessment in order to give feedback, to determine if the objectives are met. In this regard, Teacher Pırl expresses that *"Actually, to see if the course outcomes are achieved"* while Teacher Merve focuses on the aspect of feedback by stating that *"I assess learners to give feedback to them"*. Another point by Teacher Oya is as follows:

"To determine the efficiency of the instruction, whether there are any learning gaps or not, whether the learning objectives are met or not. If there are learning gaps, I use assessment to provide feedback for the students to overcome them." (Oya)

Based on the excerpts given above, it is seen that the teachers use assessment for both formative and summative purposes. Even though the most common aim stated by the teachers is to determine the degree of learning, which is has a summative role in that teachers aim to demonstrate the learners' current state regarding the learning outcomes, teachers also

used the assessment results to collect evidence about the learners' weaknesses in addition to their own practices for the sake of bridging the gap between learning and the desired outcomes by means of modifying their classroom practices. Besides, teachers see assessment as feedback to themselves, therefore, formative purposes are also adopted in the teachers' use of assessment in their classroom practices. Some views of the teachers regarding this aspect are:

"I use assessment to see learners' progress and to how much they have understood or not. It's sort of a feedback from the students to me." (Kaan)

"I guess I use it more to see my own weaknesses. Because if 80% of the students can't do something, I believe that I need to focus on that aspect more. I am missing something here; I actually get more feedback to myself." (Hale)

"For me, the purpose of assessment is definitely to see the efficiency of the instruction. For this purpose, my aim is to figure out how effective what I taught in the class was, how much of it is conveyed. I use it to determine the effectiveness of my instruction and to make necessary adjustments based on them." (Elif)

One of the responses to this question has revealed that teachers' purposes may alter depending on the age groups they are working with. That is, the developmental characteristics of the learners may have an impact on the teachers' assessment choices and why they use assessment. Teacher Rana, who is working at a primary school with young learners, expresses why she uses assessment as:

"I use assessment as a motivation tool right now because I'm working with young learners. I don't use exams that would make them anxious or worried about their grades indeed. I generally assess them to detect their weaknesses or where they don't understand" (Rana)

To get a deeper understanding of how teachers define the characteristics of language assessment, the time of the assessment should also be revealed. When the teachers are asked when they use assessment in their classrooms, their responses gathered around three codes, which are before, during, and after instruction. The frequency of the responses reveals that among 17 teachers, all the teachers use assessment during their instruction while nine of them stated that assessment is utilized at the end of a certain period as well, that is after instruction. Only two of the teachers have utilized assessment before the instruction. Teachers who utilized assessment during their instruction in the classroom believed that assessment is a part of the learning process and should not be limited to written exams or

giving grades to the learners. In this respect, Teacher Nazım puts forth that *“I use assessment constantly, not only just to give grades. I believe that formative assessment is more important to see how much is given to learners.”*. Besides, another point that is emphasized by the teachers is that assessment needs to be embedded into the classroom practices. On this issue, Teacher Giray stressed the significance of using assessment constantly as *“We use assessment all the time. Basically, that’s how it should be. After we teach, we monitor how much is learned. Otherwise, the learning will not be reinforced.”*. A similar opinion by Teacher Kaan is as follows:

“Actually, I assess learners every day in each lesson. We shouldn’t consider assessment as just a written exam. The kids’ raising their hands, their willingness to participate, their answers to the question, they are all a part of my assessment.” (Kaan)

In addition to the teachers’ rationale behind utilizing assessment embedded into their instruction, some of them adopted quizzes at the end of the coursebook units in addition to classroom activities to collect evidence of the learners’ performances in the learning process. Teachers’ opinions regarding when they use assessment are as follows:

“We generally use assessment at the end of the units. I use it more at the end of the course units after covering a new objective to determine the level of learning.” (Canan)

“I can say that I always use it with my observations about the learners’ speaking performances since we have four skills. Besides, I also use quizzes to see whether the content is clear or not, and if not, I employ different techniques and review the content. We also have written exams twice a year.” (Sevgi)

Even though they used quizzes, which can be considered as a summative assessment tool, teachers used that information to adjust their classroom practices. Based on these excerpts, it is seen that teachers employ different assessment techniques to gather information about the learners in the learning process formatively. On the other hand, similar to Teacher Sevgi’s views, more than half of the teachers have mentioned that they use assessment after the instruction with written exams as a form of summative assessment. In the K-12 education context, written exams are obligatory beginning with 4th grades. Therefore, teachers are expected to conduct midterm and final exams formally depending on the teaching hours they have to the particular grade. This obligation is revealed in the teachers’ responses as:

“I conduct written exams at certain times based on the decisions we made in the meetings with my colleagues in addition to the quizzes during the semester. Apart from that students’

participation, listening, speaking tasks and my observations are all a part of my assessment.” (Oya)

“Well, we have to use two written exams as it is mandatory. Besides, I use assessment when I need to identify the gaps in the learning with simple quizzes to see if the vocabulary items are clearly understood, can they recall them or to use simple activities in the classroom” (Rana)

Finally, two of the teachers expressed that they use assessment before the learning process for different purposes. Teacher Işın focused on determining the readiness of the learners before the process, and she also incorporated assessment during and after the instruction whereas Teacher Pırıl emphasized the coursebook selection and placement purposes of assessment for adopting assessment before instruction.

“Firstly, I use it at the beginning of the process to see the knowledge that the learners have. Then, I use it in the process to see their progress at the end of each unit. Finally, I use it at the end of the semester.” (Işın)

“We use it at the beginning of the semester so that we can select the coursebook. Besides, this may not be a proper practice, but we have classes based on the learners’ proficiency. So, it is to place the students in these classes. During the semester we have progress tests for speaking and writing skills. We also use achievement tests for grammar, vocabulary, and reading at the end of the semester.” (Pırıl)

All in all, when teachers’ expression about the purpose and the time of their assessment practices, the characteristics of the language assessment in the K-12 context have a balance between formative and summative approaches. Even though teachers have to implement written exams as a part of their institutional policy, they also made use of assessment techniques within the perspective of formative assessment by means of using assessment as a part of learning process and utilizing the information obtained from the assessment practices to determine the learning gaps and weaknesses of the learners to address those gaps by modifying their classroom instruction.

Finally, the last category, young learner assessment, has emerged from the answers given to the questions asked to the 13 teachers who teach English to learners between the ages 6 to 12. It is seen that teachers of young learners stated that young learners are highly eager and motivated to learn English. Thus, teachers mostly focused on the strengths of their students rather than focusing on their mistakes in order that the learners can acquire positive attitudes

towards learning English. Besides, the attention span of the learners is also taken into consideration by the teachers in that teachers preferred to utilize more engaging and interesting assessment practices so that the learners can have fun. Finally, visualization was also a part of their assessment practices that teachers considered while assessing young learners. Some excerpts regarding the characteristics of assessing young learners are as follows:

“First of all, they are excited and enthusiastic. Actually, they have a higher motivation, which is the most important thing for me because my former students at a high school had no motivation at all. However, with young learners, it would be very beneficial for them towards learning English if we can take advantage of their motivation in this regard.” (Rana)

“They are sometimes struggling since they are 2nd grades and just started learning English. I’m more flexible with them during the assessment. I don’t want to force them to do the task as they have just become literate in their own language.” (Oya)

“With young learners, I use songs, games, and more visuals and concrete items because the smart boards don’t appeal to them sometimes, but the touching the objects and materials attracts them more. That’s what I observed. I generally try to make them enjoy the activity and not get bored. Besides, it caught my attention that young learners are far more likely to acquire the language compared to the older learners.” (Damla)

“When they make a mistake, I don’t emphasize them. I mean, simple mistakes are not a big deal with this age group. I believe the sooner they learn English and acquire the necessary skills, the better.” (Kaan)

“They forget faster, and their focus shifts very quickly. I mean, if we have 40 minutes in a lesson, you can’t keep them focused for 40 minutes. Thus, I try to add use more visuals and games so that I can keep their attention.” (İşin)

“We keep portfolios with older students, but the portfolios for this particular age group involves colourings, drawing, and some crafts activities. I tried to add more stuff according to their interests since I want them to be motivated and interested. Besides, I really want them to have fun while doing the task, not just for the sake of doing it.” (Burcu)

As a conclusion, it is clear that teachers of young learners take their learners’ developmental characteristics and linguistic capabilities into consideration in their assessment practices. Besides, teachers were aware of the fact that young learners have high motivation, and they

utilized their motivation to promote positive attitudes towards learning English. Taking the short attention span of the young learners into account, teachers tried to keep their students interested by appealing to different senses in that they incorporated visual aids in their assessment practices. Additionally, teachers tried to increase the feeling of success by emphasizing learners' strengths without focusing too much on the learners' mistakes, creating a supporting environment that also involves the fun element.

4.4.2. Assessment of Language Skills

The assessment of language skills theme has emerged from the answers given by the teachers about the skills they assess, the methods they employ, and the tools or techniques they adopt in the classroom. Thus, the frequency of the codes under these three categories, which are language skills, methods, and tools/techniques, are presented in the table below.

Table 22

Teachers' Classroom Practices Regarding the Assessment of Language Skills

Theme	Category	Code	Frequency
Assessment of Language Skills	Language Skills	Reading	12
		Listening	12
		Speaking	11
		Grammar	9
		Vocabulary	9
		Writing	8
		Pronunciation	2
	Methods	Traditional	17
		Alternative	14
	Tools/Techniques	Pen&paper test	14
		Observation	13
		Quiz	9
		Portfolio	9
		Oral performance	9
		Checklist/Rubric	5
		Project	4
		Games	4
		Self-assessment checklist	3
		Song	3
		Digital tools	3
		Journal	2
		Competition	1
		Dictation	1

As can be seen by the frequency of the listening skills that teachers assess, reading and listening are the most frequently assessed skills followed by speaking. In addition, grammar and vocabulary, which are considered as subskills, are also highly included in the assessment practices of the teachers. However, when compared to other skills, writing has not received as much attention as reading, listening, and speaking whereas pronunciation is a skill that only two teachers have emphasized. Nevertheless, when the demographic information of the participant teachers in the interviews is taken into consideration, a distinction in the assessment practices of the teachers have been observed. Teachers who work at a middle school generally focused on reading, grammar, vocabulary, and to some extent, writing skills while listening and speaking skills are ignored for certain reasons. Some of the middle school teachers' statements on this issue are as follows:

"Reading, vocabulary, and grammar are mostly assessed. That is the standard of written exams. There is not much writing since it's a productive skill and learner profile of the school is not high especially in the stressful atmosphere of the exam. I assess listening and speaking skills during the course, yet it is not emphasized as much as the others. Not with formal exams at least since it's not included in the educational policy." (Ferhat)

"I assess listening and pronunciation informally and other skills such as grammar and reading are assessed formally with written exams. But listening and speaking skills fall behind a bit as a result of both lack of teaching hours and not being able to motivate learners in this respect." (Ahmet)

Teachers explain the reasons why speaking and listening skills are ignored are due to lack of teaching hours, learners' motivation in addition to the assessment of listening and speaking skills in the curriculum while Teacher Burcu touches upon the crowded class sizes by stating that *"Unfortunately, since our school is very crowded, we cannot spend enough time on listening and speaking skills. Therefore, we assess mostly vocabulary, reading, and grammar."* Another reason that one of the teachers reported is that the learners become anxious when it comes to speaking exams.

"I use reading passages for reading skills in addition to grammar and vocabulary. For writing, students keep a journal. Last year, I've tried to conduct a speaking exam especially because the class sizes were small, but they were too stressed about it. I had to postpone it because they were not ready. Thus, now my assessment is mostly based on reading, vocabulary, grammar, and sometimes writing." (Elif)

Additionally, even though some teachers involve all of the skills in their assessment practices, it is revealed that for reading, grammar, and vocabulary skills, formal assessment procedures are employed mostly with written exams and quizzes. On the other hand, for speaking, and listening skills, teachers mostly adopted informal assessment by means of their observations of the learners' performances in the classroom activities.

"In written exams, reading, writing, vocabulary, and grammar, but speaking, listening, and pronunciation takes place during the lesson." (Kaan)

"The expectation of the system is clear, that is, reading, grammar, and vocabulary. Because the standardized tests involve multiple choice questions, I employ these types of practices more. I use pen and paper tests for grammar, reading, and vocabulary. However, I also assess listening, writing, speaking skills. I give songs as a homework for listening. Sometimes, I ask them to make a presentation on a topic they choose for a couple minutes. For writing, we have a pen pal activity right now." (Oya)

"All of them. Sure, vocabulary, grammar, and reading are assessed with midterms and final exams in addition to our monthly quizzes. We also employ oral presentation for speaking, twice a month. Similarly, we give writing tasks as a homework, but very limited and short." (Piril)

When it comes the assessment of young learners, teachers mostly assessed listening and speaking skills of the learners with different methods and techniques, and their responses indicated that reading, writing, and grammar skills are very limited or completely ignored with this age group as a result their developmental characteristics. Some excerpt taken from the teachers' responses regarding assessing language skills of young learners are:

"Since I work at a primary school, students' reading skills develop in the later stages. I generally focus of pronunciation as the spelling and the pronunciation of the words are different. Also, vocabulary and speaking are at the center in terms of establishing a dialogue. Grammar is covered at the later stages, similar to writing. With 2nd grades, we don't do writing assessment, but use simple tasks like tracing technique." (Canan)

"It is generally listening and speaking. To assess these skills, I use dictation with simple chunks since their writing is not developed enough. The technique that I use the most is act out, I always use act out for speaking." (Rana)

“I don’t assess reading since I work at the primary level and writing as well. But I generally assess their listening with the activities in the coursebook. Also, speaking to a certain degree. No grammar. I use assessment for pronunciation a lot with questions and answers” (Damla)

All the teachers working at a high school reported that they assess four skills. Teacher Nazım expresses the importance of teaching English language and its’ constructs by stating that *“Since I teach English as a subject, I generally assess four skills.”*. In addition, Teacher Merve refers to the fact that assessing four skills in high schools is an institutional policy of the Ministry of National Education. The excerpt is given below:

“Of course, we focus on four skills as we already have to formally assess four main skills as suggested in the high school curriculum. We assess all of them, yet I also focus on vocabulary. So, four skills plus vocabulary.” (Merve)

The excerpts above reveal that the emphasis given to the language skills can be shaped by the learner’s proficiency level besides motivation, the educational policies in the curriculum, and the educational level of the schools. Since the teachers mostly adopted formal procedures for reading, grammar, vocabulary, and writing to a certain degree, their preferences on the assessment methods and techniques in addition to the assessment tools that they incorporate in their practices are also influenced by their approach. Even though all the teachers adopted traditional methods, based on their responses only a small number of the teachers preferred traditional methods and tools as the medium of their assessment. Teacher Ferhat and Damla’s views in this respect is as follows:

“Standard written exams as I have said, as an assessment method. There is not much other than that.” (Ferhat)

“I use traditional methods such as questions and answers, worksheets as homework. I don’t use modern methods that much.” (Damla)

Among the traditional assessment methods, pen and paper tests such as written exams and quizzes were the mostly preferred tools by the teachers. Similar to the emphasis given to assessing language skills, teachers mostly preferred traditional methods for reading, grammar, and vocabulary skills. Teacher Pırıl summarizes this as: *“Sure, vocabulary, grammar, and reading are assessed with midterms and final exams in addition to our monthly quizzes.”*. Some of the teachers express the reason behind adopting traditional methods as a mandatory process either as an institutional policy or the result of the exam

oriented educational system. Teachers' views about the obligatory aspect of their traditional assessment practices are as follow:

"Though not excessively, I also use tests because I believe that I have to include them in the assessment since the students have to take a standardized exam at some point in the future."
(Nazım)

"We certainly use traditional. Because we have to conduct two written exams in a semester. However, even in those written exams I try to involve visuals or dialogues just like we did in the classroom." (Rana)

When the teachers' use of different assessment methods is investigated, it is observed that they also made use of alternative methods remarkably besides traditional methods. Classroom observation was notably used by the teachers followed by portfolio assessment, performance assessment while self-assessment, games, songs, digital tools, journals, competitions, and dictation were not utilized as often compared to the others. Therefore, it can be inferred teachers adopted a balanced approach to using different assessment methods so that each method can compensate for the disadvantages of each other. In this vein, Teacher Ahmet asserts that:

"I don't think that a single assessment method is advantageous by itself, so I think of it as an eclectic method. I also implement alternative assessment methods for my students by means of performance assessment and projects. All of them have their advantages and disadvantages, yet the tests are still the most frequently used tool under these circumstances." (Ahmet)

Considering the tools and techniques that teachers adopt among alternative methods, it is clear that observation is the most common technique that teachers prefer for speaking skills followed by the oral performances, including mostly presentations besides role-plays and drama. Therefore, it can be asserted that assessment is an integrated part of the teachers' classroom practices, that is teachers collect evidence of the learners' performances continuously in the process. Teacher Kaan addresses the role of observation in his assessment practices as *"I, myself, assess to what extent learners can or can't perform during the listening and speaking activities in the lesson."* Additionally, a high school teacher, Sevgi refers to the reliability aspect of observation as *"I assess speaking by making observations on the students' participation to the tasks. So, I use observation a lot. Yet, I'm not sure how reliable they are, but it is not a part of the formal assessment."* Even though

observation can be a useful technique for assessing learners' speaking performances, relying on mostly informal and non-systematic observations may not be sufficient for an objective interpretation of the results. Accordingly, only three teachers made use of checklists or rubrics for speaking. Teacher Nazım expresses his use of rubrics as *"Usually, we have a rubric for speaking including fluency, accuracy, task completion, etc. I use these criteria."* On the other hand, many teachers relied on informal observations when assessing speaking. Teacher Oya explains why she doesn't use a scoring scale as:

"I don't use any sort of a list or a rubric while observing. I can see the outcome that I expect while listening to them. Besides, I don't want to make them feel like being assessed and realize that it's an assessment." (Oya)

In addition, observation is also mostly preferred by teachers who work at primary schools with young learners for both speaking and listening. Since writing and reading skills are very limited in the context of young learners considering the curriculum for primary schools, teachers made use of classroom observations in addition to oral performances, games, songs, and projects. Teachers of young learners express the assessment tools and techniques used in the primary school context as follows:

"Assessment is a bit different in primary schools since we don't have written exams with 2nd and 3rd grades. Thus, we are, as primary school teachers, constantly assessing learners with observations, projects, and feedback. For speaking, we don't have any other choice than classroom observations, and for listening, we have smartboards and TPR is useful. Can they act out what they hear?" (Hale)

"For 2nd and 3rd grades, I try to use games or simple tasks while assessing. Mostly based on my observations, but I don't have any observation form or anything. I focus on the listening skills for young learners with this digital game, through which I can see clearly if the students have learned or not." (Leyla)

"For example, we have dialogue completion or making dialogues, and songs. I employ songs to discover the problematic areas in their pronunciation. I also use dictation for listening with simple chunks" (Rana)

"I assign projects on some topics. They design a poster about the topic and present it if they can so that I can assess their speaking and also to enhance their motivation. Also, I use listening tasks in the coursebook" (Damla)

Based on the excerpts above, it is seen that primary school teachers concentrate on listening skills with various assessment tools such as song, games, digital games, and dictation as well as speaking skills. On the contrary to primary school teachers, when asked about their assessment methods, tools, and techniques they adopt for assessing language skills, many of the teachers did not mention which tools or techniques they use for listening, though they stated that they assess listening, except for simple tasks in the classroom in the pen and paper formats, which involve mostly gap filling, multiple choice, or true/false questions. A response provided by Teacher Sevgi in that regard is *“Well, I ask them to listen to the text and do the exercises with true/false items or fill in the blanks, and multiple-choice questions.”*. Therefore, similar to the findings that were mentioned above about the emphasis given to each language skill, it is revealed that listening can be ignored in the assessment practices of the teachers.

As a form of alternative assessment, portfolios were adopted by more than half of the teachers. Teachers believed that portfolios can be a powerful tool for both learners and themselves to see the progress the learners make. Besides, portfolios were mostly preferred for the collection of the learners works in addition to assessing writing skills though some teachers employed digital portfolios for speaking. Teachers’ opinions on the use of portfolios are as follows:

“Since language is something that develops in the process, I believe that portfolios can be a proper tool for our subject and objectives for that matter. So, I can ask for drafts from the students periodically and give feedback.” (Elif)

“Portfolios can be quite effective. For example, they drew themselves and wrote simple sentences about themselves with the decorations they make on the paper, and they put their work in their portfolios.” (Canan)

“Besides, students have a portfolio that they put their work in, and also, I also have a file that I use for my students’ videos that they send me. At the end of the year, when I show them the videos, students love it, and they can see their progress.” (Rana)

Self-assessment was one of the alternative methods that five of the teachers mentioned that they involve learners in the process of assessment for learners to take responsibility of their own learning. In addition, teachers believed that when using self-assessment checklists, the learners are objective and honest while assessing their own progress. Besides self-assessment checklists, teachers used self-assessment for grading purposes by means of

giving the students their own exam sheet so that they can grade themselves with the given answer key. In this vein, teachers' responses are as follows:

"I mostly use self-assessment for 2nd and 3rd grades. When I tell them to assess themselves, they put a happy or sad face on the checklist, and they are actually being honest with it." (Leyla)

"I use self-assessment because there is a checklist at the end of each unit in the coursebooks. There are statements such as I know how to..., and I can do... Besides, they can objectively assess their own progress as well as weaknesses and strengths." (Canan)

"If giving the quizzes that we regularly have to grade themselves so that they can see their own mistakes can be considered self-assessment, I do it frequently." (Merve)

On the other hand, some teachers expressed that they don't employ self-assessment procedures since they don't think that learners are autonomous or objective enough to assess themselves. Teacher Ferhat explains the problem as *"I don't use self-assessment since they are not autonomous yet and mostly dependent on the teacher."* while Teacher Pırıl touches upon the subjectivity of this type of assessment as *"I can't use self-assessment because even if they don't understand the subject, they always give themselves 5 points or tick the 'I know very well' section."*

Peer assessment has not received much attention except for grading purposes as Teacher Merve notes *"We have quizzes on a regular basis and to score them, I mix them up and hand them out to different students so that they can score their friends' paper."* The reason why most teachers didn't prefer peer-assessment is similar to the reasons why they didn't utilize self-assessment, that is the subjectivity. Teacher Leyla explains her problems while using peer-assessment as *"We have a book that involves peer-assessment forms. I tried to use them a couple of times, but they mostly ignored their friends' mistake or corrected them. So, it didn't work as much as self-assessment."* Thus, it can be concluded that even though self-assessment is not preferred by many teachers, the ones who adopted it as a part of their classroom practices found it useful while peer-assessment is one of the least used methods among alternative methods.

4.4.3. Assessment Results

The table below illustrates the teachers' utilization of assessment results in the K-12 context in term of their judgements based on the results and the type of feedback they adopt in their classroom practices.

Table 23

Teachers' Use of Assessment Results

Theme	Category	Code	Frequency
Assessment Results	Judgement	Planning/Adjusting classroom practices	9
		Giving feedback	8
		Grading	4
		Evaluation	2
		Making realistic judgements	1
	Feedback	Quick	17
		Oral	8
		Written	5
		Positive	4
		Corrective	4
		Teacher student conferences	4
		Delayed	3

The table above demonstrates that teachers make use of assessment results in order to plan or adjust their classroom practices and to give feedback to their learners. Based on the assessment data that they obtained with assessment practices, teachers stated that they make interpretations of the results in terms of detecting the learners' strengths and weaknesses and make proper adjustments on how to shape their classroom practices towards the problematic areas in their learning in addition to giving feedback when needed. Besides, the teachers did not only consider assessment results as a form of feedback for the students, but also to themselves. Teachers' expressions on how they interpret the assessment results are as follows:

"I can see what I do wrong. For example, if I have three classes and if something went wrong there, I could fix it with the other classes. I usually change my methods and techniques based on the results." (Damla)

"I use the results to see students' strengths and give feedback to the students if they have a problem with the course content. Basically, we determine the weaknesses and do extra activities to improve them." (Giray)

“Actually, I mostly learn from the assessment results. Were the children able to comprehend with the method I used or not? What should I change if I teach the 2nd grades again next year? So, even though I can’t give feedback to my students since we only have two teaching hours a week, I believe I utilize it very well to improve my own practices.” (Leyla)

It is seen in the teachers’ responses that only four teachers stated that they use assessment results for grading purposes while a minority of them utilized the assessment results to make a final judgement of the learners’ performances, in other words, to evaluate. In this regard, Teacher Ferhat emphasizes that he makes judgements based on the results for grading purposes as *“We have written exams and the scores of them are significant since it’s going to be in their report cards.”*. However, due to the fact that teachers had previously stated the obligation to formally assess learners with written exams based upon the institutional policies, it can be inferred teachers consider assessment as a medium of improving learning process rather than just giving grades given that the grading and evaluation purposes are not emphasized by the teachers. Similarly, Teacher Oya stresses that she makes more realistic judgements about the learners’ performances when she uses classroom activities for assessment besides conventional methods.

“Yes, we have quizzes and written exams, but in that case the students may be in a different condition that the results may not reveal their actual performance. But, with the activities we do in the classroom or performance tasks, I can observe the real outcomes more vividly, and naturally. I can make more accurate judgements.” (Oya)

It is shown that teachers utilize assessment results mostly for planning their syllabi and giving feedback to their learners. Therefore, the feedback aspect of their classroom practices is scrutinized in terms of how and when they give feedback. Teachers’ responses demonstrated that all the teachers mostly give feedback to learners’ written exams or quizzes as quickly as possible so that the learners can remember what they did in the exams and see their mistakes. In this vein, a couple of responses from the teachers are as follow:

“If we have a quiz, for example, I should grade the papers within that week not after three or four week and give them to the students so that they can remember and correct their mistakes.” (Işın)

“Generally, within a week after the written exam sometimes within the day if I can. I tell them what they’ve done wrong especially by crossing out the mistakes and writing the correct one.” (Leyla)

“Especially after the written exams, we check them together in the following lesson we have.” (Elif)

Oral feedback was also adopted by nearly half of the teachers mostly for their classroom activities. On the other hand, a minority of the teachers expressed that they used delayed feedback. Some excerpts taken from the interviews are as follows:

“Apart from the exams, for the activities we do, I give feedback, in the classroom, if possible, like ‘you did this like this, you can correct it this way.’” (Leyla)

“If we are doing speaking activities in the classroom, I don’t interrupt their speech immediately in case they get nervous or discouraged. When they complete the task, then I try to utter the correct version to make them notice.” (Oya)

Another type of feedback that is not preferred much by the teachers is written and corrective feedback. Even though they are not a popular form of feedback, teachers mostly used written and corrective feedback for the learners’ portfolios and other writing tasks. Teacher Elif explains her use of written feedback as *“If they are making a product, a written product, I ask them to bring them on certain periods and write my feedback on them with some error codes or corrections.”*. Moreover, teacher student conferences were adopted by a minority of the teachers for giving feedback to learners individually. Teachers also added that they adopt teacher and student conference only when there is an extreme case where learners performed poorly during the assessment. In this regard, Teacher Işın explains how she uses teacher student conferences as *“Generally, if a student made a lot of mistakes, I talk to them individually about the reasons and what they should do about it, and if they need to practice more.”*. The reason why individual feedback is preferred only in extreme cases is that the teachers believe that it is challenging and not practical. Teacher Giray simply expresses this issue by stating that *“Frankly, it is difficult to give feedback to 30 students one by one.”*. Another reason that is uttered by Teacher Ferhat is the classroom management problem and time related concerns while giving individual feedback.

“We don’t have the time to give feedback on writing tasks one by one. If try to do that, the other students won’t be sitting silently and wait, which is a problem in that respect.”. (Ferhat)

Furthermore, positive feedback, even though not used by many teachers, is mostly preferred by primary school teachers to encourage learners and enhance motivation. Some of the primary school teachers’ opinions are as follows:

“If they are performing a task that I give, I walk around the class asking if there is a problem and explain, but if they do something right, I say ‘good for you’ or ‘well-done’ at least. It counts as feedback after all.” (Damla)

“I try to show them their strengths and help them with their weaknesses without them noticing. They enjoy it when I give feedback with phrases such as ‘well-done’ or ‘perfect’.” (Hale)

As it is demonstrated with the excerpts taken from the teachers’ responses, it is clear that all of the teachers tried to give feedback as quickly as possible on the learners’ performances on the written exams. However, given that the teachers have difficulty with giving individual feedback and mostly used oral feedback only to the whole class in the classroom activities and in extreme cases of the learners’ poor performances, it is possible to say that teachers’ use of feedback is inadequate.

4.4.4. Planning Assessment

The table below presents the decision-making process of the teachers for planning their assessment practices in terms of the sources that teachers use, the factors that they consider, and the role of the learners in the assessment process.

Table 24

Teachers’ Decision-making Process for Planning Assessment

Theme	Category	Code	Frequency
Planning Assessment	Sources	Curriculum/Objectives	11
		Colleagues	11
		Online sources	4
		Coursebooks	3
	Factors	Reliability	14
		Construct validity	12
		Practicality	12
		Learners’ needs	8
	Learners’ Role	Washback	5
		Informing learners	17
		Involving learners	3

When asked about the sources that teachers utilize while planning their assessment practices, it is clear from their answers that most of the teachers plan their practices by taking into account the objectives and the curriculum. Similarly, in their decision-making process,

teachers often consulted to their colleagues and discussed with them to shape the common framework of their assessment practices. Some teachers express the impact of curriculum and their colleagues on their decision-making process as follows:

“We decide on our assessment practices at the very beginning. What to do for each unit or the criteria that we put on the forms, we, with my colleagues, meet on the common ground based on our meetings.” (Hale)

“I prepare my own written exams, but I definitely ask my colleague to check them since she’s been working with this grade for years. I want her to check if there’s something wrong.” (Rana)

“Of course, I personally have an objective-based approach to my assessment practices. I plan them based on the curriculum.” (Canan)

“My decision-making process is mostly about meeting the requirements of the curriculum objectives. If it’s talking about daily routines, then I try to use something that would assess that.” (Pırıl)

“I check whether it covers the curriculum or not, whether the content is appropriate or not.” (Burcu)

Moreover, teachers also made use of online sources by means of following other teachers’ practices on social media, and other supportive materials online provided by the MoNE in addition to the coursebooks they use. Below, teachers’ use of online sources and coursebook materials are given.

“I don’t get any help from any other person, but I use the tasks in the coursebooks.” (Ferhat)

“I don’t have anybody that I can get help from face to face, but there are some websites that other teachers recommend, with some interactive tools. I especially benefit from the tools on those websites. I mean, I also use social media, obviously.” (Damla)

“I benefit from EBA [the Education Information Network] and its sample questions. Also, there are some coursebooks that I use for the exams. At the same time, there are platforms created by fellow teachers on social media, I employ some the tasks on there too to plan my assessment.” (Canan)

Considering the factors affecting the decision-making process of the teachers while planning their assessment practices, it is observed that most of the teachers took the assessment principles into consideration and planned their assessment accordingly. Besides, it is

revealed that the most stated factor that teachers consider in order to shape their practices is reliability, followed by validity and practicality of the assessment. In addition to the principles of assessment, the learners' needs were also an influencing parameter that took place in teachers' decision-making process. Finally, although the positive and negative impact of their assessment practices, that is, washback, is not the primary concern for planning assessment, it is observed that some teachers included these factors in their planning process.

Teachers took the reliability into account with the written exams, and the different aspect of reliability was included in the decision-making process. Some of the teachers aimed to ensure test reliability by means of including a diversity of item types or assessment tools, consulting to a colleague, not using items that are too difficult, easy, or ambiguous while some teachers targeted the rater reliability in terms of scoring the writings anonymously, using scoring scales, or involving the learners in the scoring process. Some of the teachers' views regarding the reliability factor are as follows:

"We have group with my colleagues, I send my exams to them to tell me what they liked or didn't. Then, we make the necessary alterations on the exam together collaboratively."
(Burcu)

"For speaking, I have to assess them during the lesson, but for writing, I can pay more attention to reliability in that I check their writings twice without looking at their names."
(Pırıl)

"They may be anxious during the task and may not perform well at the time. Thus, I try to employ more than one assessment tool so that they can demonstrate what they know."
(Işın)

"I don't think that a single assessment method is advantageous by itself, so I think of it as an eclectic method.... We can get tired and there may be a problem with the scoring that we didn't notice. So, I always bring their exam sheet and give them to the students to check it and also for feedback of course." (Ahmet)

"I try to write simple, clear, and unambiguous items that wouldn't confuse the children."
(Canan)

As can be seen from the excerpts that teachers were aware of the significance of reliability and included various techniques to ensure the reliability of their assessment during their decision-making process. Besides, by adopting different assessment methods and

techniques, teachers tried to eliminate the disadvantages of utilizing only a certain type of assessment in terms of not providing enough samples of the learners' performances.

Additionally, construct validity of the assessment was another element that teachers bared in mind while planning their assessment. Most of the teachers were generally concerned about the objectives and the course content they cover in the classroom and believed that their assessment practices should mirror them. Some of the teachers' views on construct validity is as follows:

"I mean, first of all, I try to ask questions about the content we cover in the classroom for the assessment to be valid and reliable. Also, I focus on the most important parts of the content we cover." (Kaan)

"About validity, for instance, there is a unit called toys and games and this year the word 'scooter' is mentioned in the coursebook, but it wasn't in the book last year. So, I adjust the exams and don't use the same exam each year." (Leyla)

"The assessment should serve its purpose, so, I try to assess the objectives we aim to measure and design the items based on those objectives." (Canan)

Interestingly, even though most of the teachers considered reliability as a factor in their planning process and took some precautions to enhance the reliability of the assessment, construct validity was stated to be the most significant factor that teachers took into account. Teacher Burcu utters the importance that she puts on validity as *"I mostly consider if it is in line with the curriculum objectives and if the content is suitable. Actually, reliability is also important, yet validity is the first thing that I take into consideration while planning."* Therefore, it can be concluded that the decision-making process of the teachers were primarily dependent on the reflection of the course objectives and content on the assessment though they take reliability measures into account.

Another factor, which is practicality, has also been taken into consideration by many teachers to a great extent. Teachers stated that they prefer mostly practical assessment methods or tools in terms of cost, time, and feedback as a result of the large class sizes, inadequate teaching hours, and the financial considerations. Teacher Burcu touches upon the financial aspect of the assessment in line with the number of the learners she is responsible for by stating that *"Since with the class sizes of 30 to 35 and nearly 200 students as I work in a public school, it should be economical. For instance, we even consider the amount of the pictures we use on the papers in case we run out of toner."* On the other hand, Teacher Oya

explains the effects reflected on her assessment related choices in terms of feedback as *“Traditional methods and quizzes, etc. are more practical in terms of giving feedback to the students immediately. Also, we have lots of quizzes, so the cost is also important for me.”*. The teachers’ explanations about their time-related concerns in their assessment decisions are as follows:

“I prefer to use the assessment practices according to my teaching hours since we have only two hours in the primary school.” (Damla)

“Practicality matters because it shouldn’t take too much time to prepare the tool. Thus, using Web 2 tools occasionally saves me more time.” (Merve)

Moreover, learners’ needs were the factor that half of the teachers considered for their assessment plans. Elements such as the readiness and the level of the learners depending on the social and educational considerations shaped the teachers’ approach to how they plan their assessment. Teachers express the impact of their learners’ needs on their decision-making process as follows:

“For instance, some of the students didn’t participate in the lessons during the pandemic. Thus, when our assessment plan doesn’t accord with our circumstances, I have to change them. Learners’ readiness level has an impact on my assessment in that I use more simple tasks.” (Sevgi)

“Learners’ readiness really affects my decisions. I have five classes and I don’t implement the same activities for all five of them. I mean, the level of the classes varies vastly, so, I have to eliminate some activities. Thus, the level of the whole class is one of the most significant factors for me.” (Rana)

The last factor, which does not have a much influence on teachers’ decision-making process while planning their assessment practices, is washback. Although a minority of the teachers stated that their practices are influenced by washback effect, it is seen that the nation-wide standardized exams that students have to take at the end of certain education levels are reflected in the assessment practices of the teachers. Teacher Ahmet explains how the high school entrance exam shapes the content and the structure of his assessment as *“You know, 8th grades enter LGS [High School Entrance Exam] at the end of the middle school to enter a high school. I try to involve the questions and vocabulary items that students may encounter with in the exam. Even, I prioritize the items that may occur in the exam from the most to the least.”*

Based on the findings and discussions above, assessment principles such as reliability, validity, and practicality were the primary factors that influenced the teachers' assessment-related decisions. In addition to the principles of assessment, learners' needs depending on their level of proficiency and readiness were incorporated into teachers' assessment plans. It is revealed that teachers mostly referred to the practicality factor for selecting assessment methods and tools, validity factor for designing the content of the assessment tool, and reliability factor for administration and scoring process of the assessment. Overall, the assessment principles are influential to a great extent on teachers' decisions about planning assessment practices.

Learners' role in the assessment process, which is the final category in the theme planning assessment, is scrutinized. The responses in the interviews indicate that the learners did not take an active role in the assessment process in that only three teachers stated that they involved learners in the process through consulting them about the assessment type and content in addition to giving them options. Teacher Burcu expresses how she involves the learners in the process as *"I ask them to choose the content of their projects among the ten units we have in the coursebook. I give them the options, but eventually, they pick the subject themselves."* On the other hand, all of the teachers stated that they involve learners in the process by means of informing them about the assessment practices they utilized. Even though teachers expressed that they make learner-centered decisions while planning their assessment, they didn't consult or discuss with their students about the process. In this respect, teacher Sevgi utters that *"Yes, the learners' role has great importance yet if you ask me whether I consult to my students about the assessment, frankly, I don't. I only inform them about the assessment procedure."* In addition, most of the teachers stated that they informed their students about the time, content, methods or techniques, and the results of the assessment, which was typically performed within the framework of their formal exam procedures regarding the written exams. Teacher Canan and Leyla reveal the extent of how they inform their students about the written exams as follows:

"I inform the students beforehand about which subject are involved and when it will be administered. I allow a certain period for them to revise the course content of course. I also inform them about the assessment results and show them the way to review and correct their mistakes." (Canan)

"The thing is, I have never consulted the students about the assessment. Sure, I inform them about the written exams about when we're going to do it or what will be involved in the

exam. Apart from that, since I generally make observations in the classroom tasks, I don't inform them in that process.” (Leyla)

It is possible to assert that teachers were insufficient to inform their students about the criteria in their classroom assessment practices that they implement as an informal assessment procedure. The reason behind this can be explained as the teachers did not determine the assessment criteria for their assessment practices that they implement as a classroom activity. The fact that one of the assessment techniques that the teachers preferred to use the most is observation in that teachers collect evidence about learners' performances in the classroom activities through non-systematic observations can also be an indication of this result.

4.4.5. Professional Development

The table below illustrates frequencies of the ways that EFL teachers continue their professional development and their self-reflected needs regarding language assessment.

Table 25

Teachers' Professional Development

Theme	Category	Code	Frequency
Professional Development	Development Ways	Seminars	7
		Graduate courses	6
		Online courses	5
		Articles	3
	Needs	Assessment of language skills	5
		Alternative methods	4
		Technology integrated	3
		Applicable	3
		Principles of assessment	2
		Engaging	2
		Feedback	1

The table above demonstrates that teachers manage their professional development in various ways, mostly through seminars and graduate courses followed by online courses. In addition, a minority of the teachers utilized articles for their professional development. When the teachers are asked about their ways of professional development, they also mentioned that even though seminars and graduate courses have raised an awareness about the significance of the assessment for the teachers, they were generally insufficient in terms of their applicability in the classroom environment. Teachers' views about the ways of professional development and their usefulness are as follows:

“Yes, we need seminars, but the seminars offer us the ideal, how everything should be performed. However, when we are in the classroom, there is a very different scenario. For instance, even now, I’m currently taking assessment courses at a master’s program. I have learned valuable information, which is great, yet I wish I could apply them in my classes. It won’t work.” (Oya)

“I watch online videos that are that Ministry of National Education provides for in-service education. Most of them always focuses on how it needs to be or suggest an imaginary assessment for us. There are no suggestions of how assessment practices should be with class sizes of 30 to 40 students with two teaching hours a week.” (Hale)

“I have to read some articles since I’m taking an assessment course at a Masters’ program. I’m sure there are some seminars that Ministry of National Education offers, but frankly, I’m not really interested in them since I don’t find them very useful.” (Sevgi)

In addition to the quality of the in-service education, teachers touched upon the inadequacy of them in terms of their quantity. While some teachers mentioned the amount of in-service trainings regarding language assessment, some of them referred to the participant limit for these trainings. Teachers’ responses are as follows:

“I have been attending to a seminar lately regarding assessment in CEFR, and that’s all. Unfortunately, there aren’t many seminars on assessment and this one was the only one I could find in five years.” (Pırl)

“Right now, there is a seminar for language assessment based on CEFR held by the Ministry of National Education. There was also another seminar about this subject that some of my colleagues attended. However, I couldn’t attend to that one since there was a participant limit.” (Nazım)

Even though most of the teachers have attended seminars, graduate courses, online courses, or read articles regarding language assessment five of the teachers stated that they don’t continue their professional development in any ways. While some of the teachers expressed that they didn’t need to continue their professional development as they found their language assessment knowledge sufficient, others stated that professional development programs and their content is theoretical, therefore, cannot be reflected in the classroom practices. Some of the teachers’ responses regarding this issue are as follows:

"I don't think there are sound in-service education for language assessment. Even if there was, the classroom environment is not suitable for it. So, I don't know if it would be useful or not." (Ferhat)

"I'm not following any sort of books or articles right now. The truth is, it is what I learned at the time of the pre-service education and that's all. We all know the traditional methods and I've learned the alternative methods so well in the pre-service education courses. I believe my knowledge is sufficient enough for my current workplace." (Leyla)

Based on the excerpts above, it is evident that in-service trainings were inadequate both in quantity and quality. In addition, although the content of the trainings or courses have raised an awareness regarding language assessment, teachers mostly had difficulty with reflecting them in their classroom practices, which caused a bias towards these trainings on the part of the teachers, hence, decreasing the teachers' willingness to participate in in-service trainings.

When it comes to the teachers of young learners. Among 13 teachers, only two of them have continued their professional development by means of attending a seminar and taking a course in a master's program whereas 11 of them have received no in-service training for the assessment of young learners. In addition, teachers stated that they have taken some courses regarding language assessment and teaching English to young learners separately. However, the course content was not specifically designed for assessing young learners but obtained a general approach to language assessment or to young learners. Regarding the pre- and in-service trainings on assessing young learners, Teachers' statements are as follows:

"I took a course in pre-service education but not particular to young learners. The focus was not the assessment of young learners." (Elif)

"We took a course for young learners and assessment was mentioned somewhat. Besides, we also took a testing and assessment course, but I don't think young learners was involved. I also didn't attend to an in-service training." (Damla)

"I didn't participate in an in-service training, but we had a course of teaching English to young learners. However, it was mostly based on teaching language skills such as vocabulary, speaking, etc., not on assessment. Besides, the testing course we took in our final year was about assessing in general not specific to young learners." (Burcu)

The other category in the theme professional development is the needs of the teachers regarding language assessment. When the teachers are asked about what kind of a training they would like to attend, their responses varied. The first and foremost, the area that more

training is needed is the assessment of language skills. In the assessment of language skills, some of the teachers expressed their needs for speaking and writing individually while some needed training for assessing four skills. Additionally, alternative methods are the second stated area that teachers needed in-service trainings. Teachers' self-reflected needs in terms of assessment trainings are as follows:

"I would like to see different practices. What can we do besides written exams. Because I believe I need a different concept in this matter." (Elif)

"I would like to receive trainings on writing assessment since writing is a more comprehensive skill that has an order. It needs thorough consideration." (Giray)

"I wish there was a training on assessing speaking. Since the primary school students are highly eager to receive, yet I don't have the answer how. I mean, there is no training on this subject." (Damla)

"I'm curious about peer assessment. What can be done about regarding that? I want some sort of a seminar or a course regarding peer assessment and other assessment methods." (Merve)

Another training need that teachers stated is the need for more applicable and practical methods and techniques, which is related to another need that teachers touched upon, which is technology integrated assessment. Some excerpts taken from the interviews are as follows:

"Actually, it is more efficient and practical for the teachers that assignments and tasks are performed via an online tool or applications on a computer. We need to be informed about those applications." (Kaan)

"I would like to attend in trainings regarding different methods and techniques if they reflect practical concerns." (Sevgi)

"They should be more applicable. There can be a way that integrates technology into the assessment. Besides, trainings can be arranged for assessing speaking skills with different methods and techniques." (Oya)

Finally, the other training needs that are stated by a minority of the teachers are principles of assessment, giving feedback, and engaging assessment methods and techniques. Teachers express their needs as follows:

"Now that I'm working with young learners, what I struggle with the most is keeping their attention since they can get distracted easily. What kind of an assessment that children can

also enjoy could be used to draw their attention during the assessment. I would like to attend to a training in this regard.” (Canan)

“I definitely want a standardized form of feedback while assessing the students.” (Nazım)

“We all know about what validity or reliability is in theory. However, I would like to experience it in practice. For instance, the speaking exam that I design is quite unofficial. We have a curriculum and a syllabus to consult, yet I need them to be more grounded.” (Pırlı)

All in all, it is evident that teachers needed trainings in various areas of language assessment. The reason behind can be the diversity of the participant teachers in the study working at various educational levels and have different educational contexts, which may have resulted in teachers’ trainings needs to be shaped according to their own context. In addition, as it is understood from the teachers’ responses, many of the teachers mentioned the inadequacy of the trainings on language assessment in terms of quantity and limitations of participation. It is also revealed that even if they could attend to a training, they did not benefit from them due to the fact that the training content was mostly theoretical and did not contain practical concerns, hence, an exhibition of negative attitudes towards participating in the trainings is observed.

4.5. Reflection of LAK on EFL Teachers’ Classroom Practices

The fifth research question aimed to reveal whether the teachers’ LAK is reflected in their classroom practices regarding language assessment. In this vein, the participants’ knowledge level in assessing each language skills and their assessment practices are scrutinized. The teachers’ knowledge level in assessing reading was found to be adequate. In addition, their self-reported assessment practices indicated that reading was the skill that is given the most emphasis by the teachers through undertaking formal assessment procedures with pen and paper tests. Besides, the areas that teachers were knowledgeable in the scale regarding reading assessment were the genre of the reading text, methods and techniques to assess reading such as top-down approach, summarizing, the type of the test items and characteristics, adjusting the text based on the learners’ level, and enhancing reliability through collecting more samples of the learners by incorporating various item types and texts. On the other hand, teachers were not knowledgeable enough in text selection and scoring criteria in terms of involving the spelling mistakes. Likewise, they utilized traditional

methods of assessment to a great degree with quizzes and written exams focusing on reading, grammar, and vocabulary skills and gave feedback quickly to learners' mistakes in their written exams in addition to utilizing more item types and tools to gather more samples of the learners' performances. Therefore, it can be concluded that teachers' LAK in assessing reading is reflected in their practices to a great extent. The reason behind this could be related to the exam-oriented education system in which generally reading grammar and vocabulary is the focus, therefore, leading teachers to be more informed about the assessment of reading as a result of the expectation of success from the teachers and learners in high-stakes standardized tests.

When the teachers' LAK level in assessing listening, the teachers' knowledge level was insufficient even though they perceived themselves to be competent in their knowledge. The areas that teachers were knowledgeable in listening assessment in the scale were item types, the content of the text, and purposes of listening, that is, selective listening. Approaches to listening in terms of intensive and integrative, scoring procedures and purposes of discrete-point testing, scoring criteria in terms of including grammar and spelling mistakes, and text selection were the areas that teachers' knowledge was inadequate. The analysis of their assessment practices revealed that listening was ignored mostly due to time constraints and crowded classes. Besides, when they are asked about the methods, techniques, and tools for assessing language skills, the teachers did not specify their preferred techniques and tools for assessing listening skills even though they expressed that they assessed listening. In line with their perceptions and actual LAK level, a discrepancy is found between their statements and practices. Therefore, it can be inferred that their knowledge reflects their assessment practices since both their knowledge and practices are insufficient in assessing listening due to the insufficient level of knowledge in spite of high competency perceptions and not specifying their methods and techniques though stating that they assess listening.

Teachers' knowledge level in assessing writing was found to be insufficient on the contrary to their perceptions of their knowledge. The areas that teachers were knowledgeable in were the advantages and reliability aspects of analytic and holistic scoring scales, using visuals and reading texts as writing prompts, when to focus on mechanical errors, and adjusting the scoring the scale based on the learners' level and needs whereas the focus of controlled writing, when to focus on irrelevant ideas, involving more than one writing task, using opinion-based prompts, giving options to learners to write about, and scoring procedures when two raters are involved were the areas that teachers' knowledge was insufficient. When

their assessment practices are examined, it was observed that writing did not receive much attention as reading and is included in the assessment process to a certain degree. Even though teachers were aware of the use of scoring scales, most of them did not utilize a rubric to assess their learners' writing performances. Additionally, the writing tasks they employed were generally guided or controlled writing although they were not aware of the purposes of controlled writing being form focused rather than meaning, hence leading to questioning the quality of writing assessment. To conclude, it can be uttered that teachers' insufficient knowledge level transferred into their assessment practices to a certain extent.

In assessing speaking skills, teachers demonstrated an adequate knowledge in accordance with their perceptions. The areas that teachers were informed about assessing speaking skills were task types such as role-plays and discussions and incorporating various tasks to collect more samples of the learners' performances even though they were not aware that repetitions and reading aloud could be used for speaking assessment. Similarly, teachers assessed their learners' oral performances through various tasks as a classroom activity such as presentations and role-plays. Besides, teachers knew that checklists can be used as a means to score oral performances although many did not adopt a scoring scale in practice since the teachers typically made incidental observations due to limited teaching hours and large class sizes in addition to giving oral feedback based on their intuitions in the form of whole class feedback rather than individual except for extreme cases of the learners' poor performance. Considering the emphasis given to assessment of language skills, speaking skills can be ignored as well as listening in the assessment process even though teachers made use of classroom activities for speaking assessment. Thus, it can be uttered that even though teachers had adequate knowledge of assessing speaking, they couldn't reflect their knowledge in their practices to a great degree as a result of insufficient teaching hours and crowded classrooms.

All in all, even though the reflection of teachers' LAK in general on their assessment practices are observed to a moderate degree, when skill-based assessment is considered, it is seen that teachers' knowledge in assessing reading and listening are reflected on their practices whereas teachers' knowledge in assessing writing and speaking were transferred into their practices to a certain extent.

4.6. Problems of EFL Teachers in the K-12 Context Regarding Language Assessment

The sixth research question aims to demonstrate the problems that EFL teachers encounter with regarding language assessment in the K-12 context. The table below demonstrates the teachers' problems or challenges in this context in terms of teacher-related, student-related problems, and the problems in terms of the logistics of assessment.

Table 26

Teachers' Language Assessment Related Problems

Theme	Category	Code	Frequency
Problems	Teacher-related	Inadequate teaching hours	9
		Exam-oriented education system	6
	Student-related	Large class sizes	5
		Lack of motivation	4
	Logistics	Equipment	3
		Administration	2

It is shown in the table above that, the most common problem that teachers encounter with in the K-12 context is inadequate teaching hours. In addition to stating that inadequate teaching hours have led teachers to ignore assessing language skills such as listening and generally speaking since they don't have the time, it is observed that giving feedback to learners was also another result caused by insufficient teaching hours. Besides, lack of teaching hours was the biggest problem that all of the primary school teachers stated while assessing young learners. In this vein, Teacher Canan, who is a teacher at a primary school utters that *"2 hours are not enough for me. I don't think I can manage to do the things that I want to."* Some of the teachers' views regarding the problems regarding ignoring speaking assessment and giving feedback as a result of inadequate teaching hours are as follows:

"Right now, I'm having problems with the teaching hours. I can't both give instructions and assess the learners only in two lesson hours. Sometimes, run out of time and continue the following week, but then the students forget what they did. Thus, it is challenging for me in terms of giving feedback." (Damla)

"The biggest problem for me is not having enough time. For instance, I would like to assess their speaking in a more comprehensive way. Even if I give each student only a minute, I have to devote a whole lesson for assessment for 30 students in 30 minutes." (Leyla)

Another problem that teachers frequently encounter is the exam-oriented education system and its results. The teachers of young learners emphasized that the effects of exam-oriented

education system their assessment practices in that they have to conduct a written exam formally for 4th grades in addition to the standardized exams that young learners start taking even in primary school context, which doesn't match with their course objectives and curriculum related concerns. Besides, the other teachers also believed that written exams are a problem in terms of their insufficiency to reflect learners' actual language skills, hence giving unrealistic results in addition to not incorporating alternative methods since the learners are not familiar with different assessment methods and techniques as a result of exam-oriented education system. Teachers' views on the problems caused by exam-oriented education system are as follows:

"Some of my students can speak perfectly. They can understand the listening texts very well, yet they can't reflect their performance on the written exams, hence get poor grades. Unfortunately, I have to take that grade into consideration as a formal assessment. This is for 4th grades, but even though we don't have formal exams with 2nd and 3rd grades, they still take standardized tests. So, you are automatically involved in the system after a while, because if the students fail the test, it is perceived as the teachers' failure." (Hale)

"Written exams are a problem for me in terms of assessing learners' performances. Because I know that when a student is not capable of getting 90 points in an exam, but it happens with written exams. However, with practices such as performance tasks, I can see the actual performance of the student. It gives me more realistic results." (Oya)

"I work at a primary school, but the students are taking standardized tests even now and learning testing strategies. Besides, it is not appropriate for their age to be sitting down for hours and doing multiple choice questions. On the one hand, we have to focus on speaking and listening based on the curriculum. On the other hand, we have to implement a written exam for 4th grades including gap filling items, matching, and assess them based on their scores. I don't think that this is the proper approach, we need more diversity." (Canan)

"The biggest problem for me is the students' unfamiliarity with different assessment techniques. I wanted to conduct a speaking exam since I had a small number of students. I explained that I would use a rubric, but they panicked a lot. When I saw that they were anxious, I didn't do it. The fact that the students are more exam-oriented can discourage us teachers when we try to involve different methods." (Elif)

In addition to insufficient time and exam-oriented education system, large class sizes were another problem that teachers encounter with their assessment practices in terms of being

responsible for a great number of learners resulting in challenges of giving feedback, classroom management, and ignoring the assessment of speaking and listening skills. Some of the teachers' opinions in this regard are as follows:

"If had the chance, I would like to assess listening and speaking, but with 35 students in a class and spending the half of the time with managing the class, it doesn't seem possible. Therefore, not including listening and speaking assessment in the process properly is a problem for me." (Burcu)

"Checking the students' assignments in crowded classes is a bit problematic not to mention the noise that occurs. For instance, the other students won't be waiting in silence while I'm giving feedback to a student. It becomes a serious problem while giving feedback." (Ferhat)

Additionally, learners' lack of motivation was a problem that teachers stated as a student-related problem. The attitudes towards learning English and willingness to participate in the lessons were effective in teachers' choices of assessment practices. Teachers' views on the negative impact of lack of motivation on the part of the learners are as follows:

"When the students have a negative attitude towards not only English but also the concept of school, it is tiring and challenging for me to adopt alternative methods since they are unwilling to participate and lack most of the fundamental skills." (Ahmet)

"They sometimes don't want to participate in the lesson. They don't care about what we do in the class or don't say anything even though they know the answer. So, it becomes a challenge for me to figure how their performances are." (Hale)

The final category is the logistics of assessment regarding the problems teachers face with their assessment practices. It is seen that a minority of the teachers considered the lack of equipment as a problem since it influenced their practices in terms of hindering the implementation of their assessment plans while some mentioned the technical problems with the quality of the listening intervening with the assessment procedures. In this vein, Teacher Damla touches upon the copying problems by stating that *"I sometimes have difficulty finding sources or when I'm implementing an activity, I have to make copies of them. I have problems with the equipment since my school is not fully equipped."* Another issue that teachers faced is administration related problems in that the teachers have to readminister the same exam for the student who didn't participate in the assessment. The reason behind this is organizing a similar atmosphere and finding a proper time for conducting the exam again considering the learners that the teachers are responsible for, hence, workload. In this

regard, Teacher Pırl explains the problems of administration as *“For instance, the students couldn’t complete the speaking or writing task and I gave them feedback on it. It is challenging for me to administer the same exam again because I have 250 students and 40 teaching hours a week. It is hard to find the appropriate time both for me and them.”*.

All in all, based on the teachers’ responses, the most vital problem is the lack of time. As a consequence, listening and speaking skills are not emphasized in the assessment practices or completely ignored. In addition to the inadequate teaching hours, teachers also mentioned their problems with giving proper feedback in crowded classroom environments. When the teachers were asked about the language skills they mostly assess in their classes, it was revealed that listening and speaking skills were not stressed as much as reading and writing skills in addition to grammar and vocabulary and mostly based on the teachers’ incidental observations in the classroom in addition to providing insufficient feedback to learners except for their written exams. Therefore, it can be inferred that lack of time and large class sizes are critical problems that teachers encounter with since they negatively influence their assessment related choices and practices. Another frequently mentioned problem is the impact of exam-oriented education system both on learners and teachers, namely washback. The importance given to tests and exams in the educational system in the K-12 context have resulted in learners’ unfamiliarity with different assessment methods and techniques other than written exams, and therefore, teachers become hesitant to utilize alternative assessment methods in their assessment practices. Furthermore, it is understood that primary school teachers are also suffering from the exam-oriented system in that even in the context of young learners, the learners are taught testing strategies and took standardized tests, which have led teachers to adopt tools that are not suitable for young learner assessment as a result of the expectations from teachers and learners. Besides, it is clear that there is a mismatch between the curriculum expectations and the obligation to conduct written exams for young learners as an institutional policy. Finally, learners’ lack of motivation, readministering the exams, and lack of equipment in the institutions have a negative impact on the teachers’ assessment plans and practices.

CHAPTER V

CONCLUSION AND IMPLICATIONS

5.1. Conclusion

This study investigated the language assessment literacy of the EFL teachers in their knowledge base and its' reflection in their assessment practices in the K-12 context of Turkey. In this vein, the teachers' language assessment knowledge is investigated with a scale that is developed by Ölmezer-Öztürk and Aydın (2018) and semi-structured interviews are utilized to explore the assessment practices of the teachers. This study also explored the impact of the teachers' demographic information on their LAK level and whether there is a difference between their perceived self-competency in each language skills and their actual LAK level or not. Another aim of the study is to explore to what extent is the teachers' LAK is reflected on their assessment practices. Finally, the problems and challenges that the EFL teachers encounter with in the K-12 context of Turkey regarding language assessment is an element that this study is concerned with.

The impact of assessment on learning process is indisputable. The awareness about assessment has raised considerably with the importance given to assessment and advancement in the approaches to teaching and assessment. Therefore, teachers are expected to be literate in language assessment since it is unlikely to employ sound assessment practices and make proper judgements about the learners without bearing the qualifications of language assessment literacy (Fulcher, 2012). LAL typically comprises of both theory and practice. In other words, teachers need to have a degree of knowledge regarding language assessment and the skills to reflect their knowledge on their assessment practices (Inbar-Lourie, 2008). Besides, the knowledge base of LAL is considered as a prerequisite for teachers to mirror their knowledge of theoretical considerations into their classroom practices appropriately (Xu & Brown, 2016). In line with these considerations, the general

and skill-based LAK is investigated. The results indicated that EFL teachers had sufficient knowledge in language assessment supporting the finding of other studies that focused on the EFL teachers' knowledge in higher education context (Hakim, 2015; Jannati, 2015). However, in another study by Ölmezer-Öztürk and Aydın (2019a), in which the teachers' LAK was inadequate in the context of tertiary education. Therefore, it can be concluded that the EFL teachers in the K-12 context have adequate LAK considering the contextual differences between the aforementioned studies, all of which focused on the LAK in the higher education context. When it comes to the skill-based LAK, the findings illustrated that the teachers had sufficient knowledge respectively in assessing reading and speaking, reading being the highest, and insufficient knowledge in assessing writing and listening, listening being the lowest. In line with the results of another study by Ölmezer-Öztürk and Aydın (2019), in which the knowledge of the teachers was the highest in reading assessment and the lowest in listening assessment even though they had low levels of knowledge in the assessment of these skills. Similarly, Genç, Çalışkan and Yüksel (2020) investigated the LAK level of high school EFL teachers in assessing writing and speaking and found that teachers were knowledgeable in assessing speaking whereas their knowledge in assessing writing was significantly low. Thus, it can be commented that the EFL teachers in the primary, middle, and high schools in Turkey are more knowledgeable in reading and speaking assessment while insufficient knowledge is illustrated in writing and listening assessment even though they demonstrated a higher degree of LAK in general.

When the impact of the teachers' demographic characteristics are examined, it is revealed that no significant effect was found on the teachers' LAK according to their, gender, level of experience, the BA program being graduated, whether having a BA, MA, and PhD degree, working at a private or a state school, the educational level of the workplace whether primary, middle, or high school, the geographical region they work in, taking a language assessment course in pre-service education, and attending professional development programs. These findings support the results of other studies regarding gender, year of experiences, educational background in different educational contexts (Büyükkarcı, 2016; Jannati, 2015; Öz & Atay, 2017; Xu & Brown, 2017). On the other hand, a contradiction is displayed with a study in terms of the effect of the type of the institution being private or state and experience level by Farhady and Tavassoli (2018). It was found that the teachers who have more experience and work at a state institution had higher levels of knowledge in language assessment. The reason behind this difference in findings can be related to the

number of the participants in this very study, in which only a minority of the teachers worked at a private institution. As a consequence, it can be uttered that the EFL teachers' LAK is possibly affected by the type of their institution. In conclusion, the contextual and experiential factors are not influential in shaping the EFL teachers' LAK in the K-12 context of Turkey.

In the literature, a discrepancy is observed between the teachers' perceptions and their LAL and knowledge (Farhady & Tavassoli, 2018; Işık, 2021). Regarding the competency level of the teachers as they perceive themselves to be in assessing each language skill and their actual LAK, the results indicated that in general, teachers' perceptions are reflected in their knowledge in that they generally felt competent and indicated high LAK level. However, when the language skills are focused individually, it is shown that the teachers' perceived self-competency in assessing reading and speaking is in accordance with their actual knowledge level while their self-competency in assessing writing and listening did not indicate a correlation, therefore a misperception between their actual knowledge level and self-competency occurred in that even though the teachers perceived themselves to be competent in assessing writing and listening, their knowledge level was found to be low.

This study also investigated the assessment practices that teachers perform in their classrooms in primary and secondary education contexts. Based on the teachers' responses to the questions in the semi-structured interviews, the teachers defined their language assessment practices around five concepts of the characteristics of language assessment, assessment of language skills, assessment results, planning assessment, professional development. The results show that the teachers mostly used assessment with the aim of determining the degree of learning. Besides, another common aim that teachers possessed was to diagnose learners' weaknesses and adjust their classroom practices accordingly in addition to monitoring the learning process. Even though the teachers utilized assessment for giving feedback to learners, determining if the objectives are met or not, and enhancing motivation, they were not among the most common purposes of assessment. In line with the findings of Önalın and Gürsoy (2020), who also found that teachers made use of assessment for reflecting on their assessment practices and designing syllabi in addition to determining achievement and giving feedback to learners. When the time of the assessment is scrutinized, it is seen that all of the teachers employed assessment during their instruction formatively. Besides, teachers also adopted summative practices at the end of certain periods with written exams as a formal procedure of their institutional policy. On the other hand, assessment

before instruction was preferred by only a minority of the teachers for placement purposes, coursebook selection, and determining the readiness level of the learners. It can be inferred that a balanced approach is adopted by the teachers given that they both employed formative and summative purposes in addition to utilizing assessment embedded in their classroom practices on the contrary to what Gu (2014) found in a study that Chinese EFL teachers considered assessment as a form of evaluation rather than a part of their classroom instruction. Therefore, it can be concluded that EFL teachers in the K-12 context utilized assessment for both formative and summative purposes and considered assessment as a part of their classroom instruction rather than just giving grades and evaluation, but to improve learning and teaching by adjusting their instruction based their learners' performances. The results support the findings of other studies in the literature (Jannati, 2015; Önalán & Gürsoy, 2020). Additionally, in the young learner context, the teachers believed that young learners are highly motivated towards learning English. Therefore, they adopted engaging tasks and mostly focused on learners' strengths by ignoring simple mistakes during their assessment so that learners can build positive attitudes towards learning English. Besides, the developmental characteristics of young learners are taken into consideration by the teachers in terms of their low attention span and using visual aids in assessment practices.

It is found that teachers incorporated most of the language skills in their assessment practices for reading, listening, speaking, grammar, vocabulary, and writing. Besides, all of the teachers adopted traditional methods and alternative methods were employed by most of the teachers supporting the findings of Giraldo (2019). Among traditional methods written exams and quizzes were the most popular tools that teachers used formally for assessing learners' reading, grammar, and vocabulary. For alternative tools and techniques, classroom observation, portfolios, oral performances were the most common techniques that teachers utilized while checklists/rubrics, projects, games, self-assessment, songs, digital tools, journals, competitions, and dictation was not frequently used by the teachers. Similar to the findings of Tanyer and Susoy (2018), for speaking skills, oral performances of the learners in role-plays and presentations are mostly relied on the teachers' observations which are typically incidental since most of the teachers did not utilize a checklist or a rubric as a scoring scale, which may lead to subjectivity of the interpretations based on the learners' performances (Fulcher, 2003). For writing assessment, teachers generally employed portfolio assessment while for listening most of them did not mention how they assess the listening performances of their students even though they stated that they assess listening

except for primary school teachers, who adopted different tools and techniques such as games, songs, and dictation for listening assessment. As a consequence, a difference in the assessment practices in terms of emphasis given to language skills and assessment techniques used between the teachers who work at primary, middle and high schools. While primary school teachers emphasized the listening and speaking skills, the reading and writing skills remained very limited as guided by the curriculum. However, reading, grammar, vocabulary, and writing to a certain degree were the main focus of the middle school teachers in their assessment practices while speaking and listening skills can be ignored during the process confirming the results of a study by Arslan and Üçok-Atasoy (2020). Listening assessment can be ignored because of the complexities of the spoken language such as informal language, dialect, etc., and the challenge to define the construct of listening (Buck, 2001). On the other hand, high school teachers incorporated the assessment of four skills as a part of their institutional policy. Consistent with the findings of Yan, Zhang, and Fan (2018), commenting on the impact of contextual factors on teachers' assessment practices, it is concluded that the assessment practices of the teachers in terms of the emphasis given to language skills and selection of methods, techniques, and tools is influenced by the contextual factors such as the educational level of the workplace.

As for the use of assessment results, teachers mostly utilized them to make judgements about their classroom practices, learners' weaknesses to give feedback coinciding with the findings of Önalın and Gürsoy's (2020) study in the Turkish K-12 context. In addition, grading, evaluation, and making realistic judgements about the learners' performances were among the decisions that a minority of the teachers made based on the assessment results. When the form of feedback is examined, the results indicate that all of the teachers gave feedback as quickly as possible to the learners' written exams and quizzes. Oral feedback was another common feedback type that teachers preferred in their classroom practices, which generally takes place as a whole class feedback rather than individual in that individual feedback was only adopted when there is an extreme case of poor performance due to time constraints and classroom management problems. In addition to quick and oral feedback, the other types of feedback that were not adopted by the majority were written, positive, corrective, teacher student conferences, and delayed feedback. Even though positive feedback is not a common feedback type, it is utilized by the teachers in the young learner context. As a conclusion, it is observed that the teachers' use of feedback is inadequate for the assessment practices other than written exams since they mostly relied on whole class oral feedback rather than

individual feedback on the contrary to what Tsagari and Vogt (2017) found in their study in that individual feedback is generally preferred by the teachers in Germany, Cyprus, and Greece. An explanation can be made upon this issue as even though the teachers adopted both formal and informal methods and adopted assessment as embedded in their daily instruction with classroom activities, they were mostly based on the intuitions of the teachers based on their observations, which might have led to inadequate feedback as a result of not making sound interpretations of the learners' performances through non-systematic collection of assessment data.

EFL teachers' assessment practices in terms of how they plan their assessment by means of the sources they use, the factors that contributes to their decision-making process, and the role of the learners in the assessment process is examined. The results displayed that teachers' assessment practices mostly based on the curriculum and curriculum objectives in that they plan their assessment practices in light of the curriculum. Besides, most of the teachers either consulted to their colleagues or determined the language assessment procedure together. Teachers also made use of online sources and coursebooks to a certain degree through observing the classroom practices of other EFL teachers on social media, and online materials provided by the MoNE. Therefore, it can be understood that since the teachers' assessment practices were stated to be in line with the course objectives and curriculum and discussed their practices for planning, the validity and reliability of their assessment practices are increased.

Likewise, reliability and construct validity in addition to practicality were the factors that are taken into account in their decision-making process for their assessment plans. Teachers utilized various techniques to ensure reliability such as adopting different methods, techniques, and item types in order to gather more samples of the learners. Thus, it is clear that teachers eliminated the disadvantages of using only a certain method or technique. Additionally, even though the reliability was the most frequently stated factor, teachers put forth that the validity of the assessment is their primary concern in their assessment plans, which can be linked to the main source that they use which was the curriculum and objectives. Furthermore, teachers generally preferred more practical methods and instruments in terms of time, cost, and feedback due to the large class sizes, inadequate teaching hours, and financial considerations. In addition to the assessment principles, learners' needs based on their language proficiency and readiness were also taken into consideration in teachers' decision-making process. Finally, it is observed that washback,

though not to a great extent, was influential in the teachers' assessment-related choices in terms of the impact of nation-wide standardized exams and its reflection on teachers' assessment plans, which is a subject of discussion in the foreign language education context in Turkey in terms of the impact of test-based practices as a consequence of high stakes exams that learners have to take (Hatipoğlu, 2016). In line with the Hudaya's (2017) study, focusing on the application of assessment principles, in which the principles of assessment were taken into consideration in their assessment practices, it is evident in this study that teachers' decision-making process for planning assessment is considerably influenced by the assessment principles and learners' needs to a certain extent. However, it is seen that teachers considered practicality for selecting assessment methods and tools, validity for designing the content of the assessment tools, and reliability for scoring and administration-related concerns.

The result of this study is in accord with a study by Önalán and Gürsoy (2020) in terms of the learner involvement in the assessment process, which was found to be at a moderate degree except when self-assessment is utilized. The learners' role in the assessment process was another concern of the study in that to what extent learners are involved in the assessment process. Even though self-assessment was employed to a certain degree, learners are not actively involved in the assessment other than being informed about the time, content, and techniques and tools of assessment they utilized for the formal assessment practices with written exams though teachers made learner-centered decisions for their assessment plans and incorporated self-grading procedures to a certain extent.

For their professional development in language assessment, teachers mostly attended to seminars and took graduate courses in a master's program in addition to taking online courses by means of watching videos on a platform provided by the MoNE as a part of their in-service education. Besides, scientific articles were also used by a minority of the teachers who are studying at a master's program. However, it is seen that there were not many trainings provided in the language assessment field especially regarding the assessment of young learners, and the participation can be limited. Even though teachers believed that seminars and graduate courses have raised an awareness about the importance of assessment, they were not found useful within the perspective of classroom assessment since they were mostly theoretical and lacked practical concerns supporting the findings of a study by Ölmezer-Öztürk and Aydın (2019b), and Işık (2021). As a result, a bias occurred on the part of the teachers in that their willingness to participate in in-service trainings decreased. Thus,

it is shown that in-service trainings are inadequate both in quantity and quality. However, since the quality of in-service trainings in language assessment is out of the scope of this research, the conclusions are drawn upon the teachers' self-reflected perceptions about the quality of in-service trainings. Based on the lack of trainings in language assessment, teachers' training needs varied vastly. Teachers mostly needed trainings in assessing language skills and particularly in speaking besides how to adopt more alternative methods in their practices. Additionally, teachers needed to be informed about the ways to incorporate technology in their assessment practices in addition to more practical and applicable since they didn't consider the existing trainings efficient to be reflected in their assessment practices. Besides, assessment principles, giving feedback, and more engaging assessment methods were the training needs that teachers expressed. It is clear that teacher need trainings in various areas of language assessment, which could be related to the diversity of the context they teach, hence varying training needs.

LAL entails various components including knowledge of theoretical considerations and skills to facilitate assessment-related activities (Inbar-Lourie, 2013a). In other words, LAL is considered as the reflection of the theoretical knowledge into practice (Malone, 2013). The knowledge base of LAL is the most crucial component of LAL in that sound assessment practices entail a degree of assessment knowledge (Fulcher, 2012; Xu & Brown, 2016). Therefore, the extent to what teachers' LAK is reflected in their practices has been examined. It is found that teachers had sufficient knowledge in assessing reading and speaking skills. However, while teachers could project their knowledge in assessing reading into their practices regarding reading assessment, their knowledge was reflected in their practices in terms of assessing speaking to a certain extent. In addition, teachers' knowledge was insufficient in assessing writing and listening. When the reflection of their insufficient knowledge in their practices are investigated, it is seen that teachers mirrored their knowledge in assessing listening into their practices to a great degree since listening was ignored during the assessment process. Besides, since the teachers perceived themselves to be competent in assessing listening and expressed that they assessed listening of their learners, a discrepancy is observed between their perceptions and LAK and practices. Writing assessment, which was the skill that teachers indicated inadequate knowledge, did not receive much attention as reading in teachers' practices. Besides, although teachers were aware of the scoring scales only a few of them adopted rubrics for performance assessment. Teachers mostly adopted controlled writing tasks for assessing writing, yet it was seen that

they were not familiar with the focus of controlled writing being form focused rather than to convey new information and meaning, which lead to questioning the quality of writing assessment. Therefore, it can be concluded that teachers' knowledge was reflected on their practices for assessing writing to a certain extent. As a conclusion, EFL teachers in the K-12 context of Turkey had an adequate level of LAL to a certain degree.

Finally, the most common problem that EFL teachers in the K-12 context encounter with is the time constraints due to inadequate teaching hours, which led to ignoring the assessment of listening and typically speaking. Besides, as a result of insufficient time and large class sizes, teachers could not give proper feedback to their learners individually since they were responsible for the assessment of a large number of students in line with findings of other studies by Chan (2008) and Ballıdağ and İnan-Karagül (2021). Another problem was the negative impact of exam-oriented education system on the expectations from the teachers and learners in that learners were not familiar with different assessment methods and techniques leading the teachers to adopt more traditional methods even though they believe that written exams are considered as inappropriate tools to assess learners' actual language performances. Besides, there was a mismatch between the curriculum expectations and the expectations from the teachers and learners in the primary school context in that even 2nd and 3rd grades take high stakes standardized tests even though formal examination begins at 4th grade. Therefore, teachers considered it as a crucial problem since it leads to a dilemma between whether to meet the expectations of the society or the considerations of young learner assessment suggested in the curriculum. Another problem that resulted in utilization of more traditional methods and not making proper judgements of the learners' performances was the learners' lack of motivation. Finally, the lack of technical equipment of the institutions and readministering the assessment since the learners did not participate were among the problems that teachers faced in relation with the logistics of the assessment.

5.2. Pedagogical Implications

This study investigated the LAK and practices of the EFL teachers in the Turkish K-12 context. Based upon the results, pedagogical implications are drawn upon the problematic areas in teachers' language assessment literacy based on their knowledge and practices. The teachers utilized both formative and summative, formal and informal, traditional and alternative methods in a balanced approach. However, due to inadequate teaching hours and large class sizes, it is shown that reading, grammar, and vocabulary skills are typically

assessed with written exams as a formal procedure while speaking, writing, and listening are assessed informally based on the classroom observations without incorporating a scoring scale, which may lead to inappropriate judgements about the learners' performances and questioning of the reliability of the assessment without scoring criteria. Therefore, in order to make sound interpretations of the learners' performances, teachers need to adopt more reliable procedures by utilizing checklists and rubrics more. Besides, since the teachers' judgements are based on their incidental observations for assessing speaking skills rather than keeping records of the learners' performances, a suggestion that can be made on this issue is that the institutions can support the use of technological applications and tools where learners' speaking performances are assessed, so that more practical applications in terms of time can be included in the assessment process. Therefore, in-service trainings can be implemented in technology integrated assessment and the use of proper technological tools. Besides, MoNE can provide these technological applications for teachers to utilize in their classrooms so that speaking skills are incorporated into classroom activities more systematically to collect assessment data on learners' speaking skills since they can be ignored in the assessment process even though teachers have an adequate knowledge in assessing speaking.

Due to time constraints and crowded classrooms, teachers had problems with giving feedback individually for the assessment tasks they adopt as a classroom activity. As a conclusion, the teaching hours given to teaching English language needs to be increased, especially in the primary schools, where there is only 2 teaching hours for English language, for the teachers to incorporate more age-appropriate assessment tasks in consideration of their developmental characteristics. Another suggestion can be made for the policy makers in that similar to the abovementioned suggestions, the appropriate technological tools and applications can be supplied by the MoNE to the use of EFL teachers so that they can give more individualized feedback rather than whole class oral feedback to the learners' performances and works.

The assessment practices of the teachers in the context of young learners were considerate of the learners' developmental characteristics in terms of motivation, short attention span, enhancing positive attitudes by focusing on success rather than failure, hence positive feedback was preferred, and incorporating more engaging tasks. Besides teachers adopted a variety of assessment tasks for listening and speaking such as songs, games, digital games, act-out, etc. as suggested by the curriculum. Additionally, even though teachers did not adopt

an exam-based approach to assessment, they had to implement written exams as a part of the institutional policy, and the standardized test that learners take even in primary schools led to a dilemma between the teachers' approaches to assess young learners and the expectation of success as a result of exam-oriented education system. As a result, this might lead to negative washback on the young learners and decrease their motivation and attitudes towards learning English. Therefore, it is suggested that teachers keep focusing on the progress of the learners with classroom assessment tasks formatively, and standardized tests should be eliminated in the context of young learners.

Another implication can be made upon the listening skills, in which teachers' knowledge and practices were insufficient though indicated high perceptions of their knowledge and practices. Therefore, teachers need to develop their theoretical knowledge in assessing listening and their practices accordingly. As a result, it is crucial that in-service trainings are provided for the teachers with a focus on listening assessment.

Finally, the in-service trainings were inadequate in quantity regarding language assessment, and the quality of the existing trainings is questioned in that the content of the in-service trainings was found to be theoretical rather than practical in addition to adopting a broad approach to language assessment or assessment superficially without emphasizing a specific areas of language assessment. Thus, based on the needs expressed by the teachers, the number of the in-service trainings with regard to language assessment needs to be increased. It can be suggested that the content of the trainings should focus on a specific area of language assessment rather than in general such as assessment of language skills individually, giving feedback, scoring procedures and use of scoring scales, alternative assessment methods, technology-integrated assessment, and self- and peer-assessment that incorporates more applicable and practical concerns.

5.3. Suggestion for Further Research

This study is interested in the LAK level and practices of the teachers working at the primary, middle, and high school in the K-12 context of Turkey within the academic year of 2021-2022. 202 EFL teachers have participated in the knowledge scale for the quantitative phase of the study while for the qualitative phase, 17 EFL teachers have been interviewed with semi-structured interviews. Based on this, suggestions for further research are provided for the researchers.

Even though there are some studies regarding LAL, LAK, and assessment practices of the teachers, the context of these studies was interested in the tertiary education. Considering the limited number of research conducted in the K-12 context regarding LAL, more studies should be carried out in order to validate the LAL of the EFL teachers in their knowledge base and assessment practices in this particular context.

The data collection in this study is limited to the LAKS to determine EFL teachers' knowledge level in language assessment and the semi-structured interviews to reveal their assessment practices. Since the number of the participants in the scale is limited to 202 EFL teachers, the results may not be generalized to the larger population. Thus, further research can be performed with more participants in order to obtain more reliable results regarding EFL teachers' LAK. Besides, the teachers' assessment practices are limited to their self-reported practices in the interviews. Thus, to be able to make more in-depth discussion of the teachers' assessment practices, further research can incorporate classroom observations as a data collection tool.

Finally, the problematic areas in the EFL teachers' language assessment practices were determined as the lack of emphasis given to listening and speaking skills, the reliability of the interpretations and judgements of learners' performances based on the teachers' intuitive observations in classroom activities, in other words, the reliability of the scoring procedure, and not providing adequate feedback individually except for the learners' performances on written exams. Therefore, a distinctive study can be conducted focusing on these problematic areas in teachers' language assessment practices.

REFERENCES

- Abeywickrama, P. (2018a). Purposes of listening assessment. In J. I. Lontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. <https://doi.org/10.1002/9781118784235.eelt0624>
- Abeywickrama, P. (2018b). Techniques for assessing listening. In J. I. Lontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. <https://doi.org/10.1002/9781118784235.eelt0623>
- Akbari, R. (2012). Validity in language testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyanoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 30-36). USA: Cambridge University.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. <https://doi.org/10.1093/applin/14.2.115>
- Alderson, J., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching*, 34(4), 213-236. doi:10.1017/S0261444800014464
- Alderson, J., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35(2), 79-113. doi:10.1017/S0261444802001751
- Arslan, R. Ş., & Üçok-Atasoy, M. (2020). An investigation into EFL teachers' assessment of young learners of English: Does practice match the policy? *International Online Journal of Education and Teaching (IOJET)*, 7(2), 468-484.
- Ary, D., Jacobs, L.C., Sorensen, C., Razavieh, A. (2010). *Introduction to Research in Education* (8th ed.). California: Wadsworth.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University.

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University.
- Bailey, A. L. (2017). Assessing the language of young learners. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language education: Language testing and assessment* (3rd ed., pp. 323-342). Cham: Springer. https://doi.org/10.1007/978-3-319-02261-1_22
- Bailey, A. L., Heritage, M., & Butler, F. A. (2013). Developmental considerations and curricular contexts in the assessment of young language learners. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1). London: John Wiley & Sons.
- Ballıdağ, S., & İnan-Karagül, B. (2021). Exploring the language assessment literacy of Turkish in-service EFL teachers. *Balıkesir University the Journal of Social Sciences Institute*, 24(45), 73-92. DOI: 10.31795/baunsobed.909953
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2) 113-123. <https://doi.org/10.1093/elt/ccy055>
- Black, P. (1998). *Testing: Friend or foe? Theory and practice of assessment and testing*. London: Falmer.
- Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, 21(1), 49-97. DOI: 10.1080/03057269308560014
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74. DOI: 10.1080/0969595980050102
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90. <https://doi.org/10.1177/003172171009200119>
- Boyles, P. (2005). Assessment literacy. In M. H. Rosenbusch (Ed.), *New visions in action. National assessment summit papers* (pp. 18-24). Ames: IA: Iowa State University.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191. doi:10.1017/S0267190500003536

- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge, UK: Cambridge University.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. New York: Pearson Longman.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. D. (2012). Classical test theory. In G. Fulcher & F. Davidson (Eds.), *Routledge handbook of language testing* (pp. 303–15). New York, NY: Routledge.
- Brown, J., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675. Doi: 10.2307/3587999
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University.
- Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), 107-120.
- Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Eğitim ve Öğretim Dergisi*, 6(3), 333-346. <https://doi.org/10.14527/pegegog.2016.017>
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford: Oxford University.
- Chan, Y. C. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, 7(1), 37-62.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272. doi:10.1017/S0267190599190135.
- Chapelle, C. A. (2012). Reliability in language assessment. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Malden, MA: Wiley-Blackwell.

- Chapelle, C. A., & Voss, E. (2013). Evaluation of language tests through validation research. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 3, pp. 1081-1097). London: John Wiley & Sons.
- Cheng, L., & Fox, J. (2017). *Assessment in the language classroom: Teachers supporting student learning*. London, UK: Palgrave Macmillan.
- Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, 20, 147-161.
- Coombe, C. (2018). *An A to Z of second language assessment: How language teachers understand assessment concepts*. London, UK: British Council.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyanoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20-29). NY, USA: Cambridge University.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3). <https://doi.org/10.1186/s40468-020-00101-6>
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Cumming, A. (1996). Introduction: The concept of validation in language testing. In A. Cumming & R. Berwick (Eds.), *Validation in language testing: Modern language in practice 2* (pp. 1-14). Clevedon: Multilingual Matters.
- Davies, A. (1990). *Principles of language testing*. Oxford, England: Blackwell.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. DOI: 10.1177/0265532208090156.
- 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. doi:10.1080/0969594X.2010.516643.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford.

- Douglas, D. (2010). *Understanding language testing*. New York, NY: Routledge.
- Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning and assessment for learning*. Winnipeg, MB: Minister of Education, Citizenship and Youth.
- Falsgraf, C. (2005). Introduction. In M. H. Rosenbusch (Ed.), *New visions in action. National assessment summit papers* (pp. 5-8). Ames, IA: Iowa State University.
- Farhady, H. (2012). Principles of language assessment. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp. 37-46). USA: Cambridge University.
- Farhady, H. (2018). History of language testing and assessment. In J. I. Liantas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. DOI: 10.1002/9781118784235.eelt0343
- Farhady, H., & Tavassoli, K. (2018). Developing a language assessment knowledge test for EFL teachers: A data-driven approach. *Iranian Journal of Language Teaching Research*, 6(3). 10.30466/ijltr.2018.120602
- Farhady, H., & Tavassoli, K. (2021). EFL teachers' perceptions and practices of their language assessment knowledge. *Language Testing in Asia*, 11(17). DOI:10.1186/s40468-021-00135-4
- Fox, J. (2017). Using portfolios for assessment/alternative assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language education: Language testing and assessment* (3rd ed., pp. 135-147). Cham: Springer. https://doi.org/10.1007/978-3-319-02261-1_9
- Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). *How to design and evaluate research in education*. (8th ed.) New York: Mc Graw Hill.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow: Longman Pearson.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <http://dx.doi.org/10.1080/15434303.2011.642041>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.

- Genç, E., Çalışkan, H., & Yüksel, D. (2020). Language assessment literacy level of EFL teachers: A focus on writing and speaking assessment knowledge of the teachers. *Sakarya University Journal of Education*, 10(2), 274-291. doi: <https://doi.org/10.19126/suje.626156>.
- Gipps, C., & Stobart, G. (2003). Alternative assessment. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 549-575). Dordrecht: Kluwer Academic.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179-195. <https://doi.org/10.15446/profile.v20n1.62089>.
- Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *HOW Journal*, 26(1), 35-61. <https://doi.org/10.19183/how.26.1.481>
- Giraldo, F. (2021). A reflection on initiatives for teachers' professional development through language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 23(1), 197-213. <https://doi.org/10.15446/profile.v23n1.83094>
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. New York, NY: Routledge.
- Griffie, D. T. (2012). *An introduction to second language research methods: Design and data*. Berkeley: TESL-EJ.
- Gu, P. Y. (2014). The unbearable lightness of the curriculum: what drives the assessment practices of a teacher of English as a foreign language in Chinese secondary school? *Assessment in Education: Principles, Policy & Practice*, 21(3), 286-305. <https://doi.org/10.1080/0969594X.2013.836076>
- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education & Literacy Studies*, 3(4), 42-48. doi:10.7575/aiac.ijels.v.3n.4p.42
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295-303. <https://doi.org/10.1177/026553229701400306>
- Hamp-Lyons, L. (2016). Purposes of assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 13-27). Berlin: DeGruyter Mouton.

- Hamp-Lyons, L. (2018). Assessing writing skills. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis, & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 46-80). Cyprus: Nicosia.
- Hansen Edwards, J. G. (2013). Peer assessment in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 2). London: John Wiley & Sons. <https://doi.org/10.1002/9781118411360.wbcla002>
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbooks of applied linguistics [HAL]: Handbook of second language assessment* (pp.413-427). Berlin, Germany/Boston, MA: De Gruyter Mouton.
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.
- Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal*, 51(1), 12-20. <https://doi.org/10.1093/elt/51.1.12>
- Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337-354. doi:10.1191/0265532205lt312oa
- Hasselgreen, A. (2012). Assessing young learners. In G. Fulcher & F. Davidson (Eds.), *The routledge handbook of language testing* (pp. 93-105). New York, NY: Routledge.
- Hatipoğlu, Ç. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia – Social and Behavioral Sciences* 232, 136-144. doi: 10.1016/j.sbspro.2016.10.038
- Heaton, J. B. (1990). *Writing English language tests* (3rd ed.). Cambridge: Cambridge University.
- Henning, H. (1987). *A guide to language testing: Development, evaluation and research*. Boston, MA: Heinle.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145.
- Herrera, L., & Macias, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. DOI: <http://dx.doi.org/10.14483/udistrital.jour.calj.2015.2.a09>

- Hidri, S. (2018). Discrete point and integrative testing. In J. I. Liontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. <https://doi.org/10.1002/9781118784235.eelt0375>
- Hinton, P.R., McMurray, I., & Brownlow, C. (2014). *SPSS explained* (2nd ed.). New York, NY: Routledge.
- Hubley, N. J. (2018). Reading assessment. In J. I. Liontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. <https://doi.org/10.1002/9781118784235.eelt0336>
- Hudaya, D. W. (2017). Teachers' assessment literacy in applying principles of language assessment. *The 1st Education and Language International Conference Proceedings Center for International Language Development of Unissula*. 247-260.
- Huerta-Macías, A. (2002). Alternative assessment: Responses to commonly asked questions. In J. Richards & W. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 338-343). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667190.048
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University.
- İnan-Karagül, B., Yüksel, D., & Altay, M. (2017). Assessment and grading practices of EFL teachers in Turkey. *International Journal of Language Academy*, 5(5), 168-174. <http://dx.doi.org/10.18033/ijla.3656>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. doi:10.1177/0265532208090158
- Inbar-Lourie, O. (2013a). Guest Editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301-307. DOI: 10.1177/0265532213480126
- Inbar-Lourie, O. (2013b). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Blackwell Publishing. DOI: 10.1002/9781405198431.wbeal0605.
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language education: Language testing and assessment* (3rd ed., pp. 257-270). Cham: Springer. https://doi.org/10.1007/978-3-319-02261-1_19

- Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131-146). Berlin: DeGruyter Mouton.
- Işık, A. (2021). Exploring how ELT teachers perceive and practice English language assessment. *Journal of Language and Education*, 7(1), 109-126. <https://doi.org/10.17323/jle.2021.10296>
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teachers Education*, 6(2), 26-37.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345-362. DOI: 10.1177/0265532213480334
- Jones, N. (2012). Reliability and dependability. In G. Fulcher & F. Davidson (Eds.), *The routledge handbook of language testing* (pp. 350-363). New York, NY: Routledge.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535. <https://doi.org/10.1037/0033-2909.112.3.527>.
- Kırkgöz, Y., Babanoğlu, M. P., & Ağçam, R. (2017). Turkish EFL teachers' perceptions and practices of foreign language assessment in primary education. *Journal of Education and e-Learning Research*, 4(4), 163-170. DOI: 10.20448/journal.509.2017.44.163.170
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120. DOI:10.1080/15434303.2019.1674855
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.
- Lado, R. (1961). *Language testing*. New York, NY: McGraw Hill.
- Lam, R. (2014). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197. doi:10.1177/0265532214554321
- Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. London: Routledge Falmer.

- Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Testing Research*, 15(1), 11-33. <https://doi.org/10.1177/1362168810383328>
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University.
- Lynch, B., & Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly*, 28(4), 727-743. doi:10.2307/3587557
- Mackey, A. and Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, N. J.: Lawrence Erlbaum.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344. DOI: 10.1177/0265532213480129
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research, and Evaluation*, 7(8), 1-5. DOI:<https://doi.org/10.7275/5kc4-jy05>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA, and Oxford, England: Blackwell.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi-Ankara Üniversitesi TOMER [Language Journal-Ankara University TOMER]*, 167(1), 43-60.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11. <https://doi.org/10.3102/0013189X018002005>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- MoNE-ELT Competencies. (2008). *İngilizce öğretmenleri özel alan yeterlikleri [The specialized field competencies of English language teachers]*. Retrieved from https://oygm.meb.gov.tr/meb_iys_dosyalar/2017_11/06160113_2-

YYretmen_Yeterlikleri_KitabY_ingilizce_YYretmeni_Yzel_alan_yeterlikleri_ilkY
Yretim_parYa_5.pdf (Date of access: 21.05.2022).

- MoNE (2020). *İngilizce dersi öğretmen rehber kitapçığı: Okul ve sınıf tabanlı değerlendirmeye dayalı öğretmen kapasitesinin güçlendirilmesi [Strengthening teacher capacity based on school- and classroom-based assessment: English lesson teacher's guide booklet]*. Ministry of National Education, Ankara. Retrieved from <https://odsgm.meb.gov.tr/www/okul-ve-sinif-tabanlı-değerlendirmeye-dayalı-ogretmen-kapasitesinin-guclendirilmesi-calismasi/icerik/554>.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston: Pearson Education.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380. DOI: 10.1177/0265532213480336
- O'Sullivan, B. (2012) Assessing speaking. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 234-246). USA: Cambridge University.
- Ölmezer-Öztürk, E., & Aydın, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of language assessment knowledge scale (LAKS). *Language Testing in Asia*, 8(20). <https://doi.org/10.1186/s40468-018-0075-2>
- Ölmezer-Öztürk, E., & Aydın, B. (2019a). Investigating language assessment knowledge of EFL teachers. *Hacettepe University Journal of Education*, 34(3), 602-620. doi: 10.16986/HUJE.2018043465
- Ölmezer-Öztürk, E., & Aydın, B. (2019b). Voices of EFL teachers as assessors: Their opinions and needs regarding language assessment. *Journal of Qualitative Research in Education*, 7(1), 373-390. doi:10.14689/issn.2148-2624.1.7c1s.17m
- Önalın, O., & Gürsoy, E. (2020). Primary and secondary level EFL teachers' use of assessment and assessment results in Turkey. *i-managers' Journal on English Language Teaching*, 10(3).

- Ornstein, A. C. (1993). Norm-referenced and criterion-referenced tests: An overview. *NASSP Bulletin*, 77(555), 28-39. <https://doi.org/10.1177/019263659307755505>
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6(1), 1-13. <https://doi.org/10.1177/026553228900600103>
- O'Sullivan, B. (2013). Assessing Speaking. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1). London: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118411360.wbcla084>
- Öz, H. (2014). Turkish teachers' practices of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research*, 5(4), 775-785.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402. DOI: 10.1177/0265532213480337
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-24.
- Poehner, M. E. (2013). Dynamic assessment in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 2). London: John Wiley & Sons, Inc.
- Poehner, M. E., Davin, K. J., & Lantolf, J. P. (2017). Dynamic assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language education: Language testing and assessment* (3rd ed., pp. 243-256). Cham: Springer. https://doi.org/10.1007/978-3-319-02261-1_18
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4-11. DOI: 10.1080/00405840802577536
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273. DOI: 10.1080/08878730.2011.605048
- Popham, W. J. (2018). *Classroom assessment: What teachers need to know* (8th ed.). Boston: Pearson.

- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258. doi:10.1191=0265532204lt283ed
- Rogier, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Language Teaching Forum*, 3, 2-13.
- 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. DOI: 10.1177/0265532213480128
- Sevimel-Şahin, A., & Subaşı, G. (2021). Exploring foreign language assessment literacy training needs of pre-service English language teachers. *International Online Journal of Education and Teaching (IOJET)*, 8(4), 2783-2802.
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117. <https://doi.org/10.1023/A:1006985528729>
- Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing*, 2(1), 31-40. <https://doi.org/10.1177/026553228500200104>
- Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, 18(1), 1-22.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23-27.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765. <https://doi.org/10.1177/003172170208301010>
- Tanyer, S., & Susoy, Z. (2018). Assessment of young language learners: Perceptions and practices of Turkish EFL pre-service and in-service teachers. *The Journal of Language Teaching and Learning*, 8(2), 127-149.
- Tavassoli, K., & Farhady, H. (2018). Assessment knowledge needs of EFL teachers. *Teaching English Language*, 12(2).

- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. doi:10.1017/S0267190509090035
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. doi: 10.1177/0265532213480338
- The American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), & the National Education Association (NEA). (1990). *Standards for teacher competence in the educational assessment of students*. Retrieved from <https://buros.org/standards-teacher-competence-educational-assessment-students>
- Thorndike, R. M., & Thorndike-Christ, T. (2014). *Measurement and evaluation in psychology and education* (8th ed.). Pearson Education Limited.
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20-27. DOI: 10.1080/00405840802577569
- Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language education: Language testing and assessment* (3rd ed., pp. 359-372). Cham: Springer.
- Tsagari, D., & Vogt, T. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Turner, J. (1998). Assessing Speaking. *Annual Review of Applied Linguistics*, 18, 192-207. doi:10.1017/S0267190500003548
- Tuzcu-Eken, D. (2016). Testing in English classes in vocational schools of higher education. *Electronic Journal of Vocational Colleges*, 6(1), 38-47. Retrieved from <https://dergipark.org.tr/tr/pub/ejovoc/issue/36628/416986>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. DOI:10.1080/15434303.2014.960046
- Wagner, E. (2013). Assessing listening. In A. Kunnan (Ed.), *The companion to language assessment* (Vol. 1, pp. 47-63). Oxford, England: Wiley-Blackwell.

- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69. <https://doi.org/10.1177/026553229301000103>
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research context and methods* (pp. 19–36). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University.
- Weigle, S. C. (2012). Assessment of writing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Malden, MA: Wiley-Blackwell.
- William, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, 21(2), 15-20.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <http://dx.doi.org/10.1016/j.tate.2016.05.010>
- Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Yan, X., Zhang, C., & Fan, J. J. (2018). “Assessment knowledge is important, but ...”: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168. <https://doi.org/10.1016/j.system.2018.03.003>
- Yastıbaş, A. E., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessments. *Journal of Language and Linguistic Studies*, 14(1), 178-193.



APPENDIX



Appendix A. Ethics Committee Approval

Evrak Tarih ve Sayısı: 22.02.2021-E.33889



**T.C.
GAZİ ÜNİVERSİTESİ
Etik Komisyonu**

Sayı : E-77082166-302.08.01-33889
Konu : Bilimsel ve Eğitim Amaçlı

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 28.01.2021 tarihli ve 80287700-302.08.01- 16512 sayılı yazı.

İlgi yazınız ile göndermiş olduğunuz, Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı, İngiliz Dili Eğitimi Bilim Dalı **Yüksek Lisans Öğrencisi Gamze SATGUN'un, Doç.Dr.Asuman AŞIK'ın** danışmanlığında yürüttüğü *"An Investigation of EFL Teachers' Language Assessment Knowledge and Practice: The Case of K-12 School in Turkey"* adlı tez çalışması ile ilgili konu Komisyonumuzun **16.02.2021** tarih ve **03** sayılı toplantısında görüşülmüş olup,

İlgilinin çalışmasının, yapılması planlanan yerlerden izin alınması koşuluyla yapılmasında etik açıdan bir sakınca bulunmadığına oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi ve gereğini rica ederim.

Araştırma Kod No: 2021 - 207

Prof. Dr. İsmail KARAKAYA
Komisyon Başkanı

Ek:1 Liste

**GAZİ ÜNİVERSİTESİ
ETİK KOMİSYONU KATILIM LİSTESİ**

TOPLANTI TARİHİ : 16.02.2021		TOPLANTI SAYISI : 03	
ADI – SOYADI		İMZA	
Prof. Dr. İsmail KARAKAYA Başkan			
Prof.Dr.Kemal ÖZTEMEL Başkan Yardımcısı			
Prof.Dr.C.Haluk BODUR			
Prof.Dr.Seçil ÖZKAN			
Prof.Dr.Cevriye TEMEL GENCER			
Prof.Dr.İsmet YÜKSEL			
Prof.Dr.Aymelek GÖNENÇ			
Prof.Dr.Gülay BAYRAMOĞLU			
Prof.Dr.Makbule GEZMEN KARADAĞ			
Prof.Dr.Zehra GÖÇMEN BAYKARA			
Prof.Dr.İlyas OKUR			
Doç.Dr.Nihan KAFA			
Doç.Dr.Melek Gülşah ŞAHİN			

Appendix B. Permission from the Ministry of National Education

Evrak Tarih ve Sayısı: 31.12.2021-E.252626

Evrak Tarih ve Sayısı: 30.12.2021-E.252501



T.C.
MİLLÎ EĞİTİM BAKANLIĞI
Strateji Geliştirme Başkanlığı

Sayı : E-49614598-605.01-40118068
Konu : Araştırma Uygulama İzin Talebi

29.12.2021

DAĞITIM YERLERİNE

İlgi: a) Gazi Üniversitesi Rektörlüğü'nün 15/11/2021 tarihli ve E-17311665-044-215250 sayılı yazısı.
b) Millî Eğitim Bakanlığı'nın 21/01/2020 tarihli ve 2020/2 Nolu Araştırma Uygulama İzinleri Genelgesi.

İlgi (a) yazı ile Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı Yüksek Lisans Programı öğrencisi Gamze SATGUN'un "An Investigation of EFL Teachers' Language Assessment Knowledge and Practice: The Case of K-12 Schools in Turkey" konulu çalışmasına veri sağlamak amacıyla anket çalışması yapma izin talebine ilişkin ilgi yazı ve ekleri Bakanlığımız tarafından incelenmiştir.

Bakanlığımıza bağlı resmi/özel okul ve kurumlarda öğrenci, öğretmen ve okul yöneticilerinin katılımıyla yapılması planlanan uygulamanın covid-19 tedbirlerine uyulması ve denetimi il/ilçe millî eğitim müdürlükleri ve okul/kurum idaresinde olmak üzere, kurum faaliyetlerini aksatmadan, gönüllülük esasına göre; onaylı bir örneği Bakanlığımızda muhafaza edilen ve uygulama sırasında da mühürlü ve imzalı örnekten çoğaltılan, veri toplama araçlarının <https://forms.gle/yE6sbtceDf9usTfP8> adresinden online olarak uygulanmasına ilgi (b) Genelge doğrultusunda izin verilmiştir.

Gereğini bilgilerinize rica ederim.

Mehmet Fatih LEBLEBİCİ
Bakan a.
Başkan

Ek: Onaylı Veri Toplama Araçları (23 Sayfa)

Dağıtım:
Gereği:
B Planı

Bilgi:
Gazi Üniversitesi Rektörlüğüne

Appendix C. Permission E-mail from the Authors of LAKS



gamze satgun <@gmail.com>

29 Mar 2021 Pzt 17:04



Alıcı: i

Sayın Elçin ÖLMEZER-ÖZTÜRK ve Sayın Belgin AYDIN,

Gazi Üniversitesi İngiliz Dili Eğitimi Anabilim Dalı yüksek lisans öğrencisiyim. Doç. Dr. Asuman AŞIK danışmanlığında yürütmekte olduğumuz yüksek lisans tezimde, MEB'e bağlı okullarda çalışmakta olan İngilizce öğretmenlerinin yabancı dilde ölçme değerlendirme bilgi düzeylerinin ve uygulamalarının incelenmesini amaçlamaktayız. Bu doğrultuda, geliştirmiş olduğunuz 'Language Assessment Knowledge Scale' adlı ölçeği kullanmak için izninizi istiyoruz.

Şimdiden teşekkür ederim. İyi çalışmalar diliyorum.

Gamze SATGUN



Belgin AYDIN

29 Mar 2021 Pzt 23:03



Sevgili Gamze, bahsettiğin scale Elçin Hoca'nın doktora tezi kapsamında üretimiştir. Dolayısıyla izin hakkı tamamen ona ait...



Elçin ÖLMEZER ÖZTÜRK <el@lu.edu.tr>

30 Mar 2021 Sal 08:39



Alıcı: Belgin, ben

Merhabalar,

Belgin hocam her zamanki nezaketiye beni ihya etmiş,sagolsun:)

Belgin hocam sorumluluğu bana vermiş. Hocamla çok emek verdığımız olceğimizi tabii ki kulanabirlirsiniz.biz çalışırken çok keyif aldık.

Size de aynı keyifle çok başarılar dilerim...

İyi çalışmalar

Appendix D. Language Assessment Knowledge Scale (LAKS)

Ölmezer-Öztürk, E., & Aydın, B. (2018). Towards measuring language teachers' assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS), *Language Testing in Asia*, 8(20). <https://doi.org/10.1186/s40468-018-0075->

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 20 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. What is the type of the institution you work at?
a) a state school b) a private school
6. Where are you working right now?
a) primary school b) secondary school c) high school
7. Write the city you work in
_____.
8. Did you have a separate testing/assessment course in pre-service education?
a) yes b) no
9. Have you attended any professional development programs/courses/training on language assessment?
a) yes b) no
10. 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
11. Please write your institutional 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.			
2. When asking several questions about a reading text, all the questions are independent of each other.			
3. Cloze test is used for assessing the main idea of the text.			
4. In a reading exam, using a text learners have encountered before is not a problem.			
5. One reading text is enough to be included in a reading exam.			
6. The language of the questions is simpler than the text itself.			
7. Errors of spelling are penalized while scoring.			
8. Taking vocabulary difficulty into 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.			
10. The more items a reading text is followed, the more reliable it becomes.			
11. Using the same words in the correct option as in the text is not a problem.			
12. Simplification of reading texts is avoided.			
13. Reading texts in a reading exam include various genres (essay, article, etc.).			
14. In top-down approach, assessment is on overall comprehension of the reading text.			
15. Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.			
ASSESSING LISTENING			
16. Using reading texts for listening purposes poses a problem.			
17. Including redundancy (e.g. what I mean to say is that) in a listening text poses a problem.			
18. Any type of listening text is used for note-taking.			
19. Spelling errors are ignored in scoring the dictation.			
20. Errors of grammar or spelling are penalized while scoring.			
21. A listening cloze test is a way of selective listening.			
22. Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative testing.			
23. Scoring in note-taking is straightforward.			
24. In discrete-point testing, comprehension is at the literal/of local level.			
25. Using dictation diagnostically in assessing listening skills does not pose a problem.			
26. Giving learners a transcript of the listening text is a valid way of assessing listening skills.			
27. Dictation is a kind of discrete-point testing.			
28. Inference questions based on intelligence are avoided in listening tests.			
29. Asking learners to listen to names or numbers is called intensive listening.			
30. In selective listening, learners are expected to look for certain information.			

ASSESSING WRITING

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

ASSESSING SPEAKING

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

Appendix E. Interview Consent Form

BİREYSEL GÖRÜŞME FORMU

Bu form K-12 (ilkokul, ortaokul, lise) düzeyinde görev yapmakta olan İngilizce öğretmenlerinin yabancı dilde ölçme değerlendirme uygulamalarının tespitine yönelik hazırlanmış olup Gazi Üniversitesi İngiliz Dili Eğitimi Bölümü'nde bir yüksek lisans tezi için kullanılacaktır. Görüşme formu iki kısımdan oluşup ilk kısım katılımcıların demografik bilgileriyle ilgili soruları, ikinci kısım ise yabancı dil eğitiminde ölçme değerlendirme uygulamaları ile ilgili soruları içermektedir. Görüşme sırasında sorulacak sorular İngilizce öğretmenlerinin yabancı dilde ölçme değerlendirme uygulamalarının anlaşılması bakımından önem taşımaktadır. Görüşme sorularına vereceğiniz cevapların gerçekçi ve içten olması araştırmanın geçerliği ve güvenilirliği açısından büyük öneme sahiptir. Görüşme yaklaşık 30 dakika sürecek ve çevrimiçi ortamlarda gerçekleştirilecektir. Kişisel bilgileriniz kesinlikle gizli tutulacak olup gerek araştırma süresince gerek yayımlandığında herhangi bir kişi/kuruluş ile paylaşılmayacaktır. Çalışmaya katılım gönüllülük esasına dayalı olup görüşme sırasında herhangi bir sebep göstermeden çekilebilirsiniz. Görüşme sırasında araştırmacı tarafından ses kaydı alınacak olup görüşme sonunda ses kaydını dinlemeniz ve onaylamanız için tarafınıza gönderilecektir. Görüşmeye katılmadan önce ya da sonrasında aklınıza gelebilecek soruları araştırmacıya istediğiniz zaman sorabilirsiniz. Araştırma hakkında daha sonra bilgi almak isterseniz aşağıdaki e-posta adresinden araştırmacıya ulaşabilirsiniz. Araştırmaya katılmayı tercih ediyorsanız lütfen aşağıya imza atınız.

Katkılarınız için şimdiden teşekkürler.

İmza:

Tarih:

Yüksek Lisans Öğrencisi Adı-Soyadı: Gamze Satgun

E-mail: ξ

Danışmanı: Doç. Dr. Asuman Aşık

Appendix F. Semi-structured Interview Questions

TÜRKÇE	ENGLISH
1. KISIM-KİŞİSEL BİLGİLER	PART 1-DEMOGRAPHIC INFORMATION
1. Cinsiyetiniz?	1. Gender?
2. Yaşınız?	2. Age?
3. Kaç yıldır İngilizce öğretmenliği yapıyorsunuz?	3. How long have you been working as an EFL teacher?
4. Şu an hangi okulda, kademedeki ve sınıflarla çalışıyorsunuz?	4. Which school, levels, and grades are you working with right now?
5. Daha önce hangi okulda, kademedeki ve sınıflarla çalıştınız?	5. Which schools, level, and grades did you work before?
6. Şu an çalışmakta olduğunuz okulda sınıflarınızda ortalama kaç öğrenci var?	6. How many students are there in a classroom on average?
7. Şu an çalışmakta olduğunuz okulda haftalık kaç ders saatiniz var?	7. How many teaching hours do you have in a week at your current workplace?
8. En son mezun olduğunuz eğitim derecesi nedir?	8. What is your educational level?
9. Hangi lisans programından mezun oldunuz?	9. Which BA program did you graduate from?
10. Hizmet öncesi eğitiminiz süresince ölçme değerlendirme üzerine eğitim/kurs/ders vb. aldınız mı?	10. Did you take a course on language assessment in pre-service education?
11. Hizmet içinde ölçme değerlendirme üzerine kurs/seminer vb. eğitimlere katıldınız mı?	11. Have you attended any in-service trainings programs regarding language assessment?
12. E-mail adresiniz:	12. Your e-mail address:
2. KISIM- GÖRÜŞME SORULARI	PART 2- INTERVIEW QUESTIONS
1. Ölçme değerlendirme uygulamalarınızı nasıl tanımlarsınız?	1. How would you describe your assessment practices?
2. Ölçme değerlendirmeyi ne zaman kullanırsınız?	2. When do you assess your students?
3. Ölçme değerlendirmeyi ne amaçla kullanırsınız?	3. Why do you assess your students?
4. Genellikle hangi becerileri (okuma, yazma, dinleme, konuşma, kelime, dilbilgisi, telaffuz) değerlendirirsiniz? Nasıl?	4. Which language skills do you typically assess? And how?
5. Hangi ölçme değerlendirme yöntemlerini benimsiyorsunuz? Akran ve öz değerlendirme kullanıyor musunuz? Nasıl?	5. Which assessment methods do you adopt? Do you employ peer- and self-assessment? If yes, how?
6. Ne tür ölçme değerlendirme araçlarını kullanırsınız?	6. Which assessment tools do you use?
7. Ölçme değerlendirme sonuçlarını nasıl kullanır, değerlendirirsiniz?	7. How do you make use of assessment results?
8. Ölçme değerlendirme uygulamalarından sonra öğrenciye ne zaman ve nasıl dönüt verirsiniz?	8. When do you give feedback to your learners?
9. Ölçme değerlendirme uygulamalarınızı nasıl planlarsınız? Bu süreçte nelerden veya kimlerden yardım alırsınız? Karar verme sürecinizdeki etkenler nelerdir? Neleri dikkate alırsınız? Belirtke tablosu kullanır mısınız?	9. How do you plan your assessment? Do you get any help from other sources or people? What are the factors in your decision-making process? What do you consider? Do you use a table of specification?

10. Ölçme değerlendirilmede geçerlik, güvenirlik, ekonomiklik, kullanılabilirlik gibi nitelikleri dikkate alıyor musunuz? Bu nitelikleri sağlayabilmek için neler yapıyorsunuz?	10. Do you take assessment principles of validity, reliability, practicality in to consideration? How do you ensure these principles?
11. Ölçme değerlendirme sürecinde öğrencinin yeri nedir? Öğrencilere değerlendirmeyle ilgili danışıyor musunuz? Öğrencileri değerlendirmeyle ilgili bilgilendiriyor musunuz? Hangi konularda öğrenciye danışıyor veya öğrenciyi bilgilendiriyorsunuz?	11. What is the role of the learners in the assessment process? Do you inform or consult to your learners? About what?
12. Ölçme değerlendirme uygulamalarınızda ne gibi problemlerle karşılaşıyorsunuz? Sebepleri nelerdir?	12. What are the problems you face in your assessment practices? What are the reasons?
13. Ölçme değerlendirmeyle ilgili mesleki gelişiminizi nasıl sağlıyorsunuz? (Seminer, kitap, makale vb.) Bu alanda katılabileceğiniz hizmet içi eğitimler var mı? Varsa, bunlar nelerdir? Yoksa, olmasını ister misiniz? Hangi konuda?	13. How do you continue your professional development? Are there any in-service trainings you can attend in this subject? Which area of assessment do you want to have a training?
14. Yabancı dilde ölçme değerlendirme standartlarıyla ilgili K-12 öğretmenlerine sağlanan bir kaynak bulunuyor mu? Varsa belirtir misiniz? Yoksa, böyle bir kaynağa ihtiyaç olduğunu düşünüyor musunuz?	14. Are there a source provided for teachers regarding the standards of language assessment? If there are, explain. If there are not, do you think you need a source in this regard?
Aşağıda belirtilen sorular 6-12 yaş aralığındaki çocuklara İngilizce öğreten öğretmenlere yöneliktir.	The following questions are for teachers who teach English to learners between the ages 6 to 12.
15. Çocuklara yabancı dil öğretiminde ölçme değerlendirme konusunda daha önce hizmet öncesi veya hizmet içi eğitim aldınız mı?	15. Have you taken a course or a training regarding language assessment of young learners in pre- and in-service?
16. Yabancı dil olarak İngilizce öğrenen çocukları nasıl değerlendiriyorsunuz? Neleri dikkate alıyorsunuz?	16. How do you assess young learners? What do you take into consideration?



GAZİLİ OLMAK AYRICALIKTIR..