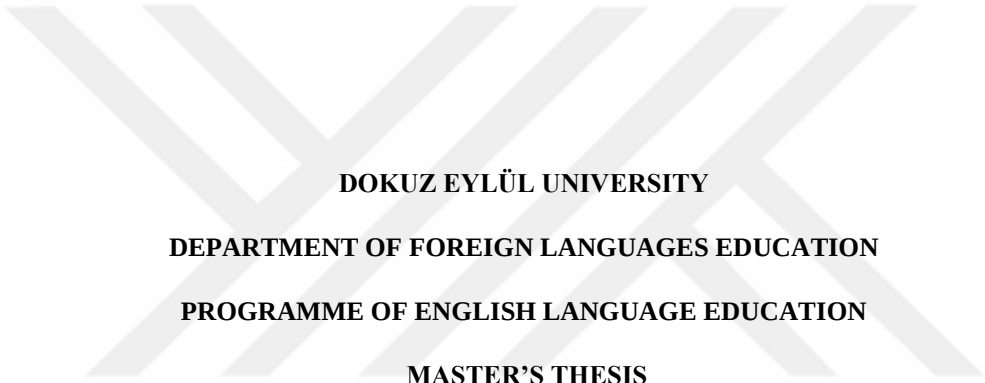




DOKUZ EYLÜL UNIVERSITY
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
PROGRAMME OF ENGLISH LANGUAGE EDUCATION
MASTER'S THESIS

**INVESTIGATING EFL INSTRUCTORS' LANGUAGE
ASSESSMENT KNOWLEDGE**

Özge DERETARLASI

İzmir
2020

T.C.
DOKUZ EYLÜL UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
PROGRAMME OF ENGLISH LANGUAGE EDUCATION
MASTER’S THESIS

**INVESTIGATING EFL INSTRUCTORS’ LANGUAGE
ASSESSMENT KNOWLEDGE**

Özge DERETARLASI

Supervisor

Assoc. Prof. Dr. Yeşim BEKTAŞ ÇETİNKAYA

İZMİR

2020

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Yüksek lisans tezi olarak sunduğum “Investigating EFL Instructors’ Language Assessment Knowledge” adlı çalışmanın içerdiği fikri izinsiz başka bir yerden almadığımı; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında ve bölümlerinin yazımında bilimsel etik ilke ve kurallara uygun davrandığımı, tez yazım kurallarına uygun olarak hazırlanan bu çalışmada kullanılan her türlü kaynağa eksiksiz atıf yaptığımı ve bu kaynaklara kaynakçada yer verdiğimi, ayrıca bu çalışmanın Dokuz Eylül Üniversitesi tarafından kullanılan bilimsel intihal tespit programıyla tarandığını ve *intihal içermediğini* beyan ederim. Herhangi bir zamanda aksinin ortaya çıkması durumunda her türlü yasal sonuca razı olduğumu bildiririm.

20.07/2020



Özge DERETARLASI

Tarih: 11/08/2020

Tez Başlığı:

'Investigating EFL Instructors' Language Assessment Knowledge'

Yukarıda başlığı belirtilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 83 sayfalık kısmına ilişkin, 08/08/2020 tarihinde **tez danışmanım tarafından** Dokuz Eylül Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı'nın sağladığı İntihal Tespit Programından (**Turnitin-Tez İntihal Analiz Programı**) aşağıda belirtilen **filtreleme tiplerinden biri** (uygun olanı işaretleyiniz) uygulanarak alınmış olan orijinallik raporuna göre, tezimin **benzerlik oranı % 11'dir**.

- <http://www.kutuphane.deu.edu.tr/turnitin-tez-intihal-analiz-programi/> adresindeki Tez İntihal Analiz Programı Kullanım Kılavuzunu okudum ☒

Filtreleme Tipi 1(Maksimum %15) ☒

Filtreleme Tipi 2(Maksimum %30) ☐

☒ Kabul/Onay ve Bildirim sayfaları hariç, ☒ Kaynakça hariç, ☒ Alıntılar dâhil, ☒ Altı (6) kelimedenden daha az örtüşme içeren metin kısımları hariç.	☐ Kabul/Onay ve Bildirim sayfaları hariç, ☐ Kaynakça dâhil, ☐ Alıntılar dâhil.
EK 1- İntihal Tespit Programı Raporu İLK SAYFA Çıktısı. ☒ EK 2- İntihal Tespit Programı Raporu (Tümü) Cd İçinde. ☐	

Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Uygulama Esasları'nı inceledim ve yukarıda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Adı Soyadı : Özge Deretarlası
Öğrenci No : 2018950185
Anabilim Dalı : Yabancı Diller Eğitimi
Programı : İngilizce Öğretmenliği
Statüsü : Yüksek Lisans ☒

Doktora ☐

ÖĞRENCİ

11.08.2020

Özge Deretarlası
(Adı Soyadı, İmza, Tarih)

DANIŞMAN

11.08.2020

Doç. Dr. Yeşim Bektaş Çetinkaya
(Unvan, Adı Soyadı, İmza, Tarih)

Açıklamalar

1: Bu formu teslim etmeden önce sizden istenen bilgileri uygun kutucuğu (☐) işaretleyerek doldurunuz.

Kullanıcı şifre vb. konusunda sorun yaşanması durumunda Üniversitemiz Merkez Kütüphanesinde bulunan Turnitin yetkilisine (Ali Taş Tel: +90 (232) 3018026 veya ali.tas@deu.edu.tr) başvurunuz.

2: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu" formu tezin ciltlenmiş ve elektronik nüshalarının içerisinde ekler kısmında yer alır.

3: Tez savunmasında düzeltme alınması durumunda bu form güncellenerek yeniden hazırlanır.

4: Turnitin-Tez İntihal Analiz Programına yükleme yapılırken Dosya Başlığı (document title) olarak **tez başlığının tamamı**, Yazar Adı (author's first name) olarak **öğrencinin adı**, Yazar Soyadı (author's last name) olarak **öğrencinin soyadı** bilgisini yazınız.

ACKNOWLEDGEMENTS

I would like to pay my special regards to my supervisor Assoc. Prof. Dr. Yeşim BEKTAŞ ÇETİNKAYA for her invaluable guidance and assistance she provided during my study. Without her persistent help, the goal of this study would not have been achieved.

I would like acknowledge the support of my family, my mother, Hatice ARAS; and my brother, Ali DERETARLASI. They kept me motivated to go on throughout this long journey. This project would not have been possible without their love and support.

I also wish to thank all the institutions and people whose assistance was a milestone in the completion of this study.

TABLE OF CONTENTS

ÖZET	vii
ABSTRACT	ix
CHAPTER 1	1
INTRODUCTION	1
1.1. Background of the study	1
1.2. Research Questions	6
1.3. Methodology	6
CHAPTER 2	8
THEORETICAL FRAMEWORK AND LITERATURE REVIEW	8
2.1. Assessment and Testing in General	8
2.1.1. Validity	10
2.1.1.1. <i>Face validity</i>	11
2.1.1.2. <i>Content validity</i>	11
2.1.1.3. <i>Construct validity</i>	11
2.1.1.4. <i>Criterion validity</i>	12
2.1.2. Reliability	12
2.2. Types of Assessment	13
2.2.1. Informal and formal assessment	13
2.2.2. Formative and summative assessment	13
2.2.3. Subjective and objective assessment	14
2.2.4. Norm-referenced and criterion-referenced assessment	14
2.2.5. Discrete-point and integrative assessment	15
2.3. Classroom Assessment	15
2.4. Language Testing and Assessment	18
2.4.1. Types of language tests	19
2.4.2. Language test design cycle	20
2.4.3. Assessing language skills	21

2.4.3.1.	<i>Assessing reading skill</i>	22
2.4.3.2.	<i>Assessing listening skill</i>	24
2.4.3.3.	<i>Assessing writing skill</i>	25
2.4.3.4.	<i>Assessing speaking skill</i>	26
2.5.	Language Assessment in Turkey	28
2.5.1.	Studies on language assessment in Turkey	28
2.6.	EFL Instructors' Language Assessment Knowledge	30
2.6.1.	Studies conducted on EFL teachers' language assessment knowledge	35
2.6.2.	Language assessment knowledge studies conducted in Turkish context	39
CHAPTER 3		42
METHODOLOGY		42
3.1.	Research Design	42
3.2.	Research Sites and Participant Selection	43
3.3.	Instruments	45
3.4.	Data Collection Procedures	47
3.5.	Data Analysis	47
3.5.1.	Quantitative research analysis	47
3.5.2.	Qualitative research analysis	48
CHAPTER 4		50
RESULTS		50
4.1.	What is the EFL instructors' current level of language assessment knowledge in terms of general and skill-based assessment?	50
4.2.	Does LAK level change according to gender and assessment unit experience of EFL instructors?	56
4.3.	Is there a relationship between EFL instructors' current level of language assessment knowledge and their year of work experience?	59
4.4.	Which practices, suggestions and beliefs of the EFL instructors who are highly knowledgeable in language assessment possess?	59

4.4.1.	Background and language assessment practices	60
4.4.2.	Beliefs	64
4.4.3.	Suggestions	65
CHAPTER 5		68
DISCUSSION AND CONCLUSION		68
5.1.	Discussion	68
5.1.1.	The language assessment knowledge level of EFL instructors.....	69
5.1.2.	The impact of gender and assessment unit experience on language assessment knowledge	71
5.1.3.	The impact of years of teaching experience on language assessment knowledge	72
5.1.4.	Language assessment practices, beliefs and suggestions of the EFL instructors with high scores in LAKS	73
5.2.	Conclusion	74
5.3.	Suggestions	75
5.4.	Limitations	78
REFERENCES		79
APPENDICES		90
APPENDIX 1. STUDENT’S CURRICULUM VITAE		90
APPENDIX 2. OFFICIAL NOTICE(S) OF ETHICAL PERMISSON.....		91
APPENDIX 3. PERMISSON E-MAIL(S) FROM THE AUTHORS OF LAKS		95
APPENDIX 4. LANGUAGE ASSESSMENT KNOWLEDGE SCALE.....		96

LIST OF TABLES

Table 1: Demographic information about the participants.....	44
Table 2: Relevant characteristics of the EFL instructors who participated in the interviews.....	45
Table 3: Reliability scores of the LAKS.....	46
Table 4: The mean of total scores of the participants.....	50
Table 5: The mean of each sub-section of the LAKS.....	51
Table 6: Descriptive results of the LAKS items.....	52
Table 7: The number of female/male participants and participants with/without assessment office experience.....	56
Table 8: The mean scores of female/male participants and participants with/without assessment office experience.....	57
Table 9: Tests of between-subjects effects.....	58
Table 10: Correlation between years of work experience and mean of scores.....	59

LIST OF FIGURES

Figure 1: Components of writing assessment.....	25
Figure 2: Stages, practices and products of speaking assessment.....	27
Figure 3: Language assessment literacy dimensions.....	31
Figure 4: Sequential explanatory design.....	42



ÖZET

YABANCI DİL OLARAK İNGİLİZCE ÖĞRETİM GÖREVLİLERİNİN DİL ÖLÇME-DEĞERLENDİRME BİLGİSİNİN ARAŞTIRILMASI

Ölçme-değerlendirme her öğrenme ortamının ayrılmaz bir parçasıdır. Yabancı dil olarak İngilizce öğretmenlerinin öğrencilerinin öğrenme gelişimlerinden emin olmak için sınıflarındaki öğrenmenin ölçülmesi için yeterli bilgi ve beceri düzeylerine sahip olmalıdırlar. Buna bakarak, İngilizce öğretmenleri için farklı amaçlar için birçok çeşitlilikte uygun ölçme araçlarını uyarlama, geliştirme ve uygulamaları elzem bir gerekliliktir. Dil yeterliği ölçme-değerlendirme bilgisi her yabancı dil öğretmenin profesyonel gelişimin ayrılmaz bir parçasıdır. Ölçme-değerlendirme bilgisinin günümüzde dil öğretmenlerinin profesyonel hayatındaki bu rolü düşünülerek, bu çalışma Türkiye'deki devlet üniversitelerinde çalışan yabancı dil olarak İngilizce öğretim görevlilerinin genel ve beceri temelli dil ölçme-değerlendirme bilgi düzeylerinin araştırılmasını amaçlamaktadır. Çalışmanın bir diğer amacı da cinsiyet, mesleki deneyim ve ölçme-değerlendirme birimi deneyiminin İngilizce öğretim görevlilerinin ölçme-değerlendirme bilgi düzeylerine etkilerini araştırmaktır. Çalışma ayrıca yabancı dilin ölçme-değerlendirmesi alanında bilgili İngilizce öğretim görevlilerinin yabancı dil ölçme çalışmaları, inançları ve önerilerini saptamayı hedeflemektedir.

Bu çalışmada nicel ve nitel verinin toplanması amacıyla bir karma yöntem araştırma deseni kullanılmıştır. Nicel veri toplama sürecinde Ölmezer-Öztürk ve Aydın (2018b) tarafından geliştirilen bir ölçek (LAKS) kullanılmıştır. Nitel araştırma için, araştırmacı tarafından geliştirilen 3 açık uçlu soruyla birlikte yarı-yapılandırılmış görüşmeler gerçekleştirilmiştir. Çalışmaya 4 devlet üniversitesinde yabancı diller yüksekokullarında çalışan 114 İngilizce öğretim görevlisi katılmıştır. Buna ek olarak, ölçekteki maddelerin en az %70'ini doğru cevaplayan 8 İngilizce öğretim görevlisi ile görüşmeler gerçekleştirilmiştir. Nicel veri analizinde, betimleyici istatistik, iki yönlü ANOVA ve Pearson korelasyon analizi testleri kullanılmıştır. Yarı-yapılandırılmış görüşmeler aracılığıyla toplanan veri kodlama ve metin analizi yöntemleriyle analiz edilmiştir.

Çalışmanın sonuçları Türkiye'deki devlet üniversitelerine bağlı yabancı diller yüksek okullarında çalışan İngilizce öğretim görevlilerinin genel ve beceri temelli olarak dil ölçme-değerlendirmede sınırlı bilgiye sahip olduklarını göstermektedir. Çalışma sonuçlarına göre, bir ölçme-değerlendirme biriminin üyesi olarak deneyim kazanmış olmanın İngilizce öğretim görevlilerinin ölçme-değerlendirme bilgi düzeylerine istatistiksel olarak önemli bir etkisi

olduğu saptanmıştır. Cinsiyetin öğretim görevlilerinin ölçme-değerlendirme bilgi düzeylerine istatistiksel olarak önemli bir etkiye sahip olmadığını saptanmıştır. Mesleki deneyim ile ölçme-değerlendirme bilgi düzeyi arasında korelasyon bulunmamıştır. Nitel veri analizi sonuçlarına göre, İngilizce öğretim görevlilerinin örnek ölçme araçlarını inceleyerek, ölçme araçlarına benzer sınıf materyalleri geliştirerek ve daha deneyimli bir soru yazarından veya meslektaşından geri bildirimler alarak dil ölçme-değerlendirme alanındaki yeterliklerini geliştirmişlerdir. Ek olarak, yabancı dil ölçme-değerlendirme bilgisinin en büyük avantajı olarak daha verimli bir öğrenme ortamının oluşturulması ileri sürülmüştür. Ayrıca, katılımcılar İngilizce öğretim görevlilerinin ölçme-değerlendirme alanında bilgi düzeylerini geliştirmeye yönelik olarak teorik ve pratik içeriğin yer aldığı eğitimleri ve yabancı dilde ölçme değerlendirmeyle ilgili gelişmelerin takibi için desteklenmelerini önermektedirler. Nicel ve nitel veri analizi sonuçları değerlendirilerek ilgili kurumlar, yabancı dil öğretim görevlileri ve gelecekteki akademik çalışmalar için öneriler sunulmuştur.

Anahtar Kelimeler: Öğretmen eğitimi, yabancı dil ölçme-değerlendirme, ölçme-değerlendirme okur-yazarlığı, dil ölçme-değerlendirme bilgisi

ABSTRACT

INVESTIGATING EFL INSTRUCTORS' LANGUAGE ASSESSMENT KNOWLEDGE

Assessment is an indispensable part of every learning context. To be able to ensure the improvement of their learners, EFL teachers need to possess adequate knowledge and skills in assessing learning in their classroom. Regarding this, it is a crucial requirement for English language teachers to adapt, construct and implement a range of proper assessment instruments for various decisions. Language assessment knowledge is, therefore, an integral facet of every language teacher's professional development. Given this role of assessment knowledge in today's foreign language teachers' professional life, this study aims to investigate knowledge level of in-service EFL instructors in general and skill-based language assessment at state universities in Turkey. Another aim of the study was to find out the impact of gender, years of work experience and experience in an assessment unit on English instructors' assessment knowledge levels. The study also aimed to find out language assessment practices, beliefs and suggestions of the EFL instructors who are knowledgeable in language assessment area.

A mixed-methods research design was used to collect both quantitative and qualitative data in this study. In the quantitative data collection process, a scale (LAKS) developed by Ölmezer-Öztürk and Aydın (2018a) was used. For the qualitative research, semi-structured interviews were conducted with 3 open ended questions developed by the researcher. A total of 114 EFL in-service instructors working at schools of foreign languages in four different state universities participated in the study. In addition, 8 EFL instructors who could answer minimum 70 % of the questions in the scale correctly were interviewed. In the analysis of quantitative data, descriptive statistics, two-way ANOVA and Pearson Correlation tests were used. The data collected through semi-structured interviews were analysed with coding and textual analysis methods.

The results of the study indicated that the in-service EFL instructors working at schools of foreign languages in state universities in Turkey have only limited knowledge in general and skill-based language testing and assessment. According to the findings, it was found that having work experience as a member of an assessment unit has a statistically significant effect on the EFL instructors' assessment knowledge. On the other hand, the findings of the study show that there is no statistically significant impact of gender difference

on assessment knowledge levels of the instructors. It is also found that there is no correlation between the years of work experience and assessment knowledge level. The findings of the qualitative data indicated that the English instructors who are knowledgeable in language assessment improved their proficiency in assessment knowledge with the help of analysing sample assessment tools, developing classrooms materials that involve test-like tasks and receiving feedback from peers or a more qualified assessment developer. In addition, the main advantage of assessment knowledge was reported as the creation of better learning atmospheres for the learners by the participants. Also, the participants suggested provision of training on theoretical and practical aspects of language assessment to EFL instructors and encouraging them to follow the new developments and trends in assessment field in order to help them improve their knowledge in assessment. Analysing quantitative and qualitative data, suggestions were presented for the related institutions, EFL instructors and further academic studies.

Keywords: Teacher education, language testing and assessment, assessment literacy, language assessment knowledge

CHAPTER 1

INTRODUCTION

1.1. Background of the study

Students' achievements and progress are an indispensable part of all learning environments. As Brown (2003) puts it, testing and assessment seem as inevitable as next day's sunrise in virtually every sort of educational setting. Teaching and testing are so closely interdependent that it is really difficult to deal with merely one of these fields without constantly considering the impact of the other (Heaton, 1990). When learners enter education systems, they are usually tested individually to identify their comprehension and skills in specific sections of the related subject (Kopriva, 2008). Thus, the learning objectives and administrative goals of a program may necessitate constant and well-planned test development processes (Brown, 1996). Regarding this, it is an essential necessity to adopt, create or adapt test items for an extensive range of decisions involving the placement, diagnostic, proficiency, and achievement decisions (Brown, 1996). Hence, language teachers and school administrations carry an immense burden of presenting the learners with appropriate testing and assessment conditions within the learners' learning atmosphere.

Assessment is directly connected to the concept of justice; hence, it requires to be justified, transparent and challengeable (Davies & Taras, 2018). Embarking such a crucial place in learning context, assessment of learners' performance is one of the most significant responsibilities of educators, since the quality of instruction in learning atmosphere is directly linked to the effectiveness of the implemented assessment (Bayat & Razaei, 2015). Therefore, it is crucial for educators to possess assessment which has been defined as 'all methods and approaches to testing and evaluation whether in research studies or educational contexts' (Kunnan, 2000, p. 1). Assessment can be integrated in instructional context designed to reach the curriculum goals and objectives, and this enables educators not only to ensure that learning is occurring, but also to benefit from the feedback they receive via assessment to support learning (McKay, 2006). Malone (2013) proposed that appropriately implemented assessment activities provide students, teachers and all other related parties with significant and necessary information about learner progress and the extent to which curriculum goals have been met in the learning atmosphere. Where several testing activities are conducted, instruction can be enhanced by learners' awareness of course goals and the important areas in a learning environment.

Assessment's traditional function has been to generate the evidence required to grade and sort (Stiggins, 2017). Assessment tools were not intended to be utilised to gauge the full range of a learner's language proficiency but they only focused on language literacy skills. However, language teachers have become highly frustrated that there seems to be a distinction between language acquisition as defined by tests that are used to exit learners from particular language programs and the academic performance required for students to thrive in mainstream academic classrooms (Kopriva, 2008). Today, many authors (e.g., Hakuta & Beatty, 2000; Kopriva, 2008; Stiggins, 2017) agree that the most appropriate methods for testing course objectives in standardised testing settings are to reflect, as much as possible, instructional and assessment techniques which naturally occur in the classroom. As cognitive psychologists have created frameworks for the scientific interpretation of learning phenomena, there has been growing interest on the small-scale learning processes by which task performance skills are gained and, eventually, on the condensed structures of the specific capabilities of which they are composed. As such structures and processes are extended upon, their mixture better represents the types of complex learning that takes place in courses (Kopriva, 2008). Assessment, then, has now been an essential tool to reflect upon those learning processes that learners have been through in their learning atmospheres.

An assessment task is essentially a tool that allows the test developer to collect evidence. Such evidence is information about the performance of the test taker, whether it is a yes/no answer or an extended participation in a dialogue (Fulcher & Davidson, 2006). In a test we gather pieces of evidence on students' knowledge from plenty of independent responses, we assess and evaluate them in some way, and report a score or letter that we assume means something about the performance of the test taker on the related construct (Bachman, 1990; Fulcher & Davidson, 2006; Stiggins, 2014). Assessment tasks may be developed particularly as tools to enhance learning and to increase the motivation of learners or especially as a way of measuring the learners' performance in a learning environment. In the first scenario, the test is tailored to the instruction that has occurred, while in the second scenario, the instruction is frequently tailored mainly to the test (Heaton, 1990).

Like it is a crucial need for a medical doctor initially to diagnose a patient's disorder, thus it is similarly needed for the educator to diagnose his/her learners' struggles and problems (Heaton, 1990). With such a function, assessment tasks can facilitate learning by their diagnostic characteristics: Tests collect evidence on what every learner has acquired, and they highlight the language areas which require further attention (Madsen, 1983). Thus, the assessment should also help the educator to determine the areas of the language curriculum

which is found too challenging by the majority of the learners. Hence, the teacher is able to evaluate the appropriateness of the instruction as well as the instructional materials and techniques he or she is using (Bachman & Palmer, 1996; Heaton, 1990). Since the test items are supposed to be representative exemplifications of the knowledge and abilities specified in the course content, and all learners are supposed to have been through to grade-level syllabus, it is expected that the assessment scores would be regarded a valid and reliable result of this content (Bachman, 1990; Kopriva, 2008). Tests can be used as a means for clarifying educational goals and, in certain circumstances, for assessing the alignment of these goals and the educational activities and materials based on them to the requirements of learners following the instructional course (Bachman & Palmer, 1996). To sum up, effective assessment procedures provide teachers with reliable results about the existence of learning in the classroom.

Assessment tasks provide a path for learners to perform skills from which one may derive the extent to which a psychological construct is in existence in those test takers. This is the purpose of test tasks in assessment and evaluation, and it justifies the claim that assessment tasks are the ‘backbone’ of measuring mental properties (Osterlind, 2002). Throughout a test task development process, at each step, we gather evidence of how the elements of the related test (i.e., test items, tasks, marking scales) is working, interpret that evidence and revise it accordingly (Cheng & Fox, 2017). Although tests might be useful devices, they are just one among plenty of practices and tasks that assessors can ultimately benefit from to evaluate learners’ performance (Brown, 2003). A test may involve a range of data-collecting tools such as expecting learners to reply to teacher-directed tasks by using interactive white-boards in a group discussion, or carrying out spoken interviews with individual learners or in groups of learners (Popham, 2009). However, no matter what the type of a test task is, whenever a learner responds to a teacher-presented question, makes a comment, or utter a new word or grammatical structure, subconsciously, the language educator assesses the learner's performance in his/her mind (Brown, 2003). In summary, various assessment tasks might find themselves a place in any learning context and, by their nature, generate feedback and evidence for the learners’ performance.

Language assessment tasks can be precious tools for creating data that are related to certain areas in language instruction (Bachman & Palmer, 1996). The preparation and use of language assessment tasks involve an understanding of both the phenomenon of language use with a communicative purpose and language ability and assessment theory (Bachman, 1990). It is clear that assessment is at the centre of the language learning programs and it is an

unavoidable need not only for teachers to see if learning is taking place but also for students to gather feedback on their improvement in their language learning journey. Thus, assessment should be integrated with language teaching so that creating a link in which one benefits and informs the other, like in any other learning environment. However, it is difficult for this reciprocal link to develop when language instructors lack effective background knowledge or professional education in language assessment to create, select and administer language assessment and interpret the results of it (Malone, 2013; Daniel & King, 2010). Hence, the quality of a language learning setting is directly affected by the classroom teachers' expertise and awareness in the fields of testing and assessment of language knowledge and performance.

A well-prepared classroom assessment helps detect the particular parts of challenge faced by the whole group or by a single learner (Heaton, 1990; Stiggins, 2017). Brown (2003) states that a professional language educator never stops to assess learners, whether these testing practices are incidental or on purpose. For an educator to feel comfortable about his/her own teaching, assessment tools are required to provide the necessary data about the efficiency of the instruction occurring in the classroom setting. Observation of the revealing pattern of academic performance on the test tasks could indicate more useful data about learner understanding, or maybe misunderstanding, of the psychological construct (Kopriva, 2008). Thus, language teachers require a concrete understanding of assessment since they guide learners to develop self-assessment skills and strategies and they adopt or develop tests for language courses, and prepare students to take tests beyond the classroom and language program (Chapelle & Douglas, 2006).

People who use testing and assessment have to comprehend what the findings of assessment tasks mean, and what assessment tasks do to the individuals who are involved in the practice of assessment (William, 2015). The need for quality testing and assessment continues to increase as our mission of guiding more learners than ever before to reach higher levels of academic success becomes increasingly pressing (Stiggins, 1995). Thus, the significant place of assessment has been dramatically acknowledged in the field of educational instruction in the last decades (Kahl, Hofman & Bryant, 2013; Looney, Cumming, van Der Kleij & Harris, 2017; Stiggins, 1995). Education reform initiatives have, in last years, recognised the necessity for higher attention to the educator's knowledge and professional development (Daniel & King, 2010). "Unless the teacher is able to identify and analyse the errors a student makes in handling the target language, he or she will be in no position to render any assistance at all through appropriate anticipation, remedial work and additional practice." (Heaton, 1990, p. 6)

While testing and assessment having a growing attention as a major component in any educational context, teacher effectiveness depends, possibly, on their comprehension of using and interpreting a variety of procedures of testing and assessment. For this reason, the present use of a huge range of learner assessment practices makes knowledge of assessment and evaluation especially essential for every classroom teacher today (Bachman & Palmer, 1996; Daniel & King, 2010; Stiggins, 2014). “In an ideal situation, teachers will have enough resources and expertise available in their program that proficiency, placement, achievement, and diagnostic tests can be developed and fitted to the goals of the program and to the ability levels and needs of the students” (Brown, 1996, p. 40). However, the problem is that plenty of classroom teachers and other related parties do not have a broad knowledge and comprehension of fundamental components of assessment, such as reliability and validity (Popham, 2009). Thus, test developers’ (mostly classroom teachers) experience, knowledge and skills have a crucial role in the preparation of quality test tasks and testing conditions.

More can be achieved in order to enhance education by developing the skills of educators than by any other individual element (Wright, Horn, & Sanders, 1997). On the other hand, language educators and administrators require the necessary tools for evaluating and reflecting on test data to be able to make appropriate decisions about teaching activities and curriculum development (Boyles, 2005). In this regard, shifting from interpreting the assessment results in a passive role to taking the role of an active user of assessment data that will influence curriculum and instruction is the essential bridge that should be created and passed to gain appropriate assessment data for each classroom teacher (Boyles, 2005). That is, it is now recognised that the teacher’s role involves both formulating and scaffolding learning based on continuous feedback from both internal and external assessment tools (Inbar-Lourie, 2008). Hence, as all language teaching programs include some testing and assessment, thus, classroom language teachers should be able to take active part in making informed judgments in selecting suitable language tests or to develop, construct, and plan appropriate their own test tasks (Bachman & Palmer, 1996). Thus, it is one of the prominent duties of language teachers to keep themselves informed about the requirements of good testing assessment processes.

It is clearly highlighted that teachers should have the necessary knowledge and skills to correctly conduct classroom assessment procedures, to interpret the results of standardized tests and report the results to stakeholders, and to prepare tests that are in line with the assessment standards (Jeong, 2013). Teachers need to be capable of efficiently collecting and utilising evidence of what learners know and able to perform to support their success and

improvement (Kahl et al., 2013). From this perspective on, learner success depends on the quality of assessment practices, and the impact of those practices relies on the effectiveness of the interpretation of assessment data that they are based upon (Jabbarifar, 2009). Thus, our aim as educators is to foster positive washback from assessments (both external and internal to our teaching context) and via assessment to enhance meaningful learning (Cheng & Fox, 2017).

Thus, the present study aims to shed light on the current assessment knowledge and needs of EFL instructors working at schools of foreign languages at universities. To improve their knowledge in assessment, language professionals need a repertoire filled with necessary qualifications and skills that help them analyse assessment data and interpret the meaning of them, respond to these findings, and implement them in curriculum and instruction development process (Boyles, 2005). Additionally, in order for educators to have necessary assessment skills and knowledge in a learning setting practices, it is crucial that they can attend the necessary and related teacher training programs in assessment (Jeong, 2013). Thus, the results of the study will potentially guide the administrative authorities and policy makers about the discrepancy between the theoretical implications of language assessment concepts and the assessment implementations of the language teachers in their classroom setting. Consequently, teachers might be provided with necessary professional training in assessment.

1.2. Research Questions

The present study aims to find answers to the following questions:

1. What is the EFL instructors' current level of language assessment knowledge in terms of general and skill-based assessment?
2. Does LAK level change according to gender and assessment unit experience of EFL instructors?
3. Is there a relationship between EFL instructors' current level of language assessment knowledge and their year of work experience?
4. Which language assessment practices, beliefs and suggestions do the EFL instructors who are highly knowledgeable in language assessment possess?

1.3. Methodology

The mixed-methods approach, which involves both quantitative and qualitative research methods, will be used in the present study. According to Leedy (1993), quantitative research aims to answer questions on relationships within measurable variables with a desire to explain, identify and control a phenomenon. In addition to this, in qualitative design, 'the

researcher seeks to establish the meaning of a phenomenon from the views of participants' (Creswell, 2014, p. 33). Furthermore, convenient sampling will be used to select the study group. Based on the purpose of the study, EFL instructors working at schools of foreign languages in four state universities are considered as appropriate to study with. In this study, a questionnaire developed by Ölmezer-Öztürk and Aydın (2018b) will be used to assess the EFL instructors' knowledge of language assessment. Within the scope of the questionnaire, EFL instructors' knowledge of assessment in reading, listening, writing and speaking skills will be investigated. Following the collection of the data, SPSS software will be used in order to analyse the finding of the questionnaire. In the second phase of the study, based on the results of the scale, 10 high scoring instructors will be interviewed to collect deeper information about their background and practices related to language assessment.

1.4. Place of Study in the Literature

There have been a few studies in Turkey which investigate EFL instructors' or teachers' knowledge of language testing and measurement (Ölmezer-Öztürk & Aydın, 2018a; Öz & Atay, 2017; Köksal, 2004). Furthermore, there are some studies conducted in various contexts. Hudaya (2017), for instance, investigated teachers' assessment knowledge and skills in terms of their' preparation level in assessing learners' performance. In another study, Fulcher (2012) evaluated 278 international EFL teachers' knowledge of assessment via an online questionnaire. The teachers were from different settings in the world from New Zealand to the Middle East. Likewise, Xu (2015) carried out a three-year longitudinal study to examine the improvement of four Chinese novice EFL teachers' classroom assessment knowledge by collecting data from 26 classroom observations, 48 interviews, and 171 participants' journal entries. All and more of these studies conducted as either quantitative or mixed-method indicate that EFL teachers encounter certain difficulties during planning their assessment tools and procedures as a result of their lack of assessment knowledge. In most of these studies, organisation of training programmes in order to aid such teachers who are in need of learning more about testing assessment processes is suggested.

In contrast with the number of studies carried out in various areas in the world, the studies searching for EFL teachers' knowledge of language assessment are quite limited in Turkish context. This present study, therefore, will determine EFL teachers' qualification in language assessment in higher education in Turkey and provide suggestions for the EFL instructors, school administrators and policy makers.

CHAPTER 2

THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1. Assessment and Testing in General

The terms testing and assessment might be used with different meanings in different contexts; however, theorists generally use them interchangeably. According to Stiggins (2014; 2017), assessment, with its most basic definition, involves the process of collecting data on learner performance to be able to inform educational reforms. Likewise, Brown (2003, p.3) defines a ‘test’ as ‘a method of measuring a person's ability, knowledge, or performance in a given domain’. However, assessment is generally used as an umbrella term involving the practices of designing tests, administering tests, analysing the results and making educational decisions based on those results. Thus, today teachers have a tendency talk about assessing (instead of testing) learners since we consider the continuous measurement of learner performance as more than only measuring knowledge and abilities in a specific area at a stage in time for selecting or grading aims (Rogier, 2014).

Cheng and Fox (2017) describe two dimensions of assessment – *assessment for learning* and *assessment of learning*. The former concept is explained as the procedure of searching and analysing data for usage by students and educators to be able to identify the current level of learners in their learning journey, in which they are expected to reach and how best they can go to the target level. For instance, the Socratic approach: This is one of the earliest educational methods, and it may be considered merely as teaching via asking questions. In contemporary world, the Socratic model is regarded as the application of *assessment for learning* (Gareis & Grant, 2015). By contrast, the latter concept refers to assessments that take place after learning has happened, to decide the degree of learning.

Assessment, is mostly divided into two main parts; assessment which is prepared to enhance learning and instruction in learning settings; and assessment practices for certification purposes, public reporting, student selection and program accountability (Barber & Hill, 2014). Assessment is supposed to serve both as a learning means to guide learners to acquire more, while still certifying or verifying student improvement as required (Stiggins, 2014). Thus, instruction and assessment must be regarded together to ensure that teaching provides itself to assessment and that the assessment results can direct continuous instructional planning and development (Jabbarifar, 2009). Thus, standardised external tests and educators’ interpretations are regarded to be different yet complementary varieties of assessment and both

of them are reasonable in nature and they need to be prepared in a way that they meet advanced standards of quality (Allal, 2013).

The assessment involves various role according to some scholars. Popham (1995), who utilises both assessment and testing terms interchangeably, explained three main roles of assessment which directly influence classroom teachers: (a) public considerations of instructional usefulness, (b) assessment of educators, and (c) achievement of educational objectives. Thus, assessment of students' progress is an inseparable part of any learning context in order to gather data for the effectiveness of teaching programs, teachers or teaching methods. With a similar approach, Hughes (2003) suggests that our assessment system should be one that;

- continuously provides accurate measures of the knowledge and skills which we need information for,
- involves an influential impact on instruction and learning (in the scenarios where assessment influences instruction),
- is economical in terms of time and money.

Popham (1995) also described several critical areas surrounding assessment which educators must understand to be influential in their instructional context. Some of these areas are (a) detection of learners' strong and weak areas, (b) observation of learners' improvement, (c) grading the learners, (d) teacher-designed tests, (e) awareness of externally prepared instructional test items, and (f) identification of the educator's educational effectiveness.

The ultimate purpose of assessment is to enhance learning, rather than only giving a grade for the amount of course content that a learner has acquired (Rogier, 2014). However, if we suppose a single 'best' test exists, and we tend either to implement that test itself, or to utilise it as a sample for developing our own test, we probably come up with a test which will be ineffective for some of our learners (Bachman & Palmer, 1996). Hence, it is crucial to know that it is impossible to find a single test or assessment method that can meet all the purposes of a learning atmosphere. As Stiggins (1995) puts it, different needs and purposes require different types of practices and thus different types of assessments. It is essential to note that assessments are not either merely for general or specific purposes – every test is developed for a certain purpose – yet that there exists a continuum of particularity which has 'very general' at one end and 'very specific' at the other, and any test may fall at a point on this continuum (Douglas, 2000).

Stiggins (2008) highlights three key points: First, teachers need to always be clear about their assessment purposes while assessing their students' learning. Second, assessment must always be used to gather data for learning levels and instructional decision making. Finally, the contemporary period of accountability has misapplied the term 'assessment', requiring a reclamation of appropriate instructional assessment procedures by educators. In addition, in parallel with Stiggins's points, Cheng and Fox (2017, p. 7) state three major purposes of assessment:

- Instructional: Assessment aims to gather evidence for students' learning, skills
- Learner-centred: Assessments for both diagnostic purposes and placement purposes involve determining learners' needs to improve their learning.
- Administrative: With grading, public announcements are made to learners, their families, and principals about learner improvement and achievement.

2.1.1. Validity

Validation of assessment was defined as “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” (Messick, 1989, p. 13). It is also described as the practice of creating and collecting data to prove the ‘well-foundedness’ of interpretations regarding trait from assessment results, i.e., particularly, assessment must deal with validity that is based on evidence (Weir, 2005). If assessment results are influenced by learners' skills other than the one teacher's aim to assess, they will not be efficient indicators of this particular ability (Bachman, 1990). Hence, the test should have a purpose of being a correct measurement of the specific ability which it is expected to assess: to the amount that it assesses other knowledge and abilities at the same test administration, it is not going to be a valid test (Heaton, 1990; Hughes, 2003). To illustrate, scores from a test developed to generate information about a learner's vocabulary knowledge may not be particularly effective for placing learners' in a writing program.

In examining validity, teachers must also consider the appropriateness and effectiveness of the test result for a determined purpose (Bachman, 1990). Also, there might be more than a single task in an assessment and validity judgements will be related to the whole test not an individual test task (Weir, 2005). According to Cronbach (1971), there are three fundamental steps in validation of tests: formulation, explication, and validation. Formulation and explication phases include the process of theorising and development of a test. The validation step involves gathering the validity evidence in order to support the

interpreting and using the test results. However, the notion of validity is understood better when all of its components or types of it are evaluated at the same time.

2.1.1.1. Face validity

When an assessment task looks appropriate to other test developers, educators, moderators, and test takers, it may be regarded that the test has face validity (Heaton, 1990). Face validity is not exactly a scientific indicator and is not considered as providing proof for construct validity (which conforms the most of validity notion of tests), yet it can still be very important within the instructional context (Hughes, 2003). Especially, the learners' motivation is enhanced when a test task has high face validity, since plenty of learners will study harder when the test task looks appropriate.

2.1.1.2. Content validity

This type of validity is related to an elaborate investigation of the specific curriculum objectives and of the knowledge or skills assessed. Tests must be so designed as to include a representative selection of the classroom course content, the association between the test tasks and objectives of the course always being salient (Heaton, 1990). In other words, an assessment task is regarded as having high content validity when it contains an appropriate sample of the target structures and abilities aimed to be learned (Hughes, 2003). The examination of content validity of a test, usually experts of the field are counselled. These may include subject teachers, curriculum designers and other test developers. These judges are requested to look at the texts that have been chosen for inclusion in the test and analyse them in terms of their reflection within the target content area (Fulcher & Davidson, 2006).

2.1.1.3. Construct validity

This kind of validity acknowledges the availability of particular instructional theories or constructs behind the accomplishment of target content and skills (Heaton, 1990). To be able to prove that a test has high construct validity, we must measure certain identified characteristics of it according to a theory of learning. Cronbach and Meehl (1955, p. 284) explain construct validity as an interlocking rules system which generates a theory as a 'nomological network'. The notion of a nomological network is described by the scholars giving examples from language assessment. Firstly, it involves plenty of constructs, and the names of them are abstract. For example, in language teaching and assessment, 'fluency' and 'accuracy' are very well-known constructs. Secondly, this network includes the observable variables – the behaviours that we can observe and assess directly, while we cannot observe learners' 'fluency' and 'accuracy' directly (Fulcher & Davidson, 2006).

The main issue with the construct validity is creating a definition for what a ‘construct’ is (Fulcher & Davidson, 2006). Kerlinger and Lee (2000, p. 40) shed a light on this problem by giving some explanations for the term. That is, concepts transform into constructs when they are so clearly defined that they become ‘operational’ (or observable) – we can assess them in a test by associating the term to some act of behaviour that is observable (whether this is choosing the correct item by ticking a box or performing some interactive conversation), and we can determine the part of a construct in a learning theory that links one construct to another. For instance, language tests are designed to assess particular theoretical constructs such as ‘reading comprehension skill’, ‘use of a certain grammar structure’, ‘listening comprehension skill’ etc. Because of this, recently, construct validity, generally, has been referred as the main notion of validity by most assessment developers (Hughes, 2003).

2.1.1.4. Criterion validity

This kind of validity is about the extent to which learners’ performance on an assessment task is parallel with those gathered from some other independent and reliable assessment of the learners’ ability (Hughes, 2003). In such a case, the validity evidence is examined in accordance with the degree of the diagnostic link between the assessment result and the target learner output on the criterion (Fulcher & Davidson, 2006). In other words, criterion validity is related to performance on an assessment instrument correctly estimates the present or future performance on another definite measurement.

2.1.2. Reliability

While validity is described as a quality of analysis and usage of the assessment, reliability is considered as a quality of test results themselves (Bachman, 1990). Reliability is a crucially needed quality of a good assessment task, because it is to become a valid test, it has to be a reliable assessment instrument (Hughes, 2003). Teachers and test developers should construct, conduct and evaluate test results in such a way that learners’ scores acquired on an assessment at a certain time and specific place would probably be similar to the ones that would have been gathered when the same test had been conducted to exactly the same learners having the same proficiency level, yet at another time (Hughes, 2003). Thus, if a certain assessment is implemented to the same students at different times (without any practice work occurring between the two administrations) then, to the degree that it generated various results, the test is not regarded as reliable (Heaton, 1990). To put it another way, the more parallel scores we take from each test, the more reliable assessment we have (Hughes, 2003). However,

in language assessment, the issue is that testing language knowledge and skills for placement, achievement or diagnosis purposes, or testing other mental traits of learners is much more difficult to do consistently than assessing the weights of these same people (Brown, 1996). As Douglas (2010, p. 10) puts it “whatever the causes of inconsistent test performance, we have an ethical responsibility to make our tests as accurate as possible so as to give our students as fair a measurement of their abilities as we can”. Hence, language teachers should consider the reliability quality of their assessment instruments in order to provide valid and more reliable test administrations.

2.2. Types of Assessment

Assessment is categorised in accordance with their purposes for which the assessment is conducted. The distinction is generally made between;

- informal and formal assessment,
- criterion-referenced and norm-referenced assessment,
- formative and summative assessment,
- objective and subjective assessment,
- discrete point and integrative assessment.

2.2.1. Informal and formal assessment

Informal assessment involves “techniques that can easily be incorporated into classroom routines and learning activities” (Navarrete et al., 1990, p. 5). These techniques might be utilised anytime in the learning environment without intervening learning time. Informal assessment may be designed in a number of different versions, involving tasks requiring unplanned answers, as well as mentoring and other constructive feedback to the learner. A good amount of an educator’s informal assessment is involved in classroom tasks developed to elicit responses without recording learners’ scores and making definite interpretations about his learners’ performance. However, formal assessments contain more systematic and planned testing methods prepared to provide classroom teachers and learners with an appraisal and evaluation of student achievement (Brown, 2003).

2.2.2. Formative and summative assessment

Formative testing is parallel with *assessment for learning* in which assessment procedures are conducted collaboratively, and the instructional decisions are particularly about the route in which instruction and learning should go (Cheng and Fox, 2017). McKay states that formative assessment is implemented for the goal of improving students’ learning and

might be embedded in classroom teaching in a complex way, primarily in a way which is called on-the-run assessment (2006). Thus, in formative assessment a teacher should match the evidence collecting techniques that he or she uses with a learning objective so that they create effective and conclusive evidence of improvement and learning as well as learning gaps connected to the specific objective (Kahl et al., 2013). To sum up, the crucial point in such formation is the instruction (by the educator) and internalization (by the learner) of useful and constructive feedback on a student's performance, with an aim of continuation (or formation) of student learning (Brown, 2003).

Summative assessment is, on the other hand, parallel to *assessment of learning*. Summative assessment is developed in order to assess learners' improvement, learning of skills and their academic success as a consequence of a specified educational term – generally when an academic term or a course programme finishes (Cheng and Fox, 2017). A summation of the skills a learner has acquired refers to looking back and examining how well that learner has accomplished instructional goals, yet does not primarily highlight the necessary method for future improvement (Brown, 2003). This kind of assessment contains more performance-based tasks, such as open-ended questions or guided-response items and these require more time and energy to construct and score and necessitate considerable planning in establishing scoring scales and scaffolding (Kahl et al., 2013).

2.2.3. Subjective and objective assessment

The distinction between objective and subjective testing is made based on scoring methods used in each. When there is no judgement made by the scorer to assess learners' performance, then the scoring is called as objective testing. In contrast, if judgement is made by the scorer then this testing method is called as subjective testing (Hughes, 2003). True/false and multiple-choice tasks can be regarded as forms of objective testing and essay tasks and oral exams can be given as examples to subjective assessment. Thus, generally in objective assessment the answers are defined beforehand. On the other hand, subjective assessment tasks can result in a range of acceptable and alternative answers.

2.2.4. Norm-referenced and criterion-referenced assessment

Another way of grouping assessment methods is based on the ways grouping learners according to their test scores. Norm-referenced and criterion-referenced assessment are two methods related to this. Norm-referenced assessment is related one test participant's academic performance on the test to that of other individuals taking the test. Assessors are not informed directly on what the learner is capable of performing using the target content (Hughes, 2003).

In norm-referenced assessment, each learner's score is interpreted in accordance with an average score (mean), middle score (median), degree of variance in scores (standard deviation), and/or percentile rank (Brown, 2003). Brown (1996) highlights that norm-referenced assessment aids school managers and educators to make curriculum-related decisions which include evaluating learners' proficiency, selection and placement decisions.

On the other hand, criterion-referenced assessment aims to measure students' performance in relation to whether they are capable of performing some test task or set of tasks according to the objectives of the course (Hughes, 2003). Criterion-referenced assessment is developed to give learners feedback, mostly in the form of scores and/or grades, on a particular course or course objectives (Brown, 2003). This type of assessment helps educators to make classroom decisions such as diagnostic and achievement (Brown, 1996).

2.2.5. Discrete-point and integrative assessment

Another categorisation was made between discrete-point and integrative testing. Discrete-point assessment tools are developed on the estimation that language can be divided into its components and those components can be assessed successfully (Brown, 2003). These components consist of reading, listening, writing speaking main skills, and other various areas like phonology and punctuation. Thus, it is supported that a well-constructed language assessment should include all of these components individually. On the other hand, integrative testing involves a unitary perspective in relation to assessment of language components (Brown, 2003). Hence, this approach supports the idea that language performance should be considered as a whole phenomenon.

2.3. Classroom Assessment

The primary difference between large-scale educational assessment and classroom assessment is the classroom context (Fulcher & Davidson, 2006). If teachers are conducting tests in the classroom developed by other teachers or test developers, this is not regarded as classroom assessment as the assessment is not designed by the classroom teacher but by other people who are one step, and maybe plenty of steps, away from the students in the course and the educational environment of the classroom (McKay, 2006). However, for external assessment instruments, the educational setting is probably a reason for invalidity, and the necessity to maximize information generally means that test tasks and items are to be independent (Fulcher & Davidson, 2006). However, in classroom assessment the context requires that the instructional practices are integrated and in relation to the learning needs of the learners (Fulcher & Davidson, 2006).

The aim of classroom assessment is quite different from the purpose of the external assessment. Whilst the latter is usually related to assessment with the goal of student selection, the former deals with assessment with the goal of helping educators to improve their own teaching skills in teaching by making alterations in their classroom practices to help particular groups of learners in the classroom to benefit more (Heaton, 1990). Thus, the primary aim of classroom assessment and testing is to provide learners with the conditions to reveal the knowledge and skills that they have acquired instead of detecting their weaknesses or to identify the areas that they could not learn (Jabbarifar, 2009). For this reason, the external tests need to be as comprehensive as possible to be able to gather lots of samples of evidence about the learner within a fixed period of time, while a teacher can spend months or years to assess his or her students (Fulcher & Davidson, 2006). In conclusion, good classroom-based assessments are not single indicators of learning, yet real experiences of learning by themselves.

Classroom assessment is regarded as formative by its nature rather than summative by plenty of scholars (Cheng & Fox, 2017; Fulcher & Davidson, 2006; Popham, 2009). In classroom assessment, a psychosocial context exists in which classroom-level assessment and teaching are integrated, and the assessment has the purpose of encouraging learning as expected by the course learning objectives and outcomes (Cheng & Fox, 2017). Thus, the findings of classroom-level testing are conducted by educators to evaluate an underway course content, in the same way as educators would progress when classroom assessment is designed accordingly, so that not only the teacher but learners are also typically depending on assessment-elicited proof to evaluate and, if required, alter their practices accordingly (Popham, 2009). As a result, in the language classroom learning setting, classroom-based assessment provides direct feedback to the student that aids him or her to determine what is required to be learned next to be able to become a self-sufficient user of that target language in a new environment. For this reason, the feedback from classroom assessment must embrace diagnostic data, and this is not often reached in formal tests (Fulcher & Davidson, 2006).

Despite the effectiveness of most of the practical support in sources for teachers, the distinctness between the classroom assessment and assessment affecting large numbers of test takers is not considered well, leading to noticeable confusion among test developers (Fulcher & Davidson, 2006). Regarding this, practitioners require the opportunity to raise their awareness in classroom assessment to improve learning. Since, without it, learning environments cannot improve (Stiggins, 2014). The test developer is supposed to make sure that the administration of a specific test is appropriate for the defined purpose and target group

of learners. This process includes investigating the degree to which meaning of scores can be interpreted beyond the occasions of the test, to skill to generalise in non-test circumstances (Fulcher & Davidson, 2006). Thus, teachers must build up assessment skills together to comprehend that classroom assessment cannot be seen as a separate concept from the ‘real’ practice of instruction and learning, yet a principal element for effective instruction and learning (McKay, 2006).

The evidence was found that classroom teachers are in need of more training in assessment according to their self-reported lack of readiness to test and assess their learners, citing gaps in their undergraduate teacher education programs (Rogers, 1991). In line with this idea, Deneen and Brown (2016) conducted a study investigating the influence of a master’s level educational assessment class on teachers’ assessment knowledge and conceptions related to assessment which may impact their practices in assessment. The results indicated that both pre-service and in-service teachers’ factual knowledge about assessment was enhanced while their understanding of conceptions of assessment were not affected by the instruction of the assessment course. The research suggests teachers’ competency in assessment conceptions impact their implementation of knowledge and skills (Brookhart, 2011; Deneen & Boud, 2014). Moreover, regardless of their years of experience, the research by Zhang and Burry-Stock (2003) showed that teachers who received training in assessment had higher levels of self-perceived assessment competencies especially in standardised testing, interpreting test results and test revision than the teachers without such a training. Thus, the findings recommend an approach to language assessment knowledge that combines knowledge, skills, and practices in a procedural text which attempts to find a balance between what is required for both classroom and external assessment (Fulcher, 2012).

From the studies conducted with teachers of various subjects, it appears that a great number of pre-service teacher-education programs have still not efficiently addressed the need for developing the assessment knowledge and skills (Kahl et al., 2013; Şahin, 2015). Findings indicated that practising teachers lacked assessment knowledge but they were better in terms of their perceived skilfulness, and they revealed a more positive attitude to instructional assessment than pre-service educators (Alkharusi, Kazem & Al-Musawai, 2011). In another study, Mertler (2003) surveyed in-service and pre-service teachers using the Classroom Assessment Inventory (CALI) and found that practising teachers performed highest in “administering, scoring and interpreting results of assessments” and lowest in “developing valid grading procedures” while pre-service teachers scored highest in “choosing appropriate assessment methods” and lowest in “developing valid assessment grading procedures”.

However, in all cases in which scientifically significant differences were identified, the in-service teachers performed better than their pre-service counterparts.

2.4. Language Testing and Assessment

It is an inevitable need to measure the skills of language students whose in order to diagnose their needs and place them in suitable language programs, observe their progress, and make instructional decisions in relation to moving them from one kind of language program to another or graduating from the language program successfully (Kopriva, 2008). Language performance tasks are objective-oriented which means the language learner is informed about what is expected to be achieved in assessment, and these are appropriate to a particular learning context. The assessments can include reading, listening, writing or speaking and may result in a blend of these main skills of language use (McKay, 2006). The quality of assessment of language proficiency skills influences the assessment of linguistic achievement to the amount that language assessment instruments generate evidence about when and where learners lack the necessary knowledge and/or skills in target language to adequately comprehend test tasks or demonstrate their language performance in content areas (Kopriva, 2008).

The way we describe the language skills we aim to assess is inescapably associated with the characteristics of the procedures followed for elicitation of language abilities, or assessment methods teachers depend on to evaluate these abilities. However, Bachman (1990) points out the problem of identification of language constructs which conforms the foundation of any language test development activity. He states that the main problem with language assessment is that we rarely have a clear definition for language domain which unambiguously explains the content of language performance tasks from which appropriate assessment tasks can be designed, so that proving either content coverage or content relevance is difficult. Bachman (1990) elaborates his concern regarding the issues currently facing language test developers by categorising them into two general areas. First is the need for clearly specifying language skills and areas and other elements that affect linguistic performance of the learners on language tests precisely to provide a ground for test design and for the interpretation of test scores. The second point is determining how test scores from language assessment instruments behave as indicators of performance (defining the constructs). In other words, what are the scoring and measurement properties of assessment tasks of language knowledge and skills? Finding an answer to this question is particularly challenging as language tests might assess several different but interrelated skills.

Language test developers, particularly classroom educators, are supposed to take part in practices of language assessment, reflect on students' improvement in accordance with external criteria, and prepare language students for certain external examinations (Inbar-Lourie, 2008). Language test developers, as applied linguists, need to fulfil the practical needs of learners for more effective assessments of language knowledge and abilities needed in language learning, language curriculum evaluation, and for making instructional decisions about students' attained levels of proficiency with respect to a variety of instructional and employment requirements (Bachman, 1990). For this reason, all language assessment tasks must be developed according to a set of clearly defined language skills, whether this is elicited from a language course syllabus or a theory of language learning, and must benefit from some process for deriving language performance (Bachman, 1990).

2.4.1. Types of language tests

In earlier sections, it is supported that different purposes will generally require different types of assessment types and tasks. In line with language assessment purposes, Hughes (2003, p.8) describes four main aims of assessing language performance which conforms the foundation for various types of language tests:

- To assess language proficiency (language proficiency tests),
- To measure how successful learners have been in meeting the course objectives (language achievement tests),
- To detect the weaknesses and strengths of learners, to diagnose the parts they already know and they do not know (diagnosis tests),
- To place learners in teaching programme that is most appropriate for their ability (placement tests).

Kopriva (2008) suggests that following the initial placement, English proficiency tests may continue to be utilised diagnostically at least once a year to provide data about learner language competencies and skills, as well as areas of difficulty. Moreover, while designing assessment instruments for testing achievement or diagnostic purposes, previously developed placement or proficiency assessment tasks may be adjusted and used when necessary. Later, language educators may want to construct their test items by themselves for a placement test in order to make sure that the test items used to place learners into the sections of study are in line with the goals of the language program and to the content the learners are encountering in the course program (Brown, 1996).

2.4.2. Language test design cycle

Item writers must develop their test's specifications before starting with the construction of the test itself. (Alderson, Clapham & Wall, 1995). Test developers need these specifications to encapsulate the content in which all language learners will be tested. Thus, developing test specifications is the foundation for content area interpretations about current knowledge and lacking areas of students (Kopriva, 2008). The content of the test tasks and overall administration procedures should be available for the access of teachers, administrators, parents and students, who are directly affected by the interpretation of test scores. Moreover, in order for test results inferences to be utilised with confidence by the educators, it seems necessary that appropriate access to required item content should be provided for test takers, especially those with detected challenges which run counter to the techniques typically used in high-stake testing environments (Kopriva, 2008). For this reason, designing detailed test specifications, documents including all the exam rules and instructions, directions for the proctors etc. is an essential need to enable a reliable and fair assessment atmosphere.

Fulcher and Davidson (2006, p. 70) suggests that the language of course content or expected response may be categorised as followed:

Language characteristics:

- Organisational features
- Structural (syntax, vocabulary, phonology)
- Discourse (cohesion, conversational or rhetorical organisation)

Pragmatic features:

- Functional (imaginative, manipulative, ideational, heuristic)
- Sociolinguistic (dialect, register, figurative language and cultural references).

Thorndike and Hagen (1977) suggested a procedure for developing good language tests. They stated that the construction of effective assessment tasks should be based on a reasonable sequence of steps connecting the assumed language construct, to the defined linguistic performance. This test development sequence involves three phases: (1) identification and definition of the construct based on a theory; (2) description of the construct operationally, and (3) constitution of procedures for assigning scores to observations.

Mislevy et al. (1999, p. 6) (cited in Fulcher & Davidson, 2006, p. 69) highlights the three stages in task design:

- Identifying the competencies or abilities (constructs) that we aim to test
- Identifying the connection between the language performances and the constructs in situations that require their use
- Describing the characteristics of situations that generate the evidence to deduce inferences from performance to constructs

While constructing language assessment tasks, it is a necessity to develop tasks and texts that reflect the content as much as similar to the one that learners have already been familiar with and/or they are likely to encounter in their future experiences (Weir, 2005). Moreover, the difficulty of the test tasks should be parallel with the challenge level of classroom tasks in order not to cause test takers to feel exam stress. Thus, it is highly suggested that before the administration of any assessment task, another professional who has been trained in assessment or at least knows the profile of the test takers should proofread the test content. Hughes (2003, p. 58) suggests ten steps in constructing language tests:

1. Stating the testing ‘problem’,
2. Writing testing specifications clearly,
3. Constructing the test items according to the specifications,
4. Administering the test items on L1 users and editing the problematic ones accordingly,
5. Administering the assessment items on non-native speakers who are from similar backgrounds to real test takers,
6. Analysing the results of piloting and making necessary changes in needed,
7. Calibrating scoring scales,
8. Validating,
9. Developing guidelines for test takers and related staff,
10. Training the assessment staff (interlocutors, assessors, etc.) accordingly.

2.4.3. Assessing language skills

In language learning classrooms, four major skills for communication via language are often largely known as listening, reading, writing and speaking. These skills, in plenty of environments where the target language is learned for general purposes, are to be integrated appropriately. Moreover, assessment of these main skills should also be planned carefully so that the assessment environment reflects the actual language learning situation (Weir, 2005).

When language teachers propose to assess a learner's ability and competence in one main skill or an integration of four main language skills, they measure learner's linguistic competence, but they observe the learner's performance (Brown, 2003). Thus, in receptive skills (reading and/or listening) performance, the elicitation technique can be developed to expect predetermined responses and merely those responses. In productive skills (writing and/or speaking) performance, the oral or written text must be effective to elicit language production and an anticipated range of linguistic performance to which grading or rating processes apply appropriately (Brown, 2003).

Heaton (1990) points out some details about testing language skills. First, he states that test developers should focus on the kinds of assessment tasks that seem directly related to the target skill to utilise language in authentic performance, particularly interacting with other orally. Hence, assessment tasks which assess the skills to comprehend and reply effectively to suggestions, advice, requests etc. might be developed instead of story-telling or reading aloud tasks. Moreover, in writing part of a language exam, test items which require individuals to create mails, notes and text messages can be preferred rather than the traditional paragraph writing and composition tasks preferred in the earlier times. Finally, in receptive performance tests, items in which learners show their capability to draw certain information of a functional nature can be preferred to tasks assessing the understanding of trivial and unimportant details. Heaton (1990, p.8) suggests these tasks as forms of tests to assess four major skills;

- listening comprehension, short conversations, lectures and talks are given to test takers,
- speaking skill, mostly attending an interview, describing a picture, role-playing and problem-solving requiring a pair or group work,
- reading comprehension, in the form of giving students tasks to test their comprehension to understand the gist of a passage and to understand relevant information on particular parts in the passage; and
- writing skill, generally writing letters, reports, notes, instructions, etc.

2.4.3.1. Assessing reading skill

Grellet (1981) distinguishes four different ways of reading: scanning, skimming, intensive and extensive reading. She regards reading skill as a continuous process of guessing, evaluating, skimming, hypothesising, prediction and so on. As Marshall puts it, "when we read a piece of work, we are judging it" (2011, p. 29). Hence, assessment of reading skills requires appropriate instruments and methods to help learners show their reading skills as testable operations. Weir (1994, p. 2) suggests three fundamental 'operations in reading':

- Skimming: going over a text in a limited period of time;
- Reading intensively to comprehend the main ideas and useful details;
- Benefitting from knowledge of linguistic contributory skills such as understanding grammatical concepts (cause, effect etc.), discourse markers, lexical difficulty, syntactic structure and so forth.

The main problem in the practice of receptive skills is that they do not essentially, or mostly, manifest themselves directly in observable behaviour. While writing and speaking, people see and hear; while reading and listening, on the other hand, there is usually nothing to observe (Hughes, 2003, p. 136; Brown, 2003). Thus, the assessment of reading performance has to be made by inference (Brown, 2003) in order to assess the product of students' reading process. This assumption is essential to the reading assessment: if test writers are incapable of defining what they mean by the reading skill, it will be hard to devise instruments of assessing such an ability (Alderson, 2000). Urquhart (1992) suggests certain band descriptors under three headings for scales of reading assessment:

- Text factors: types of texts (argumentative, discourse etc.), discourse (cause/effect, reason/effect etc.), text (accessibility – signalling etc.)
- Task factors: complexity (shallow/deep etc.), paraphrasing
- Reader factors: flexibility (matching task to performance) and independence (choosing holistic or serialistic etc.)

Brown (2003, p. 189) defines four categories of reading assessment tasks:

1. *Perceptive*: This kind of tasks include attending to the features of larger texts such as letters, words and other graphemic symbols.
2. *Selective*: This category requires students to elicit information from a very short stretch of language: sentences, graphs, brief paragraphs, etc.
3. *Interactive*: This involves several paragraphs to one or more pages in which the reader needs to, in a psycholinguistic way, interact with the written language.
4. *Extensive*: This type of reading tasks applies to passages more than one page, involving academic essays, technical reports, professional articles, and books. The reader is required to grasp the global understanding of the reading text.

2.4.3.2. Assessing listening skill

Similar to that of reading skill, listening, as the other receptive skill, is difficult to observe as it is not reflected as an overt behaviour (Hughes, 2003, p. 136; Brown, 2003). If you do not have recorded oral language, there cannot be permanent visible product for speaking output since all those utterances you just heard have disappeared from your comprehension, and, hopefully, have been converted into meaningful intake in your brain (Brown, 2003). Hence, every language teacher should be aware that a learner's oral production skill – except for monologues, speeches, and so forth – can only be as good as his/her listening comprehension skill (Brown, 2003). Buck (2001) suggests a two-step procedure in development of listening tasks. Firstly, the item developers should prepare a list of tasks which the listeners should be able to perform. Secondly, they should prepare a list of tasks that replicate those real-life listening tasks in the assessment.

Brown (2003, p. 120) explains four common kinds of listening performance;

1. *Intensive*: Listening to process linguistic components (phonemes, stress patterns, fillers etc.) of a greater piece of language.
2. *Responsive*: Listening to a smaller piece of language (a statement, offer, etc.) in order to develop a likewise short response.
3. *Selective*: Listening to utterances of specific discourse for a few minutes to receive particular information. The aim of such processing is to search for designated information instead of comprehending the main idea of the listening text.
4. *Extensive*: Processing for grasping a general meaning of oral language. This type of listening involves a variety of practices ranging from listening to a dialogue to an academic text which require drawing a comprehensive message or purpose.

The components of listening skill assessment were described by Buck (2001, p.105). These are;

- Knowledge of the sound system,
- Understanding local linguistic meanings,
- Understanding full linguistic meanings,
- Understanding inferred meanings,
- Communicative listening ability.

2.4.3.3. Assessing writing skill

Designing a writing assessment task requires defining the target performance in testing for a particular purpose (Weigle, 2002). Having a subjective nature, assessing writing skill of language learners require classroom teachers to score learners' written production as reliable as possible. It is stated that the longer of a learner's written product in an assessment instrument, the more reliable the test is possibly to be, and thus, the more confidently the assessors can generalise the language performance of test takers on the assessment tasks (Weir, 2005).

According to Brown (2003, p. 220), written performance consists of four categories:

1. *Imitative*: This category involves the ability to perceive the spelling of words and comprehend phoneme-grapheme correspondences in the target language spelling system.
2. *Intensive*: These tasks require test takers to produce appropriate vocabulary using correct grammatical features at sentence level.
3. *Responsive*: Learners are supposed to perform linking statements into a paragraph while developing a logically meaningful two or more paragraph length texts.
4. *Extensive*: This type of writing tasks involves all the three processes requiring creation of a text with the length of an essay, a research article, or even a dissertation.

According to McNamara (1996) there are certain factors of assessment for writing skill (*Figure 1*). The utilisation of detailed grading scales and providing the raters with training to use these rating scales based on established criteria and practices have recently become the essential elements in developing a consistent approach to scoring written performance (Fulcher & Davidson, 2006). When a learner's score changes depending on the rater, the dilemma arises regarding whether it is the assessor's own personal attitude that influence the score, instead of the product of the learner (Fulcher & Davidson, 2006).

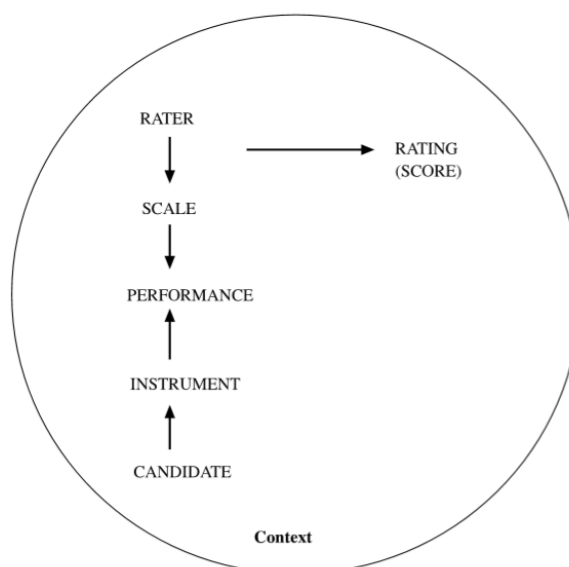


Figure 1. Components of writing assessment (adapted from McNamara, 1996, p. 9)

2.4.3.4. Assessing speaking skill

Similar to assessing writing skill, the efficiency and validity of speaking assessment is quite related to the assessment ability of the assessor. As a productive skill, while speaking can directly and empirically be observed, these observations invariably range depending on the correctness and efficiency of a learner's listening skill, which inevitably compromises the quality and reliability of a speaking performance assessment (Brown, 2003). Also, the test takers performance will differ if he/she has enough time to prepare as opposed to speaking spontaneously. In such a case, it is likely to differentiate between short and long turns (Weir, 2005). However, the language assessors are now more aware that discourse is co-constructed, and thus, the product of a language learner is partly dependent on the language performance of the interlocutor (Fulcher & Davidson, 2006). Thus, the role of interlocutor has gained a place in the literature of language assessment in the recent years.

Brown (2003, p. 142) defines five types of oral production;

1. *Imitative*: This is a completely phonetic level of speaking performance which involves a simply parrot back the utterance (imitation).
2. *Intensive*: This includes the production of short utterances of spoken language to demonstrate ability in a limited bad of grammatical, lexical or phonological features.

3. *Responsive*: These types of tasks include interaction at a limited level of short dialogues, greetings, basic requests and so forth.
4. *Interactive*: Interactive speaking tasks include longer and more complex interaction than responsive tasks.
5. *Extensive (monologue)*: Extensive speaking tasks involve oral presentations, speeches, and story-telling, which provide highly limited opportunities for interaction between the listeners and the speaker.

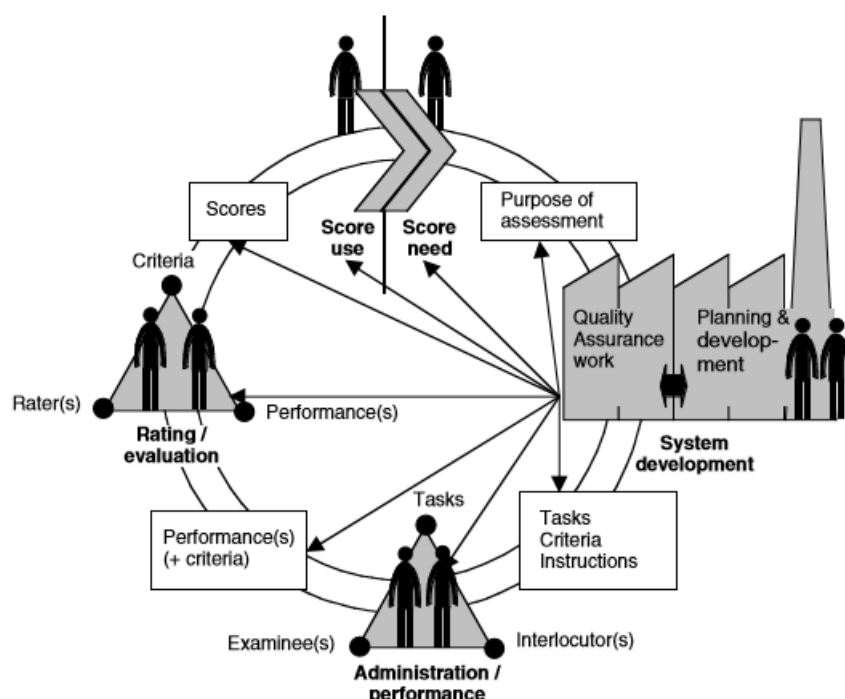


Figure 2. Stages, practices and products of speaking assessment (adapted from Luoma, 2004, p. 6)

In Figure 2, Luoma (2004, p. 6) illustrated the cycle of speaking assessment with the help of some images and terms. The connection between the related bodies is shown with arrows. On the left side, the task development phase is indicated as a factory in which ‘quality assurance’ is also shown as an important part of assessment task development. In rating/evaluation and administration/performance areas, triangles were used and related aspects were written around the corners of the triangle. Also, the human figures represent the impact of all related people within the process.

2.5. Language Assessment in Turkey

In Turkish context, language assessment tasks are prepared and administered by the classroom teachers at all public schools at primary, secondary and high school levels. However, in many private schools there has been an assessment office in charge of the preparation of summative assessment tools. Formative assessments are by nature designed and applied by the classroom language teachers. Similar to those, in higher education institutions, especially at preparatory schools which provide English courses for university students in order to prepare them English-medium majors, mostly assessment offices prepare the language tests including both summative and formative ones to provide standardisation along the different classes.

Before 90s, there was not a special course for providing undergraduate students of ELT departments with language assessment theory and practices. The candidate teachers were provided information about different assessment tasks but they were not trained on how they would develop and conduct those in their future classroom environments. Since the process of reconstruction of Faculties of Education in Turkey (1998), the universities have provided language testing and assessment courses to English teacher trainees. For this reason, many Turkish teachers of English face challenges while designing their assessment instruments for their classrooms; thus, they mostly prefer ready-made assessment tasks developed by other assessors or educators or use the test tasks provided with the course books.

2.5.1. Studies on language assessment in Turkey

The assessment methods and procedures followed in Turkish educational system have recently been reflected via an OECD report. According to the report of OECD (2019), while central regulations related to testing and assessment practices reflect the objectives of the curriculum, Turkish teachers tend to limit their classroom assessment processes with mainly multiple-choice tasks which are easy to construct and evaluate. It was also reported that these multiple-choice tests basically focus on subject areas and knowledge which are assessed in the central exams while other areas are given less importance as they are not examined in the high-stake central examinations. However, the ministry is continuously designing resources that are related to more complex assessment methods. On the other hand, this report reviewed all the subjects as well as English language so the results of it should be evaluated generally.

The studies in literature provided more specific data and findings related to the practices of Turkish EFL teachers in terms of their classroom language assessment practices. Öz (2014) conducted a study with 120 Turkish EFL teachers to collect data on their language assessment practices. The findings of the study revealed that most Turkish EFL teachers prefer conventional methods for assessment, mainly formal or summative assessment, rather than formative assessment procedures such as peer assessment and self-assessment tools. Moreover, in line with Öz's (2014) research, the teachers in Öz and Atay's (2017) study stated that they rarely benefitted from peer assessment and self-assessment methods. A study by Büyükkarcı (2014) also indicated that English teachers in primary schools in Turkey have positive attitudes toward formative assessment practices; however, they reported that they cannot use these practices in their own classes usually and effectively. Some of the derived causes of this issue are heavy work load of the teachers and their overcrowded classroom. Semiz and Odabaş's study (2016) confirms that EFL instructors reported that they were less familiar with formative assessment instruments such as self/peer assessment and portfolio; however, they also stated that they merely need basic level of training in the areas that they are not familiar with. As can be concluded from the studies, there seems to be a gap between the attitudes and beliefs of the EFL teachers in Turkey and their actual classroom assessment practices and self-reported needs.

The studies conducted to investigate the classroom measurement practices of Turkish EFL teachers revealed reflections of the competency and attitudes of these teachers to language assessment. Öz and Atay's study (2017) demonstrated that EFL teachers in Turkey preferred mostly pair work and speaking focused activities to assess their students' performance. Hence, the results showed that the English teachers were aware of the importance of oral productive skills but the findings also showed that the other three main skills of language use needed to be considered while designing assessment tools. Sarıçoban (2011) investigated the practices of English teachers at a state high school in Ankara, Turkey and found out the improvement of Turkish EFL teachers' assessment instruments regarding face validity, spelling, instruction, timing, contextualisation, scoring, legibility and reliability. Nevertheless, naming sections, washback and content and construct validity were found to be the areas that need further improvement.

The assessment course given to pre-service English language teacher candidates in Turkey is not sufficient to acquire enough knowledge and skills in foreign language assessment (Hatipoğlu, 2010). Since ELT students are not provided with all the required theoretical conceptions, key terminologies and practical reflections of these in actual

assessment instruments, they could not construct their assessment tools and evaluate their results properly. Moreover, Yastıbaş and Takkaç (2018) highlighted that in-service EFL teachers believed that they developed their language assessment knowledge and skills through work experience. However, while the teachers who graduated from ELT departments were satisfied with their assessment competency, the teachers graduating from non-ELT departments reported that they needed training in the area.

2.6. EFL Instructors' Language Assessment Knowledge

There, possibly, are plenty of contributing elements to create an effective learning atmosphere; among which educator's knowledge is probably the main element (Mellati & Khademi, 2018). "Assessment literacy" term was uttered by Richard Stiggins (1991) for the first time in assessment literature. Stiggins proposes that teachers who are assessment-literate are aware of the target content of the assessment, the purposes of the test, the procedures to follow while assessing, the potential problems with the assessment tool, and the pre-cautions to take before facing such problems. Boyles, also, defines 'assessment knowledge' as 'an understanding of the principles and practices of testing and assessment' (2005, p.18). However, Newfields (2006) defines assessment literacy differently according to the parties who are a part of the assessment process. In his perspective, for test takers, it is the knowledge of how to be successful on tests. For classroom educators, it is more about grading students correctly and fairly. And lastly for assessment developers, it includes all their practices within the scope of testing and assessment.

The goal and content of an assessment instrument is likely, to some extent, to determine who makes the best test developers (Alderson, et al. 1995). According to Popham (2004), being assessment literate is so crucial that it is considered as a "professional suicide" for educational professionals to neglect the place of it in learning environments. Xu and Brown (2016) suggested a process-oriented framework in which they describe a hierarchical model for literacy in assessment including six components of it:

1. Knowledge base
2. Assessment conceptions
3. Institutional and socio-cultural awareness
4. Assessment literacy in practice
5. Teacher learning
6. Development of teacher identity as assessor

Stiggins (1995) also suggested some guiding principles for educators to follow to develop their knowledge and awareness in the field of assessment. According to these principles, educators are to:

- Start conducting their assessment practices with a clear aim,
- Concentrate on achievement goals,
- Choose effective assessment methods,
- Effectively sample successful learner performance,
- Avoid bias and distortion.

Teachers with assessment literacy are knowledgeable about the way to assess the learning processes (Beziat & Coleman, 2015). In line with this, the quality and effectiveness of test items relies directly on the competency and expertise of the test writers (Haladyna, 2004). In other words, the educators who are literate in assessment – be they classroom teachers, school principals, course content developers, or superintendents – should be equipped with any assessment knowledge about what skills or elements of language they are assessing, why they are assessing those, how those are to be assessed to measure the achievement, how to produce samples of target outcomes, what might emerge incidentally, and what procedures to follow to prevent those unexpected problems before they appear (Stiggins, 1995). That is, in order to develop appropriate and effective instructional assessment, language teachers must be knowledgeable regarding the fundamental concepts of language assessment and how, potentially, those can inform the development of testing and assessment tasks and instructional decisions related to their implementations (Rogier, 2014).

Another term that has recently appeared in the field of language instruction is ‘language assessment literacy’. In general terms, language assessment literacy involves knowledge, competencies, and fundamental principles in assessment of linguistic performances (Fulcher, 2012). Giraldo (2017, p. 187) points out a framework describing ‘language assessment literacy’ and under the heading of ‘knowledge’ he lists the concepts of ‘awareness of applied linguistics, awareness of theory and concepts, awareness of own language assessment context.’ (*Figure 3*). Thus, language testing knowledge might be conceptualised as a construct which is based on theoretical principles of both language teaching and language assessment, is modified by socio-cultural elements, and is shaped by peoples’ considerations about language instruction and assessment (Xu, 2015). Having assessment knowledge (or literacy), hence, means having the competency and skills to develop and respond to critical queries about the aim of assessment, about the appropriateness of the

assessment tool to be used, about assessment setting, and about the evaluation of the learners' performance based on the assessment results (Inbar-Lourie, 2008).

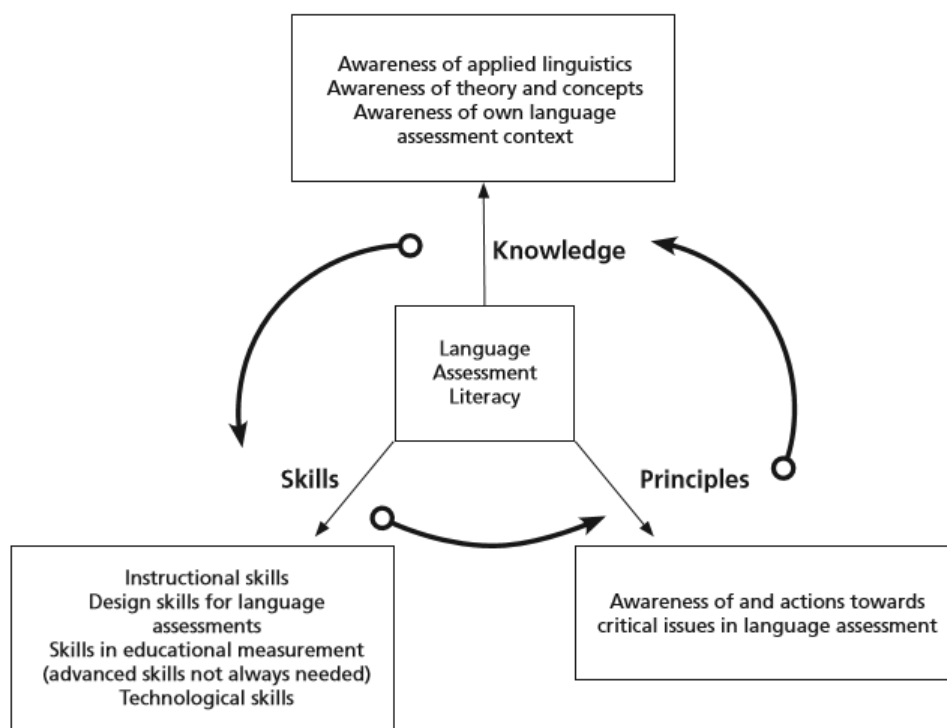


Figure 3. Language assessment literacy dimensions (adapted from Giraldo, 2017, p.187)

Effective language assessment practices in a learning setting, as mentioned earlier, play a crucial role in ensuring learners are achieving the curricular goals of the learning program (Beziat & Coleman, 2015). As a crucial guide for test developers, it is suggested that any instruction must involve three components: (1) The aims define the goals of learning context- the 'why'. (2) The procedures describe the ways of achieving these goals- the 'how'. And instructional activities are what really happens in the educational atmosphere – the 'what' (Jabbarifar, 2009; Inbar-Lourie, 2008). If teachers are willing to take the ownership of their learners and the assessment practices applied to assess learning and improvement, they are expected to be literate in assessment (Beziat & Coleman, 2015). Thus, to develop knowledge in assessment, all language educators need a repertoire which involves necessary assessment skills and strategies which will help them to interpret assessment scores, analyse the meaning of them, respond to indications of the results, and implement them in instruction and in curriculum evaluation (Boyles, 2005). Language teachers must always be motivated to be as critical of the tests that they develop as they are of the tests that are developed by external sources. Since, the fact that an assessment instrument is created by and for a particular course

curriculum does not essentially make it a quality instrument. Thus, measurement of assessment quality must be an integrated process of any assessment construction practice (Brown, 1996). However, Taylor (2009, p.24) pinpoints a crucial problem with these words;

We might even speculate that the language testing community's preoccupation with technical and measurement qualities in testing from the 1960s to 1980s, followed in turn by a similarly inward-looking focus on developing its own professional ethic during the 1990s, means that it has singularly failed to encourage wider public understanding of assessment principles and practices, that is, it has neglected to engage in the critical task of developing assessment literacy.

Educators who are less trained and less skilled in designing authentic assessment tasks consider these as more challenging to design than typical pen-and-paper assessments. Furthermore, assessment practices of language teachers are usually not in parallel with their educational objectives and tend to require low levels of mental processing (Levy-Vered & Alhija, 2015). From another perspective, the common existence of training courses for language assessment reveals that developing skills and knowledge in assessment of language skills necessitates certain additional competencies (Inbar-Lourie, 2008). Thus, Alkharusi, Kazem and Al-Musawai (2011) recommends that the instructional assessment training course be provided when teacher candidates are taking instructional practicum during the university education so that they have chances to get feedback regarding their assessment practices in an instructional context.

Kubiszyn and Borich (1996) also suggested that within plenty of other significant assessment concepts, being aware of the kind of assessment required related to the aim of the assessment, identification of goals of the course program, knowledge about essay-test development, validity, reliability, fundamental assessment task statistics, and the conclusions regarding the assignment of scores or grades as an output of instruction and educational measurement are essential issues classroom educators should address (Daniel & King, 2010). To illustrate, for achievement assessments, it is obviously important that the test developers are knowledgeable about what it is realistic and logical to expect learners have actually improved through the course curriculum. Thus, it is possible that good item developers will be the teachers who are experienced of similar learners or related subject areas, who are professionally qualified in the particular setting where the assessment instrument is being constructed (Alderson, et al. 1995). Similarly, Haladyna (2004) proposed that item developers must be experts in the subject matter, preferably in a particular area for which these item

writers will be developing test items. Since, these subject-matter experts will be writing test items, they will be in need of documenting each test item's authenticity of content and justify that there exists one and only one correct answer. In addition, they will also act as expert reviewers of their colleagues' test items.

Considering the fact that a test item involves a method of deriving behaviour or linguistic performance, together with a systematic process whereby that particular behaviour or language performance can be judged (Alderson, et al. 1995), the qualified and skilled item developer must also have a penetrating grasp of how test takers respond to test tasks, involving an understanding of how a certain test item might be mentally processed by different test takers (Osterlind, 2002). Knowing about development of test items will help teachers in being more and better informed and, thus, they will tend to make more reasonable decisions, either as test constructors, as examinees, or as the users of test results (Osterlind, 2002). Hence, by knowing the purpose, definition, and characteristic features of assessment tasks and items, teachers will reach a great amount of information about a specific test item, its development, function, and probable appropriateness. To conclude, informed and skilled test-item constructors are more likely to design better assessment tasks (Osterlind, 2002).

Popham (2010) proposed that one of the most crucial issues regarding modern teaching profession is that the teachers' knowledge of assessment is quite limited. Stakeholders are affected by language assessments, therefore; having knowledge in language assessment is more essential for them, as lack of assessment knowledge brings about inaccurate assessment practices and as a result the purposes of those assessment practices could not be met. As a consequence, it is an obvious necessity for language teachers to acquire language assessment literacy to avoid critical results of them for both teachers and learners (Bayat & Razaei, 2015). Since, the consequences of assessment are not limited to students only, yet impact the classroom teachers and other stakeholders, too (Zolfaghari & Ahmadi, 2016).

The understanding of the different theoretical concepts and practices of summative and formative testing and assessment results is a challenge for a certain number of classroom teachers. Within the scope of a study conducted in the UK by Davies and Taras (2018), it was revealed that there was mostly a good understanding of practical and theoretical concepts of summative assessment within all higher education instructors. However, there was little concordance within the lecturers in terms of understanding the aspects of formative assessment. Similarly, in another study in Canada, DeLuca and Klinger (2010) assessed 288 pre-service teachers' levels of confidence in assessment theory and practices. The findings

revealed that there is a certain need for instruction in assessment because they are not as confident in formative assessment as they are in summative assessment. In addition to these, similar situations may be observed in higher education institutions. The research with academicians including six semi-structured interviews and an examination of assessments conducted by them provided the data and findings suggest that while lecturers have an understanding about summative assessment, they have a low level of awareness of the development and administration of well-prepared formative assessment (Lees & Anderson, 2015).

2.6.1. Studies conducted on EFL teachers' language assessment knowledge

Teachers are generally so pre-occupied with designing classroom activities that they fail to develop a toolbox of comprehensive perspectives of the sequence of the learning the target language from educational objectives to practices to assessment (Jabbarifar, 2009). Research in education has confirmed that most classroom teachers are not professionally prepared to and effective in integrating assessment into their everyday teaching practice (Pastore & Andrade, 2019). Likewise, foreign language educators and school administrators require the necessary tools for interpreting and reflecting on test results to be able to make informed instructional decisions related to their classroom practices and curriculum development (Boyles, 2005). Thus, the place of building knowledge in assessment field has gained an increasing importance within the research scope of English language teaching.

Educators spend almost twenty-five to thirty per cent of their professional time dealing with the work related to assessment even without the necessary training required to conduct it effectively (Stiggins, 2014). While it is crucial for foreign language teachers to know various procedures and tools of assessing learners, plenty of the EFL educators stick to methods of language assessment used in earlier times (Ashraf & Zolfaghari, 2018). However, as Beziat and Coleman (2015) state better equipping teacher candidates for assessment in their classroom will in turn lead classroom teachers to measure learners' proficiency and develop learner success. Unfortunately, while many of today's foreign language educators graduate from their programs of teacher training, there is not any requirement for them to learn anything about assessment in language teaching (Popham, 2009). As a result of this, in recent years, a growing number of programs under professional development practices have provided input for assessment literacy for educators and/or school managers, and certain researchers and those programs have specified the required domains of content in which participants are in need of developing their assessment knowledge and skills (Ashraf & Zolfaghari, 2018). McNamara (2006) advocates an education program which involves a critical examination of assessment

and social results of it, whether those influences concern the instructional area (admission to colleges, acquiring a degree from schools etc.) or society at large (citizenship, immigration, employment etc.). Language assessors also need to develop comprehension of the opposing and usually contradicting forces that play a role between the assessment environments and testing cultures (Inbar-Lourie, 2008). To sum, language teachers are involved in various assessment practices and dealing with those require them to be knowledgeable and effective in assessing learners' language performance in their classroom atmospheres.

The research shows that teachers lack classroom assessment knowledge and skills. Tsagari and Vogt (2017) conducted a study which indicated that most language teachers from seven different countries in Europe reported that they used mostly traditional methods of assessment and they needed to receive training in language assessment. In Allal's (2013) study with 10 sixth grade teachers, many teachers showed a tendency to build their judgement methodically by collection of information while some started with first impression, based on their evaluation of student's work, which they encountered with several different types of evidence. However, the teachers participated in another study did not have a similar level of competency in assessment. In Lan and Fan (2019)'s survey-based research with 344 EFL teachers, it was found that the teachers studied were only at a functional level in terms of their knowledge in classroom assessment. They stated they need to receive more training to develop themselves procedurally and conceptually in language assessment. In a similar study (Baziat & Coleman, 2015) it was found that early childhood and primary/elementary school student teachers are in need of a greater focus on developing assessment literacy. Similarly, in Daniel and King's (2010) study with ninety-five elementary and secondary teachers, it was revealed that the assessment knowledge bases of the teachers for their classroom assessment were inadequate.

Teachers with a better understanding in assessment are likely to become more aware of the needs of their learners and their instructional practices in the classroom. The findings of Ashraf and Zolfaghari's (2018) study revealed the existence of a positive link between English teachers' language assessment knowledge and their practices related to reflective teaching. In other words, when language educators' knowledge in assessment increases, their reflection on their teaching practices increases, too (Ashraf & Zolfaghari, 2018).

The research reveals that language assessors and teachers also struggle with the content domain of language testing and assessment. Baker's (2016) research including the collaboration of admission officers and assessment professionals at higher education institutions provided various perspectives into the language assessment knowledge which was

needed for these related users and the probability of conceptualising language assessment literacy (LAL) for the users of these assessment results as a concept of competence in professional workplace. In parallel with this study, Pill and Harding (2013) investigated kinds of misconceptions related to language assessment within the discourse of authorities who make policies and other related parties during their negotiations about assessment of language proficiency. The findings revealed that there is a certain need for the language assessment communication to have an existence as a field of expertise. In order to present transparent assessment practices to the all parties (e.g. test takers, administrations, policy makers) affected by the assessment procedures and decisions, language test developers need to improve their assessment skills and understanding of related conceptions.

The research has found certain differences between language and non-language teachers' knowledge and practices in assessment. In Jeong's study (2013), it was revealed that language assessors spend significantly more of their time on theory of language assessment while non-language assessors spend more time on assessment in their classroom and test accommodations. When English teachers were compared with other subject matter teachers (in Zolfaghari & Ahmadi, 2016), they were found to be more likely to benefit from continuous (formative) assessment on its own while other teachers prefer a combination of both continuous and final (summative) assessment.

Studies carried out for the measurement of EFL educators' comprehension in productive skills' assessment show that their competency in the area is quite limited. Recent studies have found that language teachers' knowledge of language assessment makes a difference while they assess written performance of their students (Crusana, Plakansb & Gebrilca, 2016; Lam, 2019; Mellani & Khademi, 2018). Lam (2019) carried out a study with 66 secondary school English teachers and found that most of them have merely basic understanding of writing assessment concepts and practices even though they already were in possession of work experience and proper academic backgrounds. However, teachers' having assessment knowledge is likely to make a difference for those teacher assessors. Mellati and Khademi (2018) investigated assessment knowledge of EFL teachers and its influence on their present assessment practices and students' writing outcomes. According to the results of the study, English teachers' language assessment knowledge has a statistically significant influence on students' writing performance and educators' awareness in language assessment leads instructional contexts into more efficient assessment design. In Crusana, Plakansb and Gebrilca (2016), the results indicated that linguistic background and experience of teachers make a difference in teachers' writing assessment knowledge and practices. However, the

studies conducted to measure teachers' assessment knowledge for speaking are more limited. Tajeddin, Alemi and Yasaei (2018) conducted a study in order to investigate assessment knowledge of novice and experienced EFL teachers in assessing speaking and found that both group of teachers need to raise their awareness in their current knowledge of speaking assessment. Firoozi, Razavipour and Ahmadi (2019) conducted a study in Iran and their results also showed that EFL teachers are in need of training on assessing speaking and writing skills. Particularly, they need to gain knowledge in development of assessment scales for speaking as well as writing.

As for the research conducted to assess EFL teachers' assessment knowledge in reading and listening assessment, there are only few studies encountered (Marin, 2009; Ölmezer-Öztürk & Aydın, 2018a). Similar to the findings in teachers' assessment practices of productive skills, teachers' assessment repertoire is not adequate in measuring receptive skills. Marin's study (2009) revealed that teachers have a tendency to use quantitative scales to evaluate students' reading performance qualitatively. Also, teachers' lack of knowledge about methods of alternative assessment and their concerns regarding the justification of their achievement of instructional objectives were among the findings of the study. Another study conducted by Ölmezer-Öztürk and Aydın (2018a) revealed that EFL teachers need more training to improve their knowledge and skills in construction of instruments for assessing listening skill.

Several researchers have attempted to develop new instruments to collect data on language teacher's current levels of assessment knowledge. Edwards (2017) constructed a rubric to monitor pre-service and novice science teachers' summative assessment knowledge. Hudaya (2017), also, developed a 31-item (involving Likert scale, open-ended and background questions) based on principles of language assessment. The results revealed that 79 % of the in-service language teachers felt competent in assessing learners' performance in their teaching context. In addition, Farhady and Tavassoli (2018) developed a scale to measure language assessment knowledge of English language teachers in language assessment. The results of the research indicated that the majority of the EFL teachers have poor knowledge in language assessment. However, a problem about the scales is that the tendency of many instruments was to have a general focus; that is, knowledge of assessment, competencies and practices were not differentiated in relation to the background of the learner or program levels such as elementary, secondary or higher education (Looney, Cumming, Kleij & Harris, 2017).

It has now been demonstrated that language teachers' need for developing their competencies in language assessment might be fulfilled with creating opportunities for them to train themselves in the area. Fulcher (2012) conducted a research project to develop and pilot a survey instrument assessing EFL teachers' needs for training on assessment. The results showed that EFL teachers are generally aware of a range of their assessment needs which are not provided in existing tools and materials constructed with the aim of improving their competency in assessment. In another work with teachers from other subject areas, a two-week in-service training program, was found to have a substantial effect on improving teachers' understanding and knowledge of certain conceptions and processes in relation to the assessment of learner academic performance (Mertler, 2009). Teachers pre-test scores ($M = 19.57$) increased on the post-test ($M = 28.29$). They kept teacher journals throughout this two-week course period. Moreover, Gareis and Grant (2015) designed a similar focused course for teacher candidates to help the participants develop their assessment knowledge. However, it is probable that one single course in testing and assessment will be inadequate to cover everything that educators need to know (Mertler, 2004). To conclude, regarding the crucial role of assessment in classroom teaching, there is obviously a critical necessity for teachers to receive testing and assessment training during their pre-service education. The teachers who have not received such training, it is essential that they gain it as quickly as possible in their teaching environment (McKay, 2006).

2.6.2. Language assessment knowledge studies conducted in Turkish context

The research conducted in Turkish context to gather evidence on assessment knowledge of teachers is only limited to several studies and only few of them were carried out specifically with EFL teachers in Turkey. Ogan-Bekiroğlu and Suzuk (2014) conducted a study with pre-service physics teachers to assess their knowledge in theoretical and practical aspects of assessment. The results of the study revealed that the pre-service physics teachers own a high level of assessment knowledge of the theory; however, they had difficulty in internalising and reflecting in their real assessment practices. In EFL context, some evidence for the assessment knowledge of teachers has been found, too. A study conducted by Yastıbaş and Takkaç (2018) with English instructors at a Turkish foundation university showed that language teachers depend mostly on their students' needs and coursebooks while preparing their exams. The most significant result of this study is that English instructors who are Turkish possessed a critical approach to every phase of designing their tests.

However, most research found that Turkish EFL teachers' assessment knowledge were not as high as expected. Mede and Atay (2017) conducted a study with English instructors at the same context and most of them ($M = 60\%$) stated that they had limited training on classroom assessment. As all the assessment instruments are prepared by a particular office, they pointed out they "had to" use those and assess the production of students according to the scales prepared the same office. However, the teachers both supported standard assessment and they also wanted to have more autonomy in their classroom assessment. In addition, they stated that they do not feel confidence to prepare such tests as they did not get the training, thus they lack of knowledge about the assessment concepts such as validity and reliability. Moreover, Turkish EFL teachers did not consider listening as significant as the other main skills, while they gave the most importance to speaking in their classroom assessment practices (Kirkgoz, Babanoglu & Agcam, 2017). It was revealed that the teachers prefer teacher-oriented assessment practices in their classroom; thus, they mostly evaluate assessment results summatively. The study highlights the need for encouraging the teachers for embodying various and more appropriate assessment styles in their teaching environment. However, Öz and Atay (2017) found that almost all teachers who participated in their study reported that they do not use a score or rubric while evaluating their learners' classroom performance. This indicates that teachers mostly depend on their personal judgments while measuring their students' language performance.

Among the studies comparing teachers' level of assessment knowledge and years of experience, Öz and Atay (2017) found that there is not much correlation between assessment perception and experience of the teachers. Moreover, although the teachers had a certain level of assessment knowledge, they could not reflect that in their classroom assessment practices. Büyükkarcı's study (2016) also showed that Turkish EFL teachers' years of work experience had no effect on their competency in assessment. Also, an interesting finding was that the teachers with an MA degree in ELT were found to be more knowledgeable in language assessment than the ones with a PhD or BA only.

The need for teachers to be knowledgeable in assessment has never been more serious and, regarding current trends in instructional reform, its significance will only continue to increase in the future (Kahl et al., 2013). Hatipoğlu (2015) conducted research with pre-service teachers in order to identify their needs in terms of assessment training and gathered data for their perceptions about the content of their language assessment class. The implications of this study and indicated that the teachers who work in Turkish public schools do not have the basic training in testing and assessment. Similarly, in another study, Haznedar (2012) found out that

only 31.8% of English language teachers had taken a course on language assessment in their undergraduate education. Hence, lack of necessary training results in mismatches between the theory of assessment and teachers' actual practices in the classroom.

The number of studies conducted in Turkish context with the aim of assessing EFL teachers' knowledge in assessing four main language skills is quite limited. As Köksal (2004) put forward evidence, there is a gap between teaching and testing of language teachers and, moreover, teachers tend to assess receptive skills but ignore the written and oral language production of their learners. In parallel with this research, Han and Kaya (2014), in their study with 95 Turkish EFL teachers, found that the teachers considered assessing speaking as the most challenging skill among the assessment of all four main skills. They also reported that they gave less weighting to listening and writing skills than reading in their assessments. Ölmezer-Öztürk and Aydın (2018a) pointed out that EFL instructors lacked required knowledge and skills in assessment of all four main skills of language use. Furthermore, assessment of listening was found to be the most challenging of all four skills.



CHAPTER 3

METHODOLOGY

3.1. Research Design

In the present study, the mixed-methods approach will be utilised. “A mixed methods study would involve mixing within a single study; a mixed method program would involve mixing within a program of research and the mixing might occur across a closely related set of studies.” (Johnson, Onwuegbuzie, & Turner, 2007, p. 123). With the growing popularity in the fields of social and human sciences, including the combination of both quantitative and qualitative methods in its nature, mixed methods approach contributes to the evolution and development of research methodology, using the strengths of quantitative and qualitative research designs (Creswell, 2009). Thus, this research design potentially helps the researchers gain a better insight that he/she would gain from either qualitative or quantitative research on its own.

Weighting is defined as the priority given to either quantitative or qualitative method in a specific research study (Creswell, 2009). Weight emerges depending on the research method that is emphasized first or the extent of implementation of one kind of data or the other in the research study (Creswell, 2009). In the current study, the collection of quantitative data is emphasised first and it is followed by the collection of qualitative data.

Mixed-method research involves particular research strategies. Creswell (2003) identified and argued six kinds of mixed-methods research strategies. One of these strategy types, *sequential explanatory strategy* is the nature of the present research study. It is defined as the gathering and analysing quantitative data in an initial phase of research design and the gathering and analysing qualitative data in the following phase that builds up the findings of the quantitative results in the first phase (Creswell, 2009). Thus, the collection of data in the qualitative phase helps the research interpret the results of the quantitative data (Figure 4).

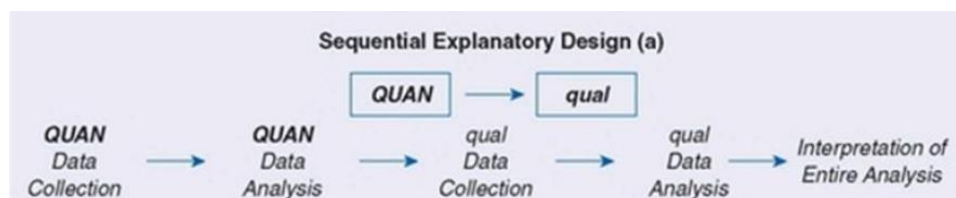


Figure 4. Sequential explanatory design (adapted from Creswell, 2009)

In this study, the data on language assessment knowledge of participants will be collected in a sequential explanatory design. It is expected that the data gathered via qualitative research will lead a more appropriate interpretation of data collected via quantitative research in the first phase. Thus, it is assumed that, having a mixed-method sequential explanatory design, the findings of the current study will build on the present research on EFL instructors' knowledge of language assessment.

3.2. Research Sites and Participant Selection

The research was conducted at school of foreign languages (SFLs) (preparatory school) at four universities in Turkey. The researcher assigned letters A, B, C and D to these universities in order for confidentiality. In SFLs in university A, there were 1512 students and 53 instructors work there as full time EFL educators. There were 3143 students in SFLs and 125 full-time EFL instructors work in university B. In university C, 1351 students studied at SFLs and 48 EFL instructors work there. Lastly, 137 EFL instructors teach and approximately 2000 students study at SFLs in university D. The data given above were retrieved from the official websites of the related universities (on 9.4.2020). The students generally receive English instruction from 20 to 25 class hours in a week in these institutions. The assessment materials are generally developed and administered by an assessment unit. The EFL instructors for these units are generally selected based on their interest, experience and education in assessment. However, when necessary any EFL instructor could be expected to prepare assessment instruments and become an active member of the assessment unit even if they do not have any experience and interest in language assessment. On the other hand, whether they prepare tests or not, generally all the EFL instructors take active roles in test administration, classroom assessment, test marking and evaluation processes.

In Turkey, university students who are not proficient enough in English need to study English preparatory school program for one year so that they can follow their lectures and use English effectively to meet the objectives of their courses in their major. In school of foreign languages (or preparatory schools), EFL learners are provided with general and academic English skills in order to prepare them for the upcoming courses provided in English language. The content of the preparatory school programmes is quite intense and learners are exposed to English around 20 hours per week in classroom settings. Learners mostly take a placement test at the beginning of the academic year and they complete the education with a certain level of English. The levels and course objectives are generally defined according to CEFR (Common European Framework). Thus, language assessment tasks are required to be prepared professionally and the results are to be interpreted correctly.

In the present research, convenience sampling was utilised while determining the participants of the study. “A convenience sample is a group of individuals who (conveniently) are available for study” (Fraenkel, Wallen & Hyun, 2012, p. 99). The target population of the study is the EFL instructors that work as regular academic staff in SFLs (or preparatory schools) in four state universities in Turkey. The state universities at which the participants work were selected according to their conveniently availability. The four state universities at which the study was conducted are all in western part of Turkey.

114 EFL instructors participated in the quantitative phase of the study. The demographic information about the participants is given in table 1. 83 of the participants were female. There were 7 international EFL instructors, two of whom were from the USA and each of the others were from Russia, Romania, Ukraine, Belarus and Egypt. Most of the participants were above 30. Moreover, a majority of the participants had an MA or PhD degree. As for the BA programs, almost half of the participants were graduates of an English Language Teaching (ELT) program. The rest of the participants generally graduated from an English Language and Literature (ELL), American Culture and Literature or Translation and Interpretation programs. A half of the participants had experience in language assessment. Also, similarly, a half of them received some sort of training in assessment in their career before.

Table 1.

Demographic information about the participants

N of participants	114
Gender	83 F / 31 M
Nationality	107 Turkish / 7 Other
Age	22-30 (12) / 31-40 (50) / 41-above (38)
Educational Background	BA (46) / MA (56) / PhD (12)
BA Programme	ELT (52) / Non-ELT (60)
Experience in Assessment Unit	Yes (58) / No (56)
Training in Assessment	Yes (56) / No (58)

8 EFL instructors participated in the qualitative phase of the study. The eight EFL instructors (six female and two male) in the study have been working for several years at

schools of foreign languages at a state university. Ahmet is pseudonym that was given to the male participant and the other names were selected and assigned to the female participants by the researcher. The ages of the participants' ranged from 30 to 50 years old and their work experience ranged from 8 to 24 years. Fulya had a BA degree in Translation and Interpretation program and Ahmet graduated from an ELL program. The rest of the participants graduated from an ELT program. Except for Öznur, Ali and Yeşim, all the other participants had a degree in an MA program. As for their MA programs, Nermin and Fatih had a master's degree in ELT, Fulya completed an MA program in Translation and Interpretation program. Deniz had her MA degree in Business Administration and Hilal's MA degree was in Journalism and Broadcasting. All the EFL instructors, except for Ahmet, had experience in an assessment unit in their institutions. Likewise, except for Ahmet, all the participants had some training in language assessment Table 2 shows the details about each EFL instructor who participated in the qualitative study.

Table 2.

Relevant characteristics of the EFL instructors who participated in the interviews

Theme	Öznur	Nermin	Ali	Fulya	Yeşim	Deniz	Fatih	Hilal
Year of experience	24	8	12	9	20	18	14	15
Educational background	BA	MA	BA	MA	BA	MA	MA	MA
BA program graduated	ELT	ELT	ELL	Translation & Interpretation	ELT	ELT	ELT	ELT
Assessment unit experience	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Training (in language assessment)	A few day inservice	Assessment course in BA/MA program	None	3-day inservice training	A few day inservice	In-service training	Pedagogical education courses	A few day inservice

3.3. Instruments

In the first phase of the present study, Language Assessment Knowledge Scale (LAKS) developed by Ölmezer-Öztürk & Aydın (2018b) was used to measure the language

assessment knowledge of the participants (Appendix 1). LAKS was validated by collecting the opinions from language teachers and academicians as the experts and a meeting with test developers was held (Ölmezer-Öztürk & Aydın, 2018b). LAKS involved 60 items with four separate sections including 15 questions related to the assessment of reading, listening, writing and speaking main skills in each section. The participants were presented with three options – *True*, *False* or *Don't know* – next to each item and they were expected to choose the option that they consider correct based on their background language assessment knowledge. The reliability of the scale was calculated and it was found reliable ($R=.75$) (Table 3).

There are four sample items (one from each part of the scale) below:

- a. *Simplification of reading texts is avoided. (Item 12)*
- b. *Any type of listening text is used for note-taking. (Item 18)*
- c. *Using visuals which guide learners for writing poses a problem. (Item 38)*
- d. *In a speaking exam, production and comprehension are assessed together. (Item 55)*

Table 3.

Reliability analysis of the LAKS

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,757	,757	60

In the second phase of the study, three interview questions were developed by the researcher with the intention to investigate the background and motivation of the 10 high scorer participants based on their LAKS scores. However, after getting the quantitative results, the number of the participants was modified. The aim of the second phase was to collect deeper information about the tools and methods that are used by the high scoring EFL instructors. Moreover, their suggestions regarding the raising the level of language assessment of EFL instructors were aimed to be collected. With this aim, 8 high scoring English instructors were visited in their institutions that they work for and the interviews were recorded with a voice recorder.

The first question was related to the participant's practices which helped them gain their present language assessment knowledge. The second question was related to the participants beliefs and their ideas on potential advantages of possessing knowledge and skills in language assessment in terms professional development of the teachers and students' learning. The last question included their suggestions for other EFL teachers to raise their awareness and knowledge in language assessment.

3.4. Data Collection Procedures

In first phase of the data collection process, the researcher applied to the ethics committee of the Institute of Educational Sciences and got approval for the study. After that, permission from administration of the study was received from the four state universities. Then the researcher visited the participants at the universities they work and they were informed about the purpose and content of the study during face-to-face gatherings. Language Assessment Knowledge Scale (LAKS) was distributed to the voluntary EFL instructors and they filled out the scale on paper. During the administration of the scale, it was made sure that each participant filled out the scale on his/her own without interacting other colleagues. The participants were requested to write their names on the scale to be able to reach the high scorers in the second phase.

In the second phase, three interview questions were constructed with the aim of collecting data on the participants language assessment practices, beliefs and suggestions related to language assessment. 8 participants were selected based on their scores on the LAKS and the threshold score was defined as 42 correct answers out of 60 questions, with the rate of 70 per cent success. All 8 participants participated in the study voluntarily. One-to-one interviews with 8 high scoring participants were held in their institutions. All the interviewees were asked three main questions and when necessary the researcher asked related follow-up questions to lead the interviews into the appropriate context. The interviews were recorded with voice recorders and the typescripts were created by listening to the audio recordings.

3.5. Data Analysis

3.5.1. Quantitative research analysis

During data analysis process of the quantitative research section, the data collected via LAKS were transferred from the paper to the SPSS 22. The reliability of the scale was measured with reliability analysis and found as .75. The data collected from LAKS were analysed using descriptive statistics in the first phase to be able to describe the general and skill-based LAK of EFL instructors. Moreover, the scores were compared and analysed across

different language skills and the high scoring participants were selected based on the total scores of the participants on the scale. To analyse the impact of gender and assessment unit experience, two-way ANOVA test was conducted. In addition, the existence of correlation between the participants' total scores and their year of work experience was analysed via Pearson correlation analysis.

3.5.2. Qualitative research analysis

In the analysis of the qualitative part, the tapescripts of the interviews with 8 high performers were analysed through transcribing and coding. The process of coding involves organising the textual data into chunks of text before deriving meaning from information (Rossman & Rallis, 1998). In this way, a coherent description of the findings of the qualitative part was made. To increase reliability, another researcher also analysed the data and the results were compared. The steps of textual data analysis suggested by Creswell (2009) were followed.

- Step 1. The data were prepared and organised. (This step involves transcribing the semi-structured interviews and arranging them according to the participants and their answers given to each question.)
- Step 2. All the data were read to get the general sense and understand the overall meaning.
- Step 3. The coding process began for a detailed analysis.

Three questions were directed to the participants who are highly knowledgeable in language assessment, based on LAKS results. The semi-structured interview questions are presented below:

1. *Can you explain how you acquired your current knowledge in language testing and assessment?*
2. *What do you think is the best method or tool for EFL instructors to improve their awareness and knowledge in language assessment? Why?*
3. *What do you think are the possible advantages of having knowledge in language assessment to you as an EFL instructor?*

Quantitative results were presented through tables and qualitative data were organised under subheadings *Language Assessment Practices, Beliefs and Suggestions*. The findings of

the study were presented as descriptive text with quotations from participants. The results were discussed and suggestions were made for EFL teachers and related institutions.



CHAPTER 4

RESULTS

The aim of the present study was to investigate EFL instructors' current language assessment knowledge level. It was also aimed to collect deeper information about the practices and suggestions of the EFL instructors who are highly knowledgeable in language assessment. The current study yielded certain significant results in relation to Turkish EFL instructors' language assessment knowledge levels. 114 EFL instructors participated in the study. Each participant answered the questions in the Language Assessment Knowledge Scale (LAKS) (Appendix-1) individually. Since there was only one correct answer in the scale, 1 point was assigned to each correct answer, 0 point for the wrong answer and 'don't know' answers were also analysed separately. The data were analysed using SPSS 22 statistics program. The descriptive analysis results of total scores acquired from the scale and the scores acquired from each part related to assessment of one main skill were illustrated. Also, the findings of two-way ANOVA and Pearson Correlation analysis were described to reveal the influence of gender, work experience and experience at an assessment unit. Finally, the qualitative part analysis was made to find out the practices, beliefs and suggestions of the participants who got high scores from LAKS.

4.1. What is the EFL instructors' current level of language assessment knowledge in terms of general and skill-based assessment?

The first research question explores the EFL instructors' language assessment knowledge in general and skill-based assessment of language. Table 4 illustrates the mean of the participants' total scores that they received from the LAKS.

Table 4

The mean of total scores of the participants

Total Scores

N		Mean	Std. Deviation	Minimum	Maximum
Valid	Missing				
114	0	33,1667	6,28232	14,00	53,00

Table 4 shows that 114 participants attended the study with no missing data. The maximum score to be acquired from the scale was 60. According to the analysis of the participants' answers to the scale, the mean of the general scores was 33,16 (SD=6,28). The maximum score received was 53 and the minimum was 14. The findings show that the general language assessment knowledge of the participants was low.

Current level of language assessment knowledge in terms of skill-based assessment was also analysed. Scores for four skills were measured based on the correct answers of the participants. Each sub-section of the scale included 15 items related to the assessment of each main skill (reading, listening, speaking and writing).

Table 5

The mean of each sub-section of the LAKS

	N	Minimum	Maximum	Mean	Std. Deviation
Reading TOTAL	114	3,00	15,00	10,1316	2,31009
Listening TOTAL	114	1,00	12,00	7,0965	2,36856
Writing TOTAL	114	1,00	13,00	7,4561	2,25436
Speaking TOTAL	114	3,00	14,00	8,5175	2,15436
Valid N (listwise)	114				

Table 5 shows that the mean of reading assessment section is 10,13, while the minimum score is 3 and the maximum score is 15. The mean of listening assessment scores is 7,09. The minimum score acquired from this part is 1 and the maximum score is 12. Writing assessment part scores are quite similar to the listening part. The mean of the scores for writing assessment is 7,45 and the minimum score is 1 while the maximum score is 13. According to the results, speaking assessment mean score is 8,51. The minimum score in assessing speaking is 3 while the highest score is 14. Thus, these findings show that the participants' performance in assessment of listening was lower compared to the other three skills. The mean score of the writing assessment section is only slightly higher than the mean score of listening assessment section, which was followed by speaking assessment mean score. These results show that the participants' knowledge level was low in all these three main skills while this level is only moderately high in reading assessment.

The responses of the participants given to each item of the scale were analysed (Table 6). Table 6 shows the number of correct, incorrect and ‘don’t know’ answers. The number of responses given to each item were presented next to each item and the correct answers were written in bold letters. The letters in front of the item numbers represent the sub-section of the scale. The sub-sections were reading, listening, writing and speaking assessment.

Table 6

Descriptive results of the LAKS items (Bold ones refer to the participants with correct answers)

Assessment of Reading	N	True	False	Don't Know
R1. Asking learners to summarize the reading text is a way of assessing their reading skills.	114	97	15	2
R2. When asking several questions about a reading text, all the questions are independent of each other.	114	50	57	7
R3. Cloze test is used for assessing the main idea of the text.	114	24	70	20
R4. In a reading exam, using a text learners have encountered before is not a problem.	114	33	68	13
R5. One reading text is enough to be included in a reading exam.	114	12	86	16
R6. The language of the questions is simpler than the text itself.	114	83	24	7
R7. Errors of spelling are penalized while scoring.	114	19	86	9
R8. Taking vocabulary difficulty into consideration is necessary in assessing reading skills.	114	99	12	3
R9. Including not stated/doesn't say along with true/false items has advantages over true/false items.	114	71	23	20
R10. The more items a reading text is followed, the more reliable it becomes.	114	49	32	33
R11. Using the same words in the correct option as in the text is not a problem.	114	28	71	15
R12. Simplification of reading texts is avoided.	114	26	60	28
R13. Reading texts in a reading exam include various genres (essay, article, etc.).	114	105	1	8
R14. In top-down approach, assessment is on overall comprehension of the reading text.	114	76	12	26

R15. Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.	114	84	13	17
Assessment of Listening	N	True	False	Don't Know
L16. Using reading texts for listening purposes poses a problem.	114	54	43	17
L17. Including redundancy (e.g. what I mean to say is that) in a listening text poses a problem.	114	20	85	9
L18. Any type of listening text is used for note-taking.	114	32	72	10
L19. Spelling errors are ignored in scoring the dictation.	114	29	81	4
L20. Errors of grammar or spelling are penalized while scoring.	114	50	54	10
L21. A listening cloze test is a way of selective listening.	114	71	16	27
L22. Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative testing.	114	38	16	60
L23. Scoring in note-taking is straightforward.	114	39	35	40
L24. In discrete-point testing, comprehension is at the literal/local level.	114	39	6	69
L25. Using dictation diagnostically in assessing listening skills does not pose a problem.	114	42	33	39
L26. Giving learners a transcript of the listening text is a valid way of assessing listening skills	114	18	79	17
L27. Dictation is a kind of discrete-point testing.	114	44	24	46
L28. Inference questions based on intelligence are avoided in listening tests.	114	68	33	13
L29. Asking learners to listen to names or numbers is called intensive listening.	114	57	36	21
L30. In selective listening, learners are expected to look for certain information.	114	107	2	5
Assessment of Writing	N	True	False	Don't Know
W31. Giving two options to learners and asking them to write about one ensure reliable and valid scoring.	114	80	22	12
W32. Analytic scoring is used to see the strengths and weaknesses of learners.	114	89	4	21
W33. The parts of a scoring scale and the scores in each part do not change for different levels of learners.	114	31	69	14
W34. When there is a disagreement between the scores of the two raters, they score the written work again.	114	79	28	7

W35. Learners are required to write about at least two tasks in the exam rather than one task.	114	22	81	11
W36. Giving restrictive prompts/guidelines to learners for the writing task is avoided.	114	28	77	9
W37. Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.	114	89	15	10
W38. Using visuals which guide learners for writing poses a problem.	114	11	88	15
W39. Holistic scoring is used to see whether the learner is proficient or not at the end of the term.	114	65	14	35
W40. Analytic scoring leads to greater reliability than holistic scoring in writing.	114	72	11	31
W41. In controlled writing, learners have the chance to convey new information.	114	34	60	20
W42. Classroom evaluation of learning in terms of writing is best served through analytic scoring rather than holistic scoring.	114	56	26	32
W43. Irrelevant ideas are ignored in the assessment of initial stages of a written work in process writing.	114	24	75	15
W44. Providing a reading text for writing is a way of assessing writing skills.	114	54	30	30
W45. Mechanical errors (e.g. spelling and punctuation) are dealt with in the assessment of later stages of a written work.	114	52	51	11
Assessment of Speaking	N	True	False	Don't Know
S46. When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.	114	62	47	5
S47. Giving learners one task is enough to assess speaking skills.	114	7	103	4
S48. Interlocutors' showing interest by verbal and non-verbal signals poses a problem.	114	25	79	10
S49. When it becomes apparent that the learner cannot reach the criterion level, the task is ended.	114	45	48	21
S50. Using holistic and analytic scales at the same time poses a problem.	114	37	36	41
S51. Reading aloud is a technique used to assess speaking skills.	114	24	68	22
S52. In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.	114	82	18	14

S53. In interactive tasks, more than two learners pose a problem.	114	40	59	15
S54. The interlocutor gives the score when the learner is in the exam room.	114	35	74	5
S55. In a speaking exam, production and comprehension are assessed together.	114	95	12	7
S56. Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills.	114	26	77	11
S57. Discussion among learners is a way of assessing speaking skills.	114	109	2	3
S58. A checklist is a means of scoring oral presentations in in-class assessment.	114	90	12	12
S59. When the focus is to assess discourse, role plays are used.	114	84	10	20
S60. In peer interaction, random matching is avoided.	114	36	60	18

According to table 6, maximum number of participants (109) answered the item 57 (*Discussion among learners is a way of assessing speaking skills.*) correctly. On the other hand, minimum number of participants (15) (*Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.*) answered the item 37 correctly. The numbers in bold indicate the correct answers. In total, 25 items were answered either incorrectly or given the answer ‘don’t know’. These suggest that particular content areas to which a low number of correct answers were given may require more attention in order to identify the gaps in the participants language assessment competency. Four items which received least correct answers are presented below:

- Item 37. *Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing skills.* (15 – correct) (Writing assessment)
- Item 22. *Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative testing.* (16 – correct) (Listening assessment)
- Item 31. *Giving two options to learners and asking them to write about one ensure reliable and valid scoring.* (22 – correct) (Writing assessment)
- Item 35. *Learners are required to write about at least two tasks in the exam rather than one task.* (22 – correct) (Writing assessment)

4.2. Does LAK level change according to gender and assessment unit experience of EFL instructors?

In this section, the effects of gender and assessment unit experience – whether having experience in an assessment unit in a preparatory school or not – on LAKS performance were analysed.

Table 7

The number of female/male participants and participants with/without assessment office experience

Between-Subjects Factors		
		N
Gender	Female	83
	Male	31
Assessment Unit	No	56
	Yes	58

Descriptive statistics were conducted to illustrate the number of male and female participants and participants with and without assessment unit experience. Table 7 indicates that there were 83 female and 31 male participants in the study. On the other hand, almost half of the participants (58) had an experience in an assessment unit in their institution while the other half of them stated that they had not had an experience in an assessment unit before.

Table 8

The mean scores of female/male participants and participants with/without assessment office experience

Dependent Variable: TOTAL

Gender	Unit	Mean	Std. Deviation	N
Female	No	31,3056	6,01341	36
	Yes	34,2128	6,26563	47
	Total	32,9518	6,28956	83
Male	No	31,8000	6,74030	20
	Yes	37,2727	3,55221	11
	Total	33,7419	6,32965	31
Total	No	31,4821	6,22602	56
	Yes	34,7931	5,94642	58
	Total	33,1667	6,28232	114

According to the results (table 8), male participants' mean score (M=33,16) is slightly higher than female participants' (M=32,95). The mean score of participants who have an assessment unit experience is 34,79 while the means of the ones who have no experience in an assessment unit is 31,48. The lowest mean score in this group belongs to female participants who had not assessment unit experience in their background (31,30). However, male participants with an assessment unit experience has the highest mean score (37,27). The statistical significance of these results is illustrated below.

Table 9

Tests of between-subjects effects

Dependent Variable: TOTAL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	398,940 ^a	3	132,980	3,602	,016	,089	10,806	,780
Intercept	95359,128	1	95359,128	2583,054	,000	,959	2583,054	1,000
Gender	66,506	1	66,506	1,802	,182	,016	1,802	,265
Unit	369,668	1	369,668	10,013	,002	,083	10,013	,880
Gender * Unit	34,648	1	34,648	,939	,335	,008	,939	,160
Error	4060,893	110	36,917					
Total	129863,000	114						
Corrected Total	4459,833	113						

a. R Squared = ,089 (Adjusted R Squared = ,065)

b. Computed using alpha = ,05

A two-way between-group analysis of variance was conducted to explore the impact of gender and assessment unit experience on LAKS. Participants were divided into two groups according to assessment unit experience. Levene's test (.36) indicated homogeneity of variance. There was a statistically significant main effect for Unit [$F(1,112) = 10,01, p = .00$]. However, there was not a statistically significant main effect for gender [$F(1,112) = 1,80, p = .18$] and interaction [$F(1,112) = .94, p = .34$]. This meant that males and females did not differ on their LAK. However, there was a difference in scores for participants who had assessment unit experience and the one who did not have that experience.

4.3. Is there a relationship between EFL instructors' current level of language assessment knowledge and their year of work experience?

Pearson correlation test was conducted to investigate the correlation between the participants' years of work experience and their level of language assessment knowledge (Table 10).

Table 10

Correlation between years of work experience and mean of scores

Correlations			
		Experience	TOTAL
Experience	Pearson Correlation	1	-,043
	Sig. (2-tailed)		,648
	N	114	114
Total Score	Pearson Correlation	-,043	1
	Sig. (2-tailed)	,648	
	N	114	114

The results show that there is no statistically significant correlation ($p < .05$) between language assessment knowledge level and year of work experience. As the correlation coefficient is just $-.04$, it shows us that work experience does not have any value in a predictive sense on the instructors' levels of language assessment knowledge. This finding reveals that these two variables are statistically independent of each other.

4.4. Which practices, suggestions and beliefs of the EFL instructors who are highly knowledgeable in language assessment possess?

This section reports the findings of a qualitative research study which investigated the language assessment practices, suggestions and beliefs of EFL instructors who are knowledgeable in language assessment. The participants were assigned pseudonyms. The findings from the study are grouped according to the main themes of each interview questions which include the participants' practices, beliefs and ideas on possible advantages of language assessment knowledge. Each group embraces findings related to the theme and all these

findings were acquired from the tape scripts of the interviews. Taken together, the analysis results provide suggestions and identify areas for improvement in assessment of language skills.

4.4.1. Background and language assessment practices

This section involves the practices that EFL instructors did for English language assessment during their regular classroom instruction or as a group of test developers under the name of an assessment unit. This part also aims to explore the background of the participants related to language assessment area.

The analysis of the qualitative data showed that one common idea shared by all the participants was that it is necessary to have a personal interest in assessment for the constructions of effective assessment processes. Four participants stated that they like preparing assessment instruments and developing themselves in this area. The other half of the participants implied that they have an interest in the area by mentioning their own extra-curricular assessment practices. Some comments are:

“Most of the time, I read and studied by myself and I was interested in the area.”
(Fatih)

“It is important to like it. Not everyone can do this. Thus, willingness is quite important. I think this job requires a high level of passion.” (Deniz)

The data indicated that the seven language instructors either took active parts in assessment units in their background or are still members of this kind of unit now. Meanwhile, five participants reported that they received feedback from either another test developer or a more experienced instructor who developed plenty of tests before. Two participants pointed out that they improved their language assessment knowledge and competency via feedback. The following comments were made:

“In my first year in the assessment office, I didn’t prepare any tests. I just observed the other test developers who actively construct test tasks. They were proofreading each other’s exam tasks and giving feedback to each other. While observing them, I paid attention to the areas they focus on and how they approach and criticise a test. During this period, I’ve learnt a lot related to the test development process.

...

When I started developing the assessment tasks, we started a new cycle. After the construction of a test, we were submitting it to our assistant director who was a quite experienced teacher and worked in the assessment unit for years. I could say that I learnt a lot from her and gained more experience in time and started to make less mistakes while preparing my exams.” (Fulya)

“We also worked together with my colleagues in our assessment unit. I think exchanging ideas and knowledge is an important part of developing your knowledge in assessment.” (Deniz)

“What I mean by experience is when I was working in the assessment unit some of the other EFL instructors were more experienced and knowledgeable than us and I learned a lot from them.” (Hilal)

The data analysis shows that a common practice among the four instructors was the development of classroom materials that include similar tasks that the students may encounter in their exams.

“During the first years of my career, I only focused on the parts in the course materials but when my students received very low grades from the first exam, I figured out that I needed to change things. Then I started preparing exam-like materials for my students. I was preparing close-test, paragraph type and other types of items to make my students familiar with the test tasks.” (Ahmet)

Four participants reported that they gained the skill of developing valid tests by analysing the content of previous exams. Some of the comments were presented below:

“When we were developing our tests in the office, we were brain-storming and trying to figure out the principles and key features of various types of questions by analysing some national and international proficiency exams such as toelf or kpss (Turkish national exam for testing English proficiency.).” (Öznur)

“From time to time, I analysed the proficiency exams of other universities. I both looked at our exams on the school website and the ones on other schools’ websites.” (Ahmet)

From the findings, four participants stated that the course on testing and assessment of foreign language that they received in their undergraduate program was useful. Two of them

highlighted the importance of practical content of their assessment course. As for the content of it, they made the comments below:

“Our language assessment course was very useful in our BA program. Our professor was very qualified. We worked on sample assessment instruments. For example, we were trying to find answers to some questions like Does this item assess our target skill? Or How can we revise and improve it?... We did a lot of practices by critically analysing the sample good, and fair examples.” (Nermin)

“We received education on assessment during my undergraduate years. In our assessment course, we focused on areas such as how a good test is prepared and how these materials are marked and evaluated. We learned such things then.” (Yeşim)

In addition to these, three participants reported they developed their knowledge in assessment theories and terminologies via courses they received during their undergraduate education. The comments below were made:

“First of all, I graduated from an ELT department so I learned some aspects of assessment in those years. ... I mean I learned most of these with the help of my lecturers and my own efforts while studying.” (Fatih)

“I think there are two aspects of assessment, theoretical and practical aspects. I learned the theoretical part of it in a faculty of education in my ELT department years ago.” (Deniz)

Three of the participants pointed out that the in-service training programs were quite effective in improving their knowledge in language assessment. The comments below were made:

“Also, when I started working at a university, I mean when I started work as an EFL instructor, we received in-service trainings.” (Yeşim)

“During summer period, we received an in-service training. Some of its content was related to language assessment. We worked on our own exams and it was a question-answer kind of a training. We examined sample assessment materials and the trainer informed us about the parts done correctly and the ones that need to be changed. The trainer gave us advice about our practices.” (Hilal)

On the other hand, one of the participants stated that the training she received related to assessment terminologies and theories was not helpful as she could not improve her practices after the training. The comments made by the participants was presented below:

“I attended two webinars and that was all. I only learned some terms from them. I did not take any training on assessment on practical level.” (Fulya)

Three participants highlighted the importance of work experience in developing knowledge in assessment. The comments below were made:

“In fact, even if you have broad knowledge in assessment theories, we learn a lot in time by gaining experience. As the student profile changes, you learn what to adjust in your assessment techniques and methods, as well. And you learn how to do this in time.” (Deniz)

Two participants reported that they asked themselves certain questions to raise their own awareness related to their assessment practices. They reported that they tried to find answers to those questions to be able to make sure that they are on the right track while assessing their learners. Some comments were presented below:

“I always think about these... What do I want to assess? And How can I assess it?” (Nermin)

“Especially, is this the right question? Does it assess what I want to assess? Or is there another way of assessing it more effectively? Can the student answer this question based his/her world knowledge?” (Fatih)

To summarise, the results of the qualitative research indicated that the participants who received high scores from the LAKS possess certain common practices and backgrounds. First of all, it was found that all the participants possessed a personal interest in language assessment field. Also, majority of the have experience in an assessment unit and they stated that receiving feedback during the process of test development has a value. A half of them also stated that they analyse various types of assessment instruments and construct classroom materials with the aim of familiarising their learners with the upcoming assessment processes. Moreover, half of the participants highlighted the place of theoretical and practical knowledge they gained during their undergraduate teacher education program. Three of the EFL instructors reported that the in-service training they received after they started working as a teacher was useful. Furthermore, three participants pointed out that work experience helped them gain a better understanding in language assessment. Finally, two participants reported

that they keep themselves aware of their assessment practices by asking themselves certain questions.

4.4.2. Beliefs

This question was designed to collect data on assessment literate EFL instructors' beliefs about the possible advantages of possessing knowledge and related skills in language assessment. All the participants highlighted the impact of being knowledgeable in assessment on classroom instruction and learning.

"If you know how to assess and evaluate, it will be easy for you to detect the needs of your learners." (Fulya)

"When you have knowledge about assessment, then your classroom tasks reflect what is tested or vice versa... You can teach learners critical thinking skills as you know how to direct your learners." (Ahmet)

"In this case, you can develop classroom materials that really serve the purposes of the instruction. For example, you ask questions that help your students grasp the underlying meaning in a piece of text." (Nermin)

Another perspective shared by the knowledgeable EFL instructors was that construction of good assessment tasks require expertise. Thus, the common idea was that there must be assessment experts among the teachers who would develop tests and the other teachers could only develop basic level assessment skills. Some related comments are:

"I think that every teacher should not prepare test materials. Being a member of an assessment unit requires expertise. You should be expert and internalise the field and so that it would be a part of your career." (Öznur)

Three participants reported that EFL instructors' awareness should be raised in relation to their own language assessment knowledge and skills. Thus, they can receive feedback on the effectiveness of their classroom practices.

"The teacher should be aware of this. Otherwise, s/he may not understand how successful his/her students were in an exam or how much they have learned." (Ahmet)

“Some teachers do not assess the skills of their learners with the appropriate method and material but they are not aware of it. If they improve their knowledge in the area, we can prevent such things.” (Fatih

To summarise, the data collected from the interviews showed that the participants believed the positive influence of language assessment knowledge in language classrooms. It is considered that the assessment knowledge of the EFL instructors help them to identify the needs of their learners, guide them more effectively and develop the classroom materials that serve the purpose of the instruction. Moreover, the participants stated that they considered assessment is an area of expertise so the assessment instruments should be developed by a specific group of EFL instructors. Lastly, three participants stated that the teachers need to be aware of their strengths and weaknesses in language assessment area.

4.4.3. Suggestions

This section aimed to investigate the suggestions and advice of knowledgeable EFL instructors in language assessment in relation to the methods to be conducted to improve EFL instructors' competency and skills in assessing foreign language skills. All the participants suggested that EFL instructors need to receive training on assessment which involves practical applications rather than presentation of theoretical conceptions. Some comments are:

“All the teachers need to get training. This is a must. Test development or language assessment can only be improved by training!” (Öznur)

“The school administration must identify the areas that require improvement in EFL teachers' competency and arrange in-service training sessions on language assessment.” (Ahmet)

“External teacher trainers could be invited to school or the instructors who have more experience in test development can organise training.” (Fulya)

The content of the training sessions on language assessment should be organised based on the needs of the institution and learner profile in the related context. The comments of one of the participants on the issue is presented below:

“...but when an external trainer is invited, he/she generally cannot provide an effective session because they don't know the institutional needs or student profile. For this reason, I think an EFL instructor experienced in language assessment should give this kind of a training.” (Fulya)

Another suggestion made by three participants was that the EFL instructors could take active roles in the processes of assessment materials development. The participants believe that competency and skills in language assessment could be improved by preparing the related instruments. Some comments made were presented below:

“In one way or another, the knowledge and skills in this area can only be developed by preparing assessment tools.” (Yeşim)

“It is difficult to learn this without teaching. It is necessary to prepare the test materials for your students and evaluate and see results.” (Fatih)

Three participants highlighted the value of collaboration and mentoring systems. They stated that the teachers who have inadequate knowledge in language assessment could be trained. Some comments are:

“There could be a program similar to internship. Thus, assessment unit members could inform the new teachers about the current practices in the institution. Then we may evaluate the understanding of these teachers.” (Yeşim)

In addition to these, three participants highlight the importance of giving and receiving feedback while constructing assessment instruments. They reported that this is one of the most effective methods of improving one’s language assessment knowledge and skills. They stated that they learned with the guidance of the instructors that they received feedback from. The comments of these participants are below:

“I think this can be developed with pair work. For example, when I start designing an assessment instrument and struggle during the process, one of colleagues should give me feedback. I should receive feedback. I think the product is not important here but the process is.” (Fulya)

Another suggestion given by two participants was that the EFL instructors could read about the trends and recent research studies in language assessment in literature. In this way, they believe that the teachers could keep themselves informed and up-to-date. Some comments of them are:

“On an individual basis, the instructors can read textbooks or articles. They can read about both theoretical and practical aspects of assessment.” (Fulya)

“There are numerous resources in literature related to language assessment. Thus, one should like reading.” (Fatih)

Another suggestion shared by two participants was that the learner scores gained from each assessment process could be analysed statistically and the findings could be shared with all the classroom instructors. Also, according to these participants, assessment knowledge and practices of the teachers should be assessed. Some of the comments are:

“Statistical analysis of assessment results could be made and shared with all the teachers. We can revise our course program based these findings.” (Hilal)

To sum up, the participants of the semi-structured interviews suggested certain practices in order to help the EFL instructors and school administrators to improve language instructors' knowledge of language assessment. The participants highlighted the significance of training in assessment of which the content is advised to focus more on practical aspects rather than theoretical sides of assessment. Also, they stated the importance of feedback received during the test development processes. Moreover, the EFL instructors should be informed about the recent trend and developments in assessment field, according to the participants.

CHAPTER 5

DISCUSSION AND CONCLUSION

This mixed study mainly aims to shed light on the present knowledge level of EFL instructors in language assessment. The study also aims to investigate the impact of gender, years of experience and assessment unit experience on English language instructors' language assessment knowledge. Moreover, the qualitative part of the research aims to find out the language assessment practices, beliefs and suggestions of the EFL instructors who received high scores in LAKS. So as to find out these, the participants for the study were selected by convenience sampling method from four state universities in Turkey. 114 EFL instructors participated in the quantitative part of the study. They were all visited in their institutions, informed about the study and requested to participate in the study. 8 high-scoring instructors participated in the face-to-face semi-structured interviews.

The Language Assessment Knowledge Scale (LAKS) developed by Ölmezer-Öztürk and Aydın (2018b) was adapted and used to measure the language assessment knowledge of the participants. The scale involved both demographic information questions and 60 true/false items which include 15 items related to the assessment of each main language skill (reading, listening, writing and speaking). In addition to 'true' and 'false' options, 'don't know' option was also presented to distinguish the participants who had incorrect knowledge about a specific area from the ones who had no idea or knowledge about it. In the second phase of the study, three interview questions constructed by the researcher were used to collect deeper information within the scope of the study.

The findings of the quantitative part were analysed by descriptive statistics, two-way ANOVA and Pearson Correlation tests. In the qualitative part, two researchers analysed the data through transcribing and coding. The findings were compared and the final results were discussed.

5.1. Discussion

The main research question of the present study was constructed to find out the language assessment knowledge level of EFL instructors working at state universities in Turkey. The next two research questions were designed to find out the impact of gender, assessment unit experience and year of work experience on language assessment knowledge of the EFL instructors. Finally, the last research question was designed to investigate the background practices, beliefs and suggestions of the EFL instructors who gained high scores in LAKS.

5.1.1. The language assessment knowledge level of EFL instructors

The findings of the present study revealed general and skill-based assessment knowledge of English language educators in schools of foreign languages in Turkish higher education institutions. When the mean of total scores received from the LAKS analysed, it is clear that EFL instructors who teach at state universities in Turkey lack necessary knowledge in language testing and assessment. The study confirmed the study conducted by Ölmezer-Öztürk and Aydın (2018b) regarding insufficient assessment skills and comprehension of language educators in Turkish higher education context. From the results, it is clear that EFL instructors are not familiar with certain assessment concepts and they have some incorrect knowledge regarding the assessment practices. Hence, it is understood that language teachers are not prepared to and effective in integrating their assessment practices into their everyday teaching practice (Pastore & Andrade, 2019). In Öz's (2014) study, which was conducted with 120 Turkish EFL teachers, it was revealed that most EFL teachers in Turkey prefer conventional methods rather than formative and alternative assessment methods. It could be concluded that EFL teachers pay inadequate attention to recent concepts and theories in language assessment field.

It is clear from the findings that EFL teachers are in need of developing their theoretical knowledge and practical skills in language assessment. Since these educators spend one fourth of their working time dealing with the work related to assessment without the necessary knowledge and training, they should learn various procedures (Stiggins, 2014), concepts and methods of language assessment to be able to conduct their responsibilities well as a teacher. The results are parallel with many studies conducted in various contexts around the world (Allal, 2013; Beziat & Coleman, 2015; Daniel & King, 2010; Lan & Fan, 2019; Tsagari & Vogt, 2017). In a study conducted in a similar context in Turkey, Semiz and Odabaş (2016) found that EFL instructors in Turkey reported that they needed a basic level of training in the areas which require further improvement. They stated that they were less familiar with concepts such as formative assessment, self-peer assessment and portfolio. One possible cause of this inadequacy in assessment knowledge levels of EFL teachers is that only around half of the EFL instructors who participated in this study possessed experience in an assessment unit. Hence, they do not construct formal assessment tools for their own learners but a group of instructors in an assessment unit prepare them. Hence, without taking an active role in test development process, the competency level of in-service EFL instructors may decrease in time. On the other hand, novice English instructors may not find an opportunity to put their

knowledge they gained in their undergraduate program into practice to reinforce their assessment skills and practices.

Another possible cause of this assessment knowledge inadequacy of EFL instructors could be related to the content of language assessment courses that pre-service teachers are provided during their undergraduate education. Hatipoğlu (2010) points out that the assessment course given to pre-service English language teacher candidates in Turkey is not sufficient to acquire enough knowledge and skills in foreign language assessment. Thus, many English teachers may graduate their teacher education programs without equipping themselves with the necessary competencies and skills needed for designing appropriate assessment tools based on the requirements of their course objectives and their learners' needs. In addition, according to Yastıbaş (2018), while English teachers who graduated ELT departments were satisfied with their level of assessment knowledge, the teachers who graduated from non-ELT departments reported that they needed training in the area. Hence, the English teachers who graduated a non-ELT department may be in need of a different training program in which they could reach content related to fundamentals of language assessment area.

The item-based analysis of the LAKS results also indicate that many language teachers have incorrect knowledge about certain assessment concepts or procedures as they selected the wrong option instead of choosing 'don't know' option. In addition to gaining competency in educational assessment, the instructors should also raise their awareness in assessment of main skills and various sub-skills of language learning. According to the results of the present study, the EFL instructors struggle most in assessment of listening (mean=7,09) and it is followed by assessment of writing skills (mean=7,45). Thus, these findings are consistent with research showing that EFL teachers lack knowledge in especially assessment of listening and writing skills in Iran (Firoozi, Razavipour and Ahmadi, 2019). In addition, Tsagari and Vogt (2017) conducted a study which revealed that most language teachers from seven different countries in Europe stated that they used generally traditional methods of assessment and they needed to receive training in language assessment. This shows us that the language educators need to raise their awareness regarding factual information about language assessment and appropriate and alternative methods of assessment.

Based on OECD report (2019), the teachers in Turkey tend to limit their classroom assessment processes with mainly multiple-choice tasks which are easy to construct and evaluate. It was also reported that these multiple-choice tests basically focus on subject areas and knowledge which are assessed in the central exams while other areas are given less importance as they are not examined in the high-stake central examinations. Even if the report

evaluated the general assessment practices including all subject areas in Turkey, one may derive similar conclusions related to language assessment tendencies of the EFL teachers in Turkey. As the learners are familiar with multiple-choice items which assess their reading, grammar and vocabulary knowledge in English in central examinations, for this reason EFL teachers in Turkish context could tend to develop tests with similar tasks and content. Thus, EFL teachers may tend to develop assessment instruments which include tasks that assess grammar and vocabulary knowledge of learners rather than developing alternative tasks assessing reading, listening, speaking and writing skills of language learners. On the other hand, Öz and Atay's study (2017) demonstrated that EFL teachers in Turkey preferred mostly pair work and speaking focused activities to assess their students' performance. Hence, the results showed that the English teachers were aware of the importance of oral productive skills but the findings also showed that the other three main skills of language use needed to be considered while designing assessment tool. Thus, further research is needed to be conducted to assess the skill-based assessment practices of EFL teachers in Turkey.

5.1.2. The impact of gender and assessment unit experience on language assessment knowledge

A two-way between-group analysis of variance test was conducted to judge the impact of gender and assessment unit experience on language assessment knowledge of EFL instructors. The findings of between-effects test indicated that there is not a statistically significant difference between male and female EFL instructors' mean scores in terms of language assessment knowledge. Although the mean scores of the male participants was slightly higher than the mean scores of the female participants, there was no statistically significant influence of gender on the LAKS scores of the EFL instructors. The results of Ölmezer-Öztürk and Aydın's study (2018b) which found that there was no impact of gender on language educators' knowledge of language assessment is in parallel with the findings of this study. However, the results contradict the claims of Asamoah, Songnalle, Sundeme and Derkye (2019) which found gender difference in formative assessment knowledge of high school teachers from various subjects in Ghana. This may be explained with the sample difference between their study and the present one so it can be concluded that, for English teachers, gender has no influence in relation to language assessment competency.

The results of the ANOVA test also revealed that having a professional background as a test developer in an assessment unit affects the language assessment competency of the EFL instructors. The mean scores of these language educators were found to be higher than the one who did not have an experience in an assessment office before. As the p value of the

effect was 0,002, it is clear that assessment unit experience creates a statistically significant effect on the level of knowledge in language assessment. At this point, the results of the present study are in line with the findings of Ölmezer-Öztürk and Aydın's study (2018b) which revealed an impact of assessment unit experience on assessment knowledge. Thus, it can be stated that the experience of constructing language assessment tools and conducting test administration procedures obviously helped those instructors to gain measurement knowledge, understanding of assessment concepts and test development skills.

5.1.3. The impact of years of teaching experience on language assessment knowledge

From the results of the Pearson correlation test, no correlation was found between the years of work experience of EFL instructors and their performance on LAKS. From this point of view, it was understood that less experienced EFL instructors can have more knowledge in language assessment. Öz and Atay (2017) also found similar results in their study with 12 EFL instructors in preparatory school in Turkey. However, in school of foreign languages in Turkey, mostly the EFL instructors who work at assessment units actually develop the assessment instruments. Thus, while teachers gain work experience in their classroom, not all of them take part in processes of formal test development. From the results of this study, only a half of the EFL instructors' assessment-related practices are limited to their own extra-curricular classroom assessment activities. Moreover, the ANOVA test results showed that the EFL instructors with assessment unit experience received higher scores in LAKS. Thus, it may be concluded that when all the EFL instructors have a chance to work as a member of an assessment unit in which they have a chance to take active roles in the processes of test administration and assessment instruments development, language assessment knowledge of English teachers may increase.

On the other hand, Alderson et al. (1995) states that it is possible that knowledgeable item developers will be the teachers who are experienced in teaching similar learners or related subject areas, who are also professionally qualified in the particular setting where the assessment instruments are being developed. However, regardless of their years of experience, the research by Zhang and Burry-Stock (2003) revealed that teachers who received training in assessment had higher levels of self-perceived assessment competencies especially in standardised assessment. Thus, we may conclude that depending on the context of the studies, years of work experience and/or the training received during in-service teaching can make a significant impact on assessment competency of the EFL instructors. Thus, more studies need to be conducted to measure this.

5.1.4. Language assessment practices, beliefs and suggestions of the EFL instructors with high scores in LAKS

The semi-structured interviews were conducted with the EFL instructors who gave correct answers to at least 70 per cent of the LAKS items. The results of the qualitative analysis showed that there were some commonalities among these EFL instructors' background practices, education and interests. First of all, it was found that these instructors generally have a background in an assessment unit. This finding is in line with the quantitative part results. Moreover, most of them had an MA degree and all of them stated that they had a personal interest in language assessment. Thus, it is obvious that being involved in construction and administration of assessment instruments equipped these teachers with the necessary knowledge and skills in language assessment area. Also, in addition to their master's degrees, they are interested in and willing to conduct practices related to assessment.

The findings of the qualitative research revealed that working on and analysing various types of assessment instruments and materials help the language instructors to develop a more professional and qualified background and competency in understanding and constructing language assessment tools. This can be explained with these language educators' awareness in the value of assessment in language learning as well as the qualification of the teacher as a well-equipped language leader in the classroom. Hence, language instructors can develop their knowledge in language measurement by analysing the assessment materials with a critical point of view, especially when there is a committee which consists of some test developers and plenty of the teachers do not design any assessment tools.

Another finding of the interviews was the EFL teachers believe the significance of receiving feedback from another assessment task developer or a more experienced teacher who had taken an active role in designing a range of tests for various academic purposes. Language educators can be provided with a chance to gain different insights and information with the help of another colleague's guidance and feedback. Hence, As McKay (2006, p. 226) pointed out, involving in an 'assessment dialogue' with each other should be a part of professional life of every language teacher since it is one of the most effective tools for improving the quality of classroom assessment procedures.

The participants of the study also pointed out that the content of the training sessions on assessment should not merely include theoretical aspects of test construction. The training courses need be organised according to the needs of the teachers and institution. On the other hand, the importance of getting necessary training on language assessment is obviously seen

fruitful by the assessment literature EFL instructors. Many studies conducted in different contexts in assessment literature are in line with this finding (Brown, 2016; Fulcher, 2012; Gareis & Grant, 2015). It is clearly seen that well-organised training sessions may fulfil the needs of the language teachers who lack required competencies in educational assessment and language assessment especially.

5.2. Conclusion

This study was conducted to investigate the language assessment knowledge level of in-service EFL teachers at state universities in Turkish context in terms of measurement of general and skill-based language performance. Another aim of the study was to investigate the background practices, beliefs and suggestions of assessment literate language instructors. In quantitative analysis part, it was found that the EFL instructors had insufficient knowledge about language assessment. When the scores of the EFL instructors for each skill was analysed with descriptive statistics, the mean scores for each skill was found low. As for the impact of gender and work experience on EFL instructors' language assessment knowledge, no statistically significant influence was found. The only statistically significant relationship was seen among the EFL instructors with respect to having assessment unit experience or not. According to the results, the EFL instructors who had experience in an assessment unit received higher scores in LAKS compared to the ones who do not have such a background.

The results of the qualitative part showed that the EFL instructors who possess a high level of competency in general and skill-based language assessment share some common characteristics. They have a personal interest in assessment, generally they have a background as an assessment unit member and they mostly received education on assessment in their undergraduate program. They also share certain practices such as preparing exam-like classroom materials, analysing different types of assessment instruments and reading recent trends and developments related to the area. The results of the qualitative research also showed that the EFL instructors who received high scores in LAKS have awareness in their own assessment practices. These English instructors reported that they believed that a higher level of assessment knowledge brings a more effective teaching and learning setting. They also regard that assessment is an area of interest so one needs to have an interest in it to be able to construct good assessment tools. Lastly, the EFL instructors who possess a good level of knowledge in language assessment suggest in-service training programs, mentoring systems and feedback cycles to help language educators improve their assessment competency.

5.3. Suggestions

From the findings of the study it is evident that Turkish EFL instructors have inadequate knowledge about language assessment. One recommendation for this issue is to organise in-service training sessions which include necessary conceptions, theoretical principles and practical aspects of language assessment. Although training on assessment for pre-service teachers is necessary, continuous training on a range of topics related to language assessment should be a significant component of any related institution's professional development program for its educators (Mertler, 2004; Herrera & Macias, 2015). For in-service EFL teachers, professional development training sessions including theoretical aspects and practical considerations of language assessment could be organised. All the EFL instructors whether they work in an assessment office or not could be supported to participate in these trainings. Given that assessment is not limited to construction, administration and marking of tests, classroom educators also need to gain competency in evaluation of learner performance. Teachers should be aware of their learners' needs, able to measure their classroom performance and give effective feedback to them. Regarding the broad scope of language assessment, it is seen that assessment include plenty of teaching and learning related activities, and it completes the classroom instruction. For this reason, classroom instructors could be provided in-service training which conforms the needs of the teachers, learner profile and curriculum objectives. The assessment training could involve fundamental domains of assessment as well as sessions on marking scales, alternative assessment techniques, self/peer assessment methods.

Knowing about development of test items will help teachers in being more and better informed and, thus, they will tend to make more reasonable decisions, either as test constructors, as examinees, or as the users of test results (Osterlind, 2002). Thus, while designing the content of training sessions on language assessment, the related educators could consider integrating the framework of Xu and Brown (2016) in their course plan. Xu and Brown' (2016) process-oriented framework describes a hierarchical model for literacy in assessment including six components of it:

- a. Knowledge base
- b. Assessment conceptions
- c. Institutional and socio-cultural awareness
- d. Assessment literacy in practice
- e. Teacher learning
- f. Development of teacher identity as assessor

From the findings, another issue to consider is that having experience in an assessment unit, which means taking active roles in test preparation processes increases the language assessment knowledge level of English teachers. The EFL instructors who do not work in an assessment unit may lead them stay unaware of the test preparation processes conducted in a particular institution in some cases. According to the institutional needs and curriculum objectives, these language instructors also need to develop and conduct certain classroom assessment processes. Thus, in fact, whether they prepare test materials or not, they should have competency for develop necessary assessment tools. In order to develop effectiveness of their classroom assessment practices and skills in interpretation of test results, assessment unit members could organise workshops or seminars in which they collaborate with the classroom instructors. Moreover, via receiving and providing feedback assessment unit instructors could work with classroom teachers so that the relationship between classroom practices and assessment tools is maintained.

Given that conducting assessment and evaluation practices is not only the responsibility of assessment unit members in preparatory schools but a part of every language teacher's classroom instruction, the teachers could lack awareness in their competency in LAK. Assessment knowledge scales similar to the ones used in this study could be administered and data on teachers' needs and expectations could be collected. The findings of this could be shared with every language teacher individually and make them aware of their own strengths and areas requiring further improvement. In this way, the teachers can do some individual work to close the gap between their knowledge and the target domain. In addition, the mean scores of every group of students and the mean score of the whole institution could be shared with the classroom educators. Also, the item analysis of every question individually and section-by-section analysis results could also be conducted and shared with the teachers. As a result of these, language teachers can raise their awareness related to their learners' performance and pay necessary attention to formative assessment tools, as well.

The findings of the qualitative study also revealed that teacher education programs in Turkish higher education institutions may have some limitations with respect to content of courses on language assessment, and pre-service EFL teachers' needs can be fulfilled with some modifications in the course content and materials. The instructional assessment training course could be provided when teacher candidates are taking instructional practicum during their university education so that they can find a context where they can get feedback regarding their future assessment practices in an instructional context (Alkharusi, Kazem & Al-Musawai, 2011). In addition, it is understood that there could be more than a single course on language

assessment. The objectives and content of the course may include both theoretical conceptions and procedures in assessment and more practical, hands-on practices that require teacher candidates to analyse, comment on and construct a range of assessment instruments for different contexts and purposes. The assessment of listening and writing skills, in which EFL instructors seem to struggle more according to the findings of the study, could be given more importance within the scope of assessment courses. The assessment and evaluation of these courses in undergraduate programs should also include various types of assessment tools in order to make pre-service teachers experience washback effect of the assessment tools as learners as well as practicing the design of these sort of assessment materials.

Ashraf and Zolfaghari's (2018) study revealed the existence of a positive link between English teachers' language assessment knowledge and their practices related to reflective teaching. Thus, another issue to consider is that even though professional development offices and school administrators in schools of foreign languages in Turkey encourage and coordinate classroom observations which involve addressing instructional practices of the teachers, these observations lack giving emphasis on assessment practices used in the classroom. The reason for this could be the supervisor's lack of assessment knowledge and the absence of sufficient teacher trainers-assessment researchers' collaboration (Babaii & Asadnia, 2019). Regarding this, the programs designed for the education of language teacher trainers could include content related to recent studies, new trends, various procedures and practices in language assessment fields. In this way, teacher educators and supervisors could give feedback on assessment practices of the classroom teachers as well as assessment unit members who construct language exams.

A conclusion made from the study is that construction of assessment materials is considered as a different area of profession which require an advanced knowledge in language performance assessment. Assessment task developers must be experts in the subject matter, preferably in a specific area for which these task designers will be developing tests (Haladyna, 2004). The EFL instructors who take active roles in the design and administration phases of language tests could be selected based on their background related to assessment practices and their personal interest in language assessment. These assessment tasks developers could be provided a range of training sessions based on their needs. They can be supported and encouraged to analyse different types of questions, work on their marking rubrics and evaluate their present assessment practices. In addition, assessment unit members could be encouraged to interact with other test developers at other schools of foreign languages and attend related professional development activities and conferences. These teachers can also arrange and lead

in-service training sessions to familiarise other teachers with their practices and expectations, and organise calibration sessions to increase the reliability of the language tests.

Besides, collaboration among language teachers should be encouraged to ensure that language learners go through a range of classroom assessment procedures based on their needs and the curriculum objectives. Engaging in a kind of ‘assessment dialogue’ with their colleagues serves as one of the most influential methods of improving classroom assessment competency (McKay, 2006, p. 144). Language teachers could cooperate and collaborate by observing each other’s assessment practices in their classroom, sharing knowledge and experience, and requesting guidance from their colleagues. The teachers could examine and give feedback on each other’s assessment materials such as guidelines, portfolio tasks, feedback sheets and so forth. Hence, they could improve their own assessment knowledge and practices by the help of ideas and feedback given by a colleague who would look at their work from a different angle.

Another approach to be followed to improve language teachers’ LAK is to reconsider and modify the institutional aims and objectives regarding the instruction and profile of the learners who could complete the course successfully. When an institution gives more importance on the summative and traditional methods of assessment only, the teachers working at this school may have a tendency to do the same only. However, if the management and executive agencies of the language education institutions could develop institutional culture which contain beliefs, values and behaviours related to quality and contemporary assessment measurements, the attitudes and practices of the institution’s educators can adapt to the same way.

5.4. Limitations

This study was conducted with in-service EFL instructors working at schools of foreign languages in four state universities in Turkey. The results may not be generalised to all English teachers in various contexts. EFL instructors who work in different types of educational courses may have different levels of language assessment knowledge. Thus, more studies should be conducted with EFL teachers at different settings.

The study was conducted with 114 EFL instructors. The number of the participants who attended the qualitative part of the was limited to 8 since not an expected number of instructors could answer minimum 70 % of the questions correctly. Hence, the qualitative data were collected from only 8 participants. In future studies, considering this, the number of quantitative part participants could be higher.

REFERENCES

- Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press
- Alderson, J. C. (2000). *Assessing Reading*. Cambridge University Press
- Alkharusi, H.& Kazem, A. & Al-Musawi, A. (2011). Knowledge, skills, and attitudes of preservice and in-service teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*. 39, 113-123. 10.1080/1359866X.2011.560649
- Allal, L. (2013). Teachers' professional judgement in assessment: a cognitive act and a socially situated practice, *Assessment in Education: Principles, Policy & Practice*, 20:1, 20-34, DOI: 10.1080/0969594X.2012.736364
- Asamoah, D., Songnalle, S., Sundeme, B. & Derkye, B. (2019). Gender Difference in Formative Assessment Knowledge of Senior High School Teachers in the Upper West Region of Ghana. *Journal of Education and Practice*, 10(6)., 54-58, DOI: 10.7176/JEP
- Ashraf, H., & Zolfaghari, S. (2018). EFL Teachers' Assessment Literacy and Their Reflective Teaching. *International Journal of Instruction*, 11(1), 425-436. <https://doi.org/10.12973/iji.2018.11129a>.
- Babaii, E. & Asadnia, F. (2019) A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice, *Reflective Practice*, 20:6, 745-760, DOI: 10.1080/14623943.2019.1688779
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press
- Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press
- Baker, B. (2016). Language Assessment Literacy as Professional Competence: The Case of Canadian Admissions Decision Makers. *The Canadian Journal of Applied Linguistics* 19(1), 63-83
- Barber, P., & Hill, M. (2014). *Preparing for a renaissance in assessment*. London: Pearson. Retrieved from: <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/PreparingforaRenaissanceinAssessment.pdf>

- Bayat, K. & Rezaei, A. (2015) Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1) pp. 139-146. DOI: <https://doi.org/10.5296/ijele.v3i1.6887>
- Beziat, T. L. R., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25-30
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National assessment summit papers* (pp. 18–23). Ames, IA: Iowa State University
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30, 3–12. <http://dx.doi.org/10.1111/emip.2011.30.issue-1>
- Brown, H. Douglas. (2003). *Language Assessment: Principles and Classroom Practices*. California: San Francisco State University
- Brown, J. D. (1996). *Testing in Language Programs*. Prentice Hall Regents, New Jersey
- Buck, G. (2001). *Assessing Listening* (Cambridge Language Assessment). Cambridge: Cambridge University Press
- Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction* 7(1), pp. 107-120. Retrieved from <https://eric.ed.gov/?id=EJ1085246>
- Chapelle, A. C. & Douglas, D. (2006). *Assessing Language through Computer Technology*. Eds. J. Charles Anderson & Lyle F. Bachman. Cambridge University Press
- Cheng, L. & Fox, J. (2017). *Assessment in the Language Classroom*. Red Globe Press
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches (2nd ed.)*. Thousand Oaks, CA: Sage
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.)*. Sage Publications, Inc
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52, 281–302

- Cronbach, L. J. (1971). Test Validation. In R. L. Thorndike (Ed.), *Educational Measurement*, pp. 443-507
- Crusana, D., Plakans, L. & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing* 28 (43–56)
- Daniel, L. G. & King, D. A. (2010) Knowledge and use of testing and measurement literacy of elementary and secondary teachers, *The Journal of Educational Research*, 91:6, 331-344, DOI: 10.1080/00220679809597563
- Davies, M. & Taras, M. (2018) Coherence and disparity in assessment literacies among HE staff. *London Review of Education*, 16 (3), 474-490
- 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
- Deneen, C., & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education*, 39, 577–591. <http://dx.doi.org/10.1080/02602938.2013.859654>
- Deneen, C. C. & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, 3(1) 1-14. <http://dx.doi.org/10.1080/2331186X.2016.1225380>
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press
- Douglas, D. (2010). *Understanding language testing*. London, UK: Hodder Education
- Edwards, F. (2017). A rubric to track the development of secondary preservice and novice teachers' summative assessment literacy, *Assessment in Education: Principles, Policy & Practice*, 24:2, 205-227, DOI: 10.1080/0969594X.2016.1245651.
- Farhady, H. & Tavassoli, K. (2018). Developing a language assessment knowledge test for EFL teachers: A data-driven approach. *Iranian Journal of Language Teaching Research* 6(3) (79-94). Retrieved from: http://ijltr.urmia.ac.ir/article_120602_f86e80ea4e4c1b739ed17a6e0481bff7.pdf
- Firoozi, T., Razavipour, K. & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing*

- in Asia 9(2). Retrieved from: <https://language-testing-asia.springeropen.com/articles/10.1186/s40468-019-0078-7>
- 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., & Davidson, F. (2006). *Language testing and assessment: An advanced resource book*. New York: Routledge. <http://dx.doi.org/10.4324/9780203449066>
- Fulcher, G. (2012) Assessment literacy for the language classroom, *Language Assessment Quarterly*, 9:2, 113-132, DOI: 10.1080/15434303.2011.642041
- Gareis, C. R. & Grant, L. W. (2015). Assessment literacy for teacher candidates. *Teacher Educators' Journal*, 4–21. Retrieved from: <https://eric.ed.gov/?id=EJ1085699>
- Giraldo, F. (2017). 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>
- Grellet, F. (1981). *Developing reading skills*. Cambridge: Cambridge University Press
- Hakuta, K. & Beatty, A. (eds.) (2000) *Testing English-language learners in U.S. schools*. Washington, DC: National Academy Press. DC: National Academy Press
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Lawrence Erlbaum Associates Publishers. Mahwah, New Jersey, London
- Han, T. & Kaya, H. İ. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education* 4(1), 77-93
- Hatipoğlu, Ç. (2010). Summative evaluation of an English language testing and evaluation course for future English language teachers in Turkey. *ELTED*, 13, 40-51
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: expectations and needs of pre-service English language teachers. *International Association of Research in Foreign Language Education and Applied Linguistics* 4(2), 111-128. Retrieved from: <http://dergipark.ulakbim.gov.tr/eltrj/>
- Haznedar, B. (2012). Perspectives on pre-service education of English language teachers in Turkish primary schools. In Y. Bayyurt & Y. Bektaş-Çetinkaya (Eds.), *Perspectives on Teaching and Learning English in Turkey* (pp. 39-57). Berlin & New York: Peter Lang

- Heaton, J. B. (1990). *Writing English language tests: Longman handbook for language teachers (New Edition)*. London: Longman Group UK Ltd
- Herrera, L. & Macías, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colomb. Appl. Linguist. Journal*, 17(2), pp. 302-312. DOI: <http://dx.doi.org/10.14483/udistrital.jour.calj.2015.2.a09>
- 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
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25, 328–402
- Jabbarifar, T. (2009). The importance of classroom assessment and evaluation in educational system. *Proceedings of the 2nd International Conference of Teaching and Learning (ICTL 2009)*. INTI University College, Malaysia
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing* 30(3) 345–362. <https://doi.org/10.1177/0265532213480334>
- Johnson, R. B., Onwuegbuzie, A. J. & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, 112-133
- Kahl, S. R., Hofman, P. & Bryant, S. (2013). *Assessment literacy standards and performance measures for teacher candidates and practicing teachers*. Prepared for the Council for the Accreditation of Educator Preparation (CAEP)
- Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of behavioral research*, 4th ed. Orlando, FL: Harcourt Brace
- 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

- Kitchen, H., et al. (2019), *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Turkey*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://doi.org/10.1787/5edc0abe-en>
- Kopriva, R. (2008). *Improving testing for English language learners*. Routledge, New York.
- Köksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education-TOJDE* 5(1). Retrieved from: https://www.researchgate.net/publication/26339362_Assesing_Teachers'_Testing_Skills_in_Elt_and_Enhancing_Their_Professional_Development_Through_Distance_Learning_on_The_Net
- Kubiszyn, T. & Borich, G. (1996). *Educational testing and measurement: Classroom application and practice* (5th ed.). New York: Harper Collins
- Kunnan, A. J. (Ed.). (2000). Fairness and validation in language assessment: *Selected papers from the 19th Language testing research colloquium*, Orlando, Florida. Cambridge, UK: Cambridge University Press
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System* 81 (78-89)
- Lan, C. & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation* 61 (112–122)
- Leedy, P. D. (1993). *Practical research: Planning and design*. New Jersey: Prentice-Hall
- Lees, R. & Anderson, D. (2015). Reflections on academics' assessment literacy. *London Review of Education* 13(3) 42-48. DOI: 10.18546/LRE.13.3.06
- Levy-Vered, A. & Alhija, F. N. (2015). Modelling beginning teachers' assessment literacy: the contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation* 21(5–6), 378–406, <http://dx.doi.org/10.1080/13803611.2015.1117980>
- Looney, A., Cumming, J., Kleij, F. & Harris, K. (2017). Reconceptualising the role of teachers as assessors: teacher assessment identity, *Assessment in Education: Principles, Policy & Practice*, <http://dx.doi.org/10.1080/0969594X.2016.1268090>

- Luoma, S. (2004). *Assessing speaking* (Cambridge Language Assessment). Cambridge: Cambridge University Press. DOI:10.1017/CBO9780511733017
- Madsen, H. S. (1983). *Techniques in testing*. Eds: Russell N. Campbell & William E. Rutherford. Oxford University Press
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*. 30(3) 329-344. <https://doi.org/10.1177/0265532213480129>
- Marin, J. H. M. (2009). Exploring teachers' practices for assessing reading comprehension abilities in English as a foreign language. *Profile* 11(2), 71-84. Retrieved from: <https://www.semanticscholar.org/paper/Exploring-Teachers'-Practices-for-Assessing-Reading-Mar%C3%ADn/536bc0cebc60d36df9b7419318cc5a50edd7939b>
- Marshall, B. (2011). *Testing English: Formative and summative approaches to English assessment*. Continuum International Publishing Group
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press
- McNamara, T. (1996). *Measuring second language performance*. London & New York: Longman
- McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing
- Mede, E. & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Ankara Tömer Dil Dergisi* 168(1). DOI: 10.1501/Dilder_00000000237
- Medland, E. (2019). 'I'm an assessment illiterate': towards a shared discourse of assessment literacy for external examiners. *Assessment & Evaluation in Higher Education*, 44(4), 565-580. <https://doi.org/10.1080/02602938.2018.1523363>
- Mellati, M. & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education* 43(6). <http://dx.doi.org/10.14221/ajte.2018v43n6.1>
- Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?*. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus

- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education* 33(1) 49-64. Retrieved from: <https://www.jstor.org/stable/41064623>
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools* 12(2) 101-113. DOI: 10.1177/1365480209105575
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp.13-103). Newyork: Macmillan
- Navarrete, C. et al. (1990). Informal assessment in educational evaluation: Implications for bilingual education programs. *National Clearinghouse for Bilingual Education*, Washington, DC. Retrieved from: <https://files.eric.ed.gov/fulltext/ED337041.pdf>
- Newfields, T. (2006). Teacher development and assessment literacy. *Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference*. (pp. 48-73)
- Ogan-Bekiroglu, F. & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice, *Curriculum Journal*, 25(3), 344-371, DOI: 10.1080/09585176.2014.899916
- Osterlind, S. J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance and other formats*. New York: Kluwer Academic Publishers
- Ölmezer-Öztürk, E. & Aydın, B. (2018a). Investigating language assessment knowledge of EFL teachers. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education)*. DOI: 10.16986/HUJE.2018043465
- Ölmezer-Öztürk, E. & Aydın, B. (2018b). 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>
- 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), pp. 775-785. Doi: 10.4304/jltr.5.4.775-785
- Öz, S. & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: perceptions and practices. *International Association of Research in Foreign Language*

- Education and Applied Linguistics* 6(1), 25-44. Retrieved from: <http://dergipark.ulakbim.gov.tr/eltrj/>
- Pastore, S. & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education* 84 (128-138)
- 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
- Popham, W. J. (1995). *Classroom assessment: What teachers need to know*. Needham Heights, MA: Allyn and Bacon
- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62, 82
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48, 4–11. DOI:10.1080/00405840802577536
- Popham, W. J. (2010). *Everything school leaders need to know about assessment*. Thousand Oaks: Corwin Press
- Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research*, 37(2), 179-192. Retrieved from: <https://psycnet.apa.org/record/1992-00151-001>
- Rogier, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Teaching Forum* 3 (3-13). Retrieved from: <https://eric.ed.gov/?id=EJ1045594>
- Rossmann, G., & Rallis, S. F. (1998). *Learning in the field: An introduction to qualitative research*. Thousand Oaks, CA: Sage
- Sarıçoban, A. (2011). A study on the English language teachers' preparation of tests. *Eğitim Fakültesi Dergisi*. 398-410. Retrieved from: https://www.researchgate.net/publication/298312811_A_STUDY_ON_THE_ENGLISH_LANGUAGE_TEACHERS'_PREPARATION_OF_TESTS
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory Into Practice* 32(2) Assessing Tomorrow's Teachers (Spring, 1993), pp. 118-126. Retrieved: <https://www.jstor.org/stable/1476329>

- Semiz, Ö. & Odabaş, K. (2016). Turkish EFL teachers' familiarity of and perceived needs for language testing and assessment literacy. Proceeding of the 3rd international linguistics and language studies conference, İstanbul, Turkey. Retrieved from https://docs.wixstatic.com/ugd/bac820_2c683311d65a478f9b10ee6dfbd1020d.pdf
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 534-539
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 238-245
- Stiggins, R. (2008). Assessment manifesto: A call for the development of balanced assessment systems. Portland, OR: ETS Assessment Training Institute
- Stiggins, R. J. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan*, 96-67. DOI: 10.1177/0031721714553413
- Stiggins, R. J. (2017). *The perfect assessment system*. ASCD
- Şahin, S. (2015). *Language testing and assessment (LTA) literacy of high school English language teachers in Turkey*. Paper presented at the 3rd ULEAD Congress, International Congress on Applied Linguistics: Current Issues in Applied Linguistics, Çanakkale, Turkey
- Tajeddin, Z., Alemi, M. & Yasaei, H. (2018). Classroom assessment literacy for speaking: exploring novice and experienced English language teachers' knowledge and practice. *Iranian Journal of Language Teaching Research* 6(3), 57-77
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics* 29 21-36
- Thorndike, R. L. & Hagen, E. P. (1977) *Measurement and evaluation in psychology and education*. 4th ed. New York: Wiley
- Tsagari, D. & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment* 6(1) 41-63.
- Urquhart, A. H. (1992). *Draft band descriptors for reading*. Report to the IELTS Research Committee. Plymouth: College of St Mark and St John.
- Weigle, S. (2002). *Assessing writing* (Cambridge Language Assessment). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511732997

- Weir, C. J. (1994). *Reading as multi-divisible or unitary: between Scylla and Charybdis*. Paper presented at the RELC, SEAMEO Regional Language Centre, Singapore. Retrieved from: <http://nflrc.hawaii.edu/rfl/PastIssues/rfl102weir.pdf>
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave, Macmillan
- William, D. (2015). Foreword: Assessment literacy. *London Review of Education* (13)3. Available at: <https://www.researchgate.net/publication/308578545>
- Wright, S.P., Horn, P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67
- Xu, Y. (2015). *Language assessment literacy in practice: A case study of a Chinese university English teacher*. Paper presented at the 37th international LTRC Conference on From Language Testing to Language Assessment: Connecting Teaching, Learning, and Assessment (March 18-20). Toronto, Canada
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58(1), 149-162
- Yastıbaş, A. E. & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessments. *Journal of Language and Linguistic Studies*, 14(1), 178-193. Retrieved from: <https://www.jlls.org/index.php/jlls/article/view/861>
- Zhang, Z. & Burry-Stock, J. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342. DOI: 10.1207/S15324818AME1604_4
- Zolfaghari, F. & Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education* (2016), 3: 1252561. DOI: 10.1080/2331186X.2016.1252561.

APPENDICES

APPENDIX 1. STUDENT'S CURRICULUM VITAE

Kişisel Bilgiler			
Adı ve Soyadı	Özge DERETARLASI		
E-postası/Web Sayfası	ozgederetarlasi@gmail.com		
Bildiği Yabancı Diller	İngilizce		
Uzmanlık Alanı	İngilizce Öğretmenliği		
Öğrenim Bilgileri			
	Üniversite	Bölüm	Yıl
Lisans	Hacettepe Üniversitesi	İngiliz Dili Öğretmenliği	2008-2012
Tez Başlığı	-		
Tez Danışmanı	-		
Akademik Eserler			
(Makale, kitap, kitap bölümü, bildiri, poster, sergi, konser vb. eserlerden kaynakça yazım kurallarına göre en fazla 10 eser yazılmalıdır.) Tezden üretilen yayınlar * ile işaretlenmelidir.			
Alanıyla İlgili Bilimsel Kuruluşlara Üyelikler			
Alanıyla İlgili Aldığı Ödüller			

APPENDIX 2. OFFICIAL NOTICE(S) OF ETHICAL PERMISSON

SDP : 199.99 - Eğitim-Öğretimle İlgili Diğer İşler
Evrak No : 23708841-199.99-8928
Evrak Tarihi : 20-01-2020

KEP 20.01.2020 E.2000004368



T.C.
İZMİR KATİP ÇELEBİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Öğrenci İşleri Daire Başkanlığı



Sayı : 24346654-199
Konu : Tez Uygulaması (Özge DERETARLASI)

DOKUZ EYLÜL ÜNİVERSİTESİ REKTÖRLÜĞÜNE
(Öğrenci İşleri Daire Başkanlığı)

İlgi : 09.01.2020 tarihli ve 3434 sayılı yazınız.

İlgi yazınıza istinaden, Üniversiteniz Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Yüksek Lisans Programı öğrencisi Özge DERETARLASI'nın "Investigating EFL Instructors' Language Assessment Knowledge" konulu tez çalışmasını Üniversitemiz Yabancı Diller Yüksekokulunda yapması Rektörlüğümüzce uygun görülmüştür.

Bilgilerinizi ve gereğini arz ederim.

e-imza/ihdr

Prof. Dr. Turan GÖKÇE
Rektör a.
Rektör Yardımcısı

Adres : İzmir Katip Çelebi Üniversitesi Çifti Ana Yolu Kavşağı 25520 Çiğli İlçesi - TÜRKİYE	Bilgi İçerir : Öğzetim Akademi - Menteş
e-posta : rektar.akademi@duku.edu.tr	Telefon : (0 232) 320 5335 Belgeleme No : (0 232) 346 0808 İnternet Adresi : www.doku.edu.tr



Bu belge, 2020 yılında İzmir Katip Çelebi Üniversitesi'nde Prof. Dr. Meriç CAĞLAR tarafından 05.02.2020 tarihinde e-imzalanmıştır.
Evrak numarası: 24346654-199.99-8928. Bu belge, 938BF109XA koda ile doğrulanabilir.

SDP : 302.08.01 - 01 - Bilimsel ve Eğitim Amaçlı
Evrak Tarih : 24-01-2020



MANİSA
CELAL BAYAR
ÜNİVERSİTESİ

T.C.
MANİSA CELAL BAYAR ÜNİVERSİTESİ
Yabancı Diller Yüksekokulu Müdürlüğü



Sayı : 86044887-302.08.01-
Konu : Anket çalışması (Özge Deretarlasi)

ÖĞRENCİ İŞLERİ DAİRE BAŞKANLIĞINA

İlgi : 14/01/2020 tarihli ve 32585336-302.08.01-E.3983 sayılı yazı.

Dokuz Eylül Üniversitesi Rektörlüğü Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Yüksek Lisans Programı öğrencisi Özge DERETARLASI'nın "Investigating EFL Instructors' Language Assessment Knowledge" konulu tez çalışması kapsamında uygulama yapma isteği uygun bulunmuştur.
Bilgilerinizi ve gereğini arz ederim

e-imzalıdır
Prof. Dr. Mahmut KARĞIN
Yüksekokul Müdürü



SDP : 300.99 - Öğrenci İşleri (Genel)
 Kıtık No : 23708841-300.99-16701
 Binyılın Evrak No ve Sayısı : 03/02/2020-E.8636

KEP 04.02.2020



T.C.
 EGE ÜNİVERSİTESİ REKTÖRLÜĞÜ
 Sosyal ve Beşeri Bilimler Bilimsel Araştırma ve Yayın Etiği
 Kurulu



Sayı : 85553214-600
 Konu : Araştırma İzni (Özge
 DERETARLASI)

DOKUZ EYLÜL ÜNİVERSİTESİ REKTÖRLÜĞÜNE
 (Öğrenci İşleri Daire Başkanlığı)

İlgi : 09/01/2020 tarihli ve 15563195-302.08.01-E.3434 sayılı yazınız.

İlgi yazınız ile "Üniversitemiz Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Yüksek Lisans Programı öğrencisi Özge DERETARLASI'nın "Investigating EFL Instructors Language Assessment Knowledge" konulu tez çalışması kapsamında uygulama yapabilmesi için gerekli iznin verilmesi hususu" belirtilmiş idi.

Anılan yazı ve ekleri Üniversitemiz Yabancı Diller Yüksekokulu Müdürlüğü'nce incelenmiş olup konu ile ilgili olarak anılan Müdürlüğümüzden alınan cevabi yazıda; "... Özge DERETARLASI'nın "Investigating EFL Instructors Language Assessment Knowledge" konulu tez çalışması kapsamında Yüksekokulumuzda uygulama yapması uygun görülmüş olup çalışmanın hangi gün ve saatlerde yapılacağı hususunda bilgi verilmesi" hususu bildirilmiştir.

Bilgilerinize arz ederim.

e-İmzalıdır
 Prof. Dr. Mehmet ERSAN
 Kurul Başkanı

Evrakın Doğrulması İçin: https://edyz.ege.edu.tr/en/Verify/Validate_Doc.aspx?V=BENNZVHLP
 Ege Üniversitesi Rektörlüğü Gençlik Cad. No:12 35040 Bornova-İzmir
 Telefon No: +90 (232) 311 21 10 Faks No: +90 (232) 339 90 90
 E-Posta: İnternet Adresi: www.ege.edu.tr

Bilgi İçin: Reyhan Ayna
 Unvan: Birim Evrak Sorumlusu
 Telefon No: 4313



Bu belge 5070 sayılı e-İmza Kanununa göre Metin CAĞLAR tarafından 05.02.2020 tarihinde e-imzalanmıştır.



T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
Yabancı Diller Yüksekokulu Müdürlüğü



Sayı : 70348285-900.99-E.3482
Konu : Tez Uygulaması

10/01/2020

EĞİTİM BİLİMLERİ ENSTİTÖSÜ MÜDÜRLÜĞÜNE

İlgi : 06.01.2020 tarih ve 26 sayılı yazınız.

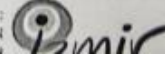
Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Yüksek Lisans Programı 2018950185 numaralı öğrencisi Özge DERETARLASI' nın Investigating EFL Instructors' Language Assessment Knowledge konulu tez çalışması kapsamında Müdürlüğümüzde tez uygulamasını yapması Yüksekokulumuzca uygun görülmüştür. Bilgilerini ve gereğini arz/rica ederim.

Prof.Dr. Kamil İŞERİ
Yüksekokul Müdür V.



Dokuz Eylül Üniversitesi – Yabancı Diller Yüksekokulu
Adres: Dokuzçesmeler Kampüsü 35160, Buca – İZMİR
Tel: 0 312 301 0 010 Elektronik Adres: iletisim@deu.edu.tr

Bilgi için İtibar:
Faruk ZEYTUN
Fakülte: 30009



APPENDIX 3. PERMISSON E-MAIL(S) FROM THE AUTHORS OF LAKS

 **Özge DERETARLASI** <ozgederetarlas@gmail.com> 2 Şub 2019 Cmt 09:33 ☆ ↩

Alıcı: elcinolmezeroturk, baydin ▼

Merhaba Belgin Hocam ve Elçin Hocam,

Dokuz Eylül Üniversitesi İngilizce Öğretmenliği bölümü yüksek lisans öğrencisiyim. Tez danışmanım Yeşim Çetinkaya'dır. Tezimde MEB'de çalışan öğretmenlerin ölçme-değerlendirme bilgi düzeylerini ölçmeyi hedefliyorum. Bu amaçla sizin geliştirmiş olduğunuz Language Assessment Knowledge Scale i tez araştırmamda kullanabilmem konusunda izninizi rica ediyorum.

Şimdiden çok teşekkür ederim. İyi çalışmalar diliyorum.

Özge Deretarlası

 **Elçin ÖLMEZER ÖZTÜRK** <elcinolmezeroturk@anadolu.edu.tr> 4 Şub 2019 Pzt 09:38 ☆ ↩

Alıcı: ben, Belgin ▼

Merhabalar,

Tezinizde geliştirdiğimiz ölçeği kullanabilirsiniz. Çalışmanızda kolaylıklar diliyorum, çıkan sonuçlardan beni de haberdar ederseniz çok mutlu olurum...

Bu arada bu ölçeği kullanarak MEB'de çalışan İngilizce öğretmenlerinden veri topluyorum ben de...

kolaylıklar diliyorum

APPENDIX 4. LANGUAGE ASSESSMENT KNOWLEDGE SCALE**LANGUAGE ASSESSMENT KNOWLEDGE SCALE****1. DEMOGRAPHIC INFORMATION**

- a. Name/Surname: _____
- b. Age: _____
- c. Sex: Female ☐ / Male ☐
- d. Nationality: TR ☐ / Other ☐
- e. Years of work experience: _____
- f. Educational background: BA ☐ / MA ☐ / PhD ☐
- g. The BA programme you graduated: ELT ☐ / Non-ELT ☐
- h. Have you ever worked for another institution/school before? YES ☐ / NO ☐
- i. Have you ever worked at the assessment unit in your institution? YES ☐ / NO ☐
- j. Have you received any training on language assessment? YES ☐ / NO ☐

2. ASSESSING READING	True	False	Don't know
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.	☐	☐	☐

3. ASSESSING LISTENING	True	False	Don't know
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/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.	☐	☐	☐
4. ASSESSING WRITING	True	False	Don't know
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.	☐	☐	☐

5. ASSESSING SPEAKING	True	False	Don't know
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.	☐	☐	☐