



**REPUBLIC OF TURKEY
CANAKKALE ONSEKİZ MART UNIVERSITY
SCHOOL OF GRADUATE STUDIES**

**DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM**

**INVESTIGATING PERCEIVED LANGUAGE TESTING AND
ASSESSMENT LITERACY LEVELS OF ENGLISH LANGUAGE
INSTRUCTORS: A SUGGESTED TRAINING SYLLABUS**

MASTER THESIS

SEDA AÇIKPORTALI

Supervisor

ASSIST. PROF. DR. ZEYNEP GÜLŞAH KANI

ÇANAKKALE – 2022



REPUBLIC OF TURKEY
ÇANAKKALE ONSEKİZ MART UNIVERSITY
SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM

**INVESTIGATING PERCEIVED LANGUAGE TESTING AND ASSESSMENT
LITERACY LEVELS OF ENGLISH LANGUAGE INSTRUCTORS: A
SUGGESTED TRAINING SYLLABUS**

MASTER THESIS

SEDA AÇIKPORTALI

Supervisor
ASSIST. PROF. DR. ZEYNEP GÜLŞAH KANI

ÇANAKKALE – 2022



T.C.
ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



Seda AÇIKPORTALI tarafından Dr. Öğr. Üyesi Zeynep Gülşah KANİ yönetiminde hazırlanan ve **21/02/2022** tarihinde aşağıdaki jüri karşısında sunulan “**Investigating Perceived Language Testing and Assessment Literacy Levels of English Language Instructors: A Suggested Training Syllabus**” başlıklı çalışma, Çanakkale Onsekiz Mart Üniversitesi Lisansüstü Eğitim Enstitüsü **Yabancı Diller Eğitimi Anabilim Dalı**’nda **YÜKSEK LİSANS TEZİ** olarak oy birliği ile kabul edilmiştir.

Jüri Üyeleri

İmza

Dr. Öğretim Üyesi Zeynep Gülşah KANİ
(Danışman)

.....

Prof. Dr. Dinçay KÖKSAL

.....

Dr. Öğr. Üyesi Burcu ÜNAL

.....

Tez No : 10449664

Tez Savunma Tarihi : 21/02/2022

.....
DOÇ.DR. YENER PAZARCIK
Enstitü Müdürü

21/02/2022

Etik Beyan

Çanakkale Onsekiz Mart Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Yazım Kuralları'na uygun olarak hazırladığım bu tez çalışmada; tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi, tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi, kullanılan verilerde herhangi bir değişiklik yapmadığımı, bu tezde sunduğum çalışmanın özgün olduğunu, bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi taahhüt ve beyan ederim.

Declaration

In this thesis, which I prepared in accordance with the Thesis Writing Rules of School of Graduate Studies of Çanakkale Onsekiz Mart University; I declared that I have obtained the data, information and documents I have presented in the thesis within the framework of academic and ethical rules; I present all information, documents, evaluations and results in accordance with scientific ethics and morals; I have cited all the works I have benefited from in the thesis by making appropriate references; I have not made any changes to the data used in the research and the data presented in this thesis is original. I have accepted all the losses of rights that may arise against me otherwise.

Seda AÇIKPORTALI

28/01/2022

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my thesis advisor Assist. Prof. Dr. Zeynep Gülşah KANI for her contributions and great support. Whenever I got lost during the writing process, she was there with her invaluable support and her constructive feedback. She helped me feel like I was accomplishing things in this rocky road. She made me believe wholeheartedly that I reached a milestone whenever I finished a chapter. All firsts are special, and I am very glad I would be the first graduated thesis student of Assist. Prof. Dr. Zeynep Gülşah KANI.

I would also express my gratitude to all the teachers at Çanakkale Onsekiz Mart University English Language Teaching Department for all the invaluable information they provided during my master's education. Historian and journalist Henry B. Adams said "A teacher affects eternity; he can never tell where his influence stops". Therefore, I would like to thank to Prof. Dr. Dinçay Köksal, Prof. Dr. Ece Zehir Topkaya, Assist. Prof. Dr. Salim Razi, Assist. Prof. Dr. Mustafa Tekin, Assist. Prof. Dr. Kürşat Cesur and Assist. Prof. Dr. M. Sercan Uztosun from whom I learned so much by attending their lessons.

I am more than grateful to my friend Utku Tönel. He always stayed after me to complete the thesis. To be honest, he was one of the prime movers in my academic career and encouraged me to improve myself academically. Ayşe, my blood sister! I couldn't thank you more for being there with your continuous support whenever I was down. I would also thank to my best friends Diyar, Buse and Büşra, who sympathised with me in this hard sledding and listened to my complaints all the time. Without you, it would have been much more difficult.

My life partner and husband, Mert... You have always been a source of motivation in my life. With your sense of humour and positive energy, I have always found enough energy to write and go on my studies. And finally, many thanks for the funny photos you took of me while I was struggling with writing the thesis. Now, we have a lot of memories reminding us of these valuable experiences.

And finally, I would like to express my sincere thanks to my dear family. My dear mom and dad... I can't thank you enough for always believing in me and your efforts to get me to where I am today. You are my biggest support in pursuing my dreams. I am grateful to you for all your efforts and for raising me as a role model to my dear brothers Emine and Remzi.

Seda AÇIKPORTALI

Çanakkale, 2022



ÖZET

İngilizce Öğretim Görevlilerinin Algılanan Dil Ölçme ve Değerlendirme Okuryazarlık Düzeylerinin Araştırılması: Önerilen Eğitim Müfredatı

Seda AÇIKPORTALI

Çanakkale Onsekiz Mart Üniversitesi

Lisansüstü Eğitim Enstitüsü

Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Zeynep Gülşah KANI

21/02/2022, 105

Araştırmanın amacı, Türkiye'nin farklı yerlerinde İngilizce hazırlık programlarında çalışan İngilizce okutmanlarının dilde ölçme ve değerlendirme eğitim düzeylerini ve eğitim ihtiyaçlarını keşfetmektir. Mezun olunan lisans programının dilde ölçme ve değerlendirme eğitim düzeylerine ve eğitim ihtiyaçlarına etkisinin araştırılması amaçlanmıştır. Karma araştırma yöntemi kullanılmıştır. Nicel veri toplama ile başlayan nitel veri ile devam eden sıralı açıklayıcı desen benimsenmiştir. Öcelikle, sınıf odaklı dil ölçme ve değerlendirme, ölçmenin amaçları ve içerik ve kavramları hakkında öğretmenlerin algıladıkları eğitim düzeylerini ve ihtiyaçlarını ortaya çıkarmak için Vogt ve Tsagari (2014) tarafından geliştirilen “Öğretmen Anketi” 80 İngilizce öğretim görevlisine uygulanmıştır. Nitel aşamada ise, eğitim ihtiyaçlarını ve dilde ölçme değerlendirme uygulamalarını daha fazla araştırmak için ankete katılanlardan 8 öğretmenle yazılı görüşmeler yapılmıştır. Anket verileri SPSS versiyon 25 ile hesaplanmıştır. Ortaya çıkan en önemli bulgulardan biri, tüm öğretim elemanlarının dilde ölçme ve değerlendirme hakkında sınırlı bilgiye sahip oldukları ve tüm alanlarda çok az eğitilmiş oldukları görülmüştür. En fazla eğitimin dilde ölçme ve değerlendirmenin içerik ve kavramlarında ihtiyaç duyulduğu algılanmasına rağmen en çok bu alanda eğitim aldıklarını belirtmişlerdir. Öğretim görevlileri, dilde ölçme ve değerlendirmeyle ilgili daha fazla eğitime ihtiyaçları olduğundan da bahsetti. İngilizce öğretmenliği bölümü mezunu olan ve olmayan öğretim görevlilerinin algılanan eğitim seviyelerinde önemli bir fark olsa da eğitim ihtiyaçlarında anlamlı bir fark bulunmamıştır. Bulgular

nitel verilerle de desteklenmiştir. Araştırmadan elde edilen bulgular göz önünde bulundurularak, anket ve görüşmelerde belirlenen tüm ihtiyaçları içeren amaca hizmet edecek şekilde dilde ölçme ve değerlendirme eğitim müfredatı hazırlanmıştır. Önerilen eğitim müfredatı, İngilizce öğretim görevlilerinin ihtiyaçlarına göre şekillendiğinden hizmet içi eğitimler için yol gösterici olacaktır.

Anahtar Sözcükler: Dilde Ölçme ve Değerlendirme, Hizmet İçi Eğitim, Sınıf Odaklı Ölçme, Ölçmenin Amaçları, Ölçme Değerlendirmenin İçeriği ve Kavramları, Alternatif Değerlendirme, Eğitim Müfredatı



ABSTRACT

INVESTIGATING PERCEIVED LANGUAGE TESTING AND ASSESSMENT LITERACY LEVELS OF ENGLISH LANGUAGE INSTRUCTORS: A SUGGESTED TRAINING SYLLABUS

Seda AÇIKPORTALI

Çanakkale Onsekiz Mart University

School of Graduate Studies

Department of Foreign Language Education / English Language Teaching Program

Thesis of Master

Supervisor: Assist. Prof. Dr. Zeynep Gülşah KANI

21/02/2022, 105

The aim of this research was to explore the perceived language testing and assessment training levels and training needs of English language instructors working in English preparatory programs in different parts of Turkey. It was aimed to investigate the effect of the program they graduated from on their training levels in and training needs for Language Testing and Assessment (LTA). A mixed-method research design was employed. A sequential explanatory design, starting with the quantitative data collection and proceeding with the qualitative data, was adopted in this research. Firstly, “Teachers’ Questionnaire” by Vogt and Tzagari (2014) was administered to 80 English language instructors to reveal the perceived training levels and needs of the instructors. With respect to the qualitative phase, written interviews were conducted with 8 instructors from those who participated in the questionnaire to further investigate their training needs and LTA practices. The questionnaire data were calculated with (SPSS). All instructors had limited knowledge about LTA and found out to be little trained in all the domains of LTA. From all the domains, the most training was perceived to be needed in the contents and concepts of LTA although they stated they received the most training in this domain. The instructors mentioned the need for further training sessions about LTA. There was a significant difference in the perceived training levels of ELT and non-ELT graduate instructors, while no significant difference was found

in their training needs. These findings were also supported by the qualitative data. An LTA training syllabus was prepared to serve the purpose. The training syllabus could be a guide for in-service training courses about language testing.

Keywords: Language Testing and Assessment, In-service Training, Classroom-focused Testing, Purposes of Testing, Content and Concepts of Testing, Alternative Assessment, Training Syllabus



TABLE OF CONTENTS

Jüri Onay	i
Etik Beyan.....	ii
Declaration	ii
Acknowledgements.....	iii
ÖZET	v
ABSTRACT.....	vii
TABLE OF CONTENTS.....	ix
THE LIST OF TABLES	xii
THE LIST OF FIGURES.....	xiv
CHAPTER I INTRODUCTION.....	1
1.0. Inroduction.....	1
1.1. Background of the Study	1
1.2. The Problem Statement.....	2
1.3. The Aims of the Study	3
1.4. Research Questions	3
1.5. The Significance of the Study	4
1.6. The Scope of the Study	4
1.7. Limitations of the Study.....	4
CHAPTER II.....	5
REVIEW OF LITERATURE	5
2.0. Introduction.....	5
2.1. Basic Terms: Testing, Assessment, Evaluation	5
2.1.1. Testing.....	5
2.1.2. Assessment.....	8
Assessing Grammar	11
Assessing Vocabulary	12
Assessing Reading	14
Assessing Listening	15

Assessing Speaking.....	16
Assessing Writing.....	17
2.1.3. Evaluation.....	18
2.2. European Language Portfolio and Self-assessment.....	20
2.3. Language Assessment Literacy.....	27
2.3.1. Language Assessment Literacy Research Conducted Abroad.....	29
2.3.2. Language Assessment Literacy Research Conducted in Turkey.....	32
2.4. Syllabus Design.....	37
2.4.1. Syllabus Design Studies.....	39
CHAPTER III.....	42
METHODOLOGY.....	42
3.0. Introduction.....	42
3.1. Research Design.....	42
3.2. Research Questions.....	43
3.3. Setting and Participants.....	43
3.4. Data Collection Tools and Procedures.....	48
3.5. Data Analysis.....	49
CHAPTER IV.....	51
FINDINGS.....	51
4.0. Introduction.....	51
4.1. Quantitative Findings.....	51
4.1.1. Results of Data Analysis for the First Research Question.....	51
4.1.2. Results of Data Analysis for the Second Research Question.....	55
4.1.3. Results of Data Analysis for the Third Research Question.....	58
4.2. Qualitative Findings.....	59
4.2.1. Interview Findings regarding Classroom-focused LTA practices of EFL instructors.....	59
4.2.2. Interview Findings regarding Purposes of Testing.....	61
4.2.3. Interview Findings regarding Contents and Concepts of LTA.....	63
4.2.4. Interview Findings regarding Self-Efficacy Beliefs of EFL Instructors.....	65
4.2.5. Interview Findings regarding Suggested Topics for LTA Training.....	66
CHAPTER V.....	68
DISCUSSION, CONCLUSION, AND SUGGESTIONS.....	68

5.0. Introduction.....	68
5.1. Discussion of the Findings.....	68
5.1.1. Discussion of the Findings of the First Research Question.....	68
5.1.2. Discussion of the Findings of the Second Research Question.....	71
5.1.3. Discussion of the Findings of Third Research Question.....	74
5.2. Conclusion.....	76
5.3. Pedagogical Implications.....	78
5.4. Limitations and Suggestions.....	79
REFERENCES.....	81
APPENDICES.....	88
APPENDIX A.....	88
THE QUESTIONNAIRE.....	88
APPENDIX B.....	92
WRITTEN INTERVIEW QUESTIONS.....	92
APPENDIX C.....	93
THE SUGGESTED TRAINING SYLLABUS.....	93
APPENDIX D.....	97
THE ETHICS COMMITTEE APPROVAL.....	97
APPENDIX E.....	98
THE QUESTIONNAIRE: WHAT IS MOST IMPORTANT IN TEACHING A LANGUAGE? (CHENG AND FOX, 2017, P. 23).....	98
APPENDIX F.....	104
ÖZGEÇMİŞ.....	104

THE LIST OF TABLES

Table No	Title	Page
1	Participants' Demographic Information	57
2	Interviewee's Demographic Information	59
3	The Normality Test Results	64
4	LTA Training Background of Instructors	65
5	Descriptive Statistics of Perceived Training Levels on Classroom-focused Language Testing and Assessment (LTA)	66
6	Descriptive Statistics of Perceived Training Levels on Purposes of Testing	67
7	Descriptive Statistics of Perceived Training Levels on Content and Concepts of LTA	68
8	Perceived Training Levels in LTA Domains	69
9	Descriptive Statistics of Perceived Training Needs on Classroom-focused Language Testing and Assessment (LTA)	70
10	Descriptive Statistics of Perceived Training Needs on Purposes of Testing	71
11	Descriptive Statistics of Perceived Training Needs on Content and Concepts of LTA	72
12	Perceived Training Needs in LTA Domains	73
13	Interview Findings-Classroom Focused LTA Practices	74
14	Interview Findings-Purposes of Testing	78
15	Interview Findings-Assessment of different language skills	80



THE LIST OF FIGURES

Figure No	Title	Page
1.	Data Collection Procedures and The Research Design	17
2.	Relationship between assessments/measurement/tests, their use for evaluation, and the consequences of assessment use	24
3.	Continuum of syllabi	49

CHAPTER I

INTRODUCTION

1.0. Introduction

This chapter starts with a brief discussion on the language assessment literacy research in the ELT world and discusses the underpinning topics. It is followed by the problem statement, the aim of the study, and research questions. The significance and the scope of the study are then presented. The chapter finishes with the limitation of the study.

1.1. Background of the Study

Teaching and assessing the knowledge of students are two interrelated areas in an educational context (Sevimel and Subaşı, 2019). Therefore, language testing and assessment have gained great importance in the field of English language teaching (Vogt and Tsagari, 2014), since there is a shift from traditional ways of assessing to more dynamic, student-centred, and alternative types of assessment. The teachers who have been already dealing with assessment procedures on a daily basis are expected to be up to date with the testing and assessment strategies and put their knowledge into practice. Thus, language assessment literacy holds center stage in the last decade. Being assessment literate- i.e., knowing different assessment procedures and using numerous and suitable assessment tools in different contexts- is of vital importance. Thanks to this, teachers can improve the quality of education, help improving the knowledge of students and achieving the goals of a curriculum.

A considerable amount of literature has been published on assessment literacy knowledge and training needs of English language instructors working at universities. The main aspects of the reviewed literature are assessment literacy perceptions, training levels, training needs, and

assessment practices (Fulcher, 2012; Giraldo Aristizabal, 2018; Hakim, 2015; Hasselgreen et al, 2004; Jannati, 2015; Vogt and Tsagari, 2014; Xu and Brown, 2017).

The research conducted in Turkey regarding perceived assessment literacy levels of teachers indicated that although the teachers received training or had little assessment knowledge, they stated they did not put their knowledge into practice. The recent evidence suggests that teachers need continuous, in-service trainings and workshops to be upto-date and improve their assessment literacy (Ballıdağ, 2020; Büyükkarcı, 2016; Mede and Atay, 2017; Öz and Atay, 2017, Yastıbaş and Takkaç, 2018).

Some of the research aimed to investigate the relationship between demographic information of teachers and their assessment knowledge (Büyükkarcı, 2016; Hasselgreen, 2004; Jannati, 2015; Kaya, 2020; Öz and Atay; 2017; Xu and Brown, 2017). However, in the reviewed literature, it is seen that very little attempt has been made to investigate the graduated program's impact on perceived training levels and training needs regarding language testing and assessment literacy.

Since language assessment literacy is still in its infancy, there is a need for more research (Ballıdağ, 2020; Büyükkarcı, 2016; Mede and Atay, 2017). On that account, the current research was conducted to contribute to the growing area of research by exploring the perceived training levels and training needs of EFL instructors working in English preparatory programs at non-state universities in Turkey and proposed a thorough language assesment training syllabus.

1.2. The Problem Statement

Since most English language instructors are provided with ready-to-use testing and assessment tools, and as there are test developers dealing with testing procedures in some institutions, the instructors are not generally interested in testing practices and may not develop

their knowledge regarding language testing and assessment. In this regard, exploring the instructors' perceived LTA training levels and needs has the utmost importance for improving the quality of the language education by suggesting a training syllabus. This thesis was written with these concerns to contribute to the English language research field.

1.3. The Aims of the Study

The purpose of this thesis is to reveal perceived language assessment literacy levels of English language instructors working in English preparatory programs in Turkey and suggest a training syllabus according to their needs. This designed syllabus includes theoretical and practical knowledge enabling them to improve their assessment skills and encourage them to adapt effective strategies for each language skill and their classroom practices.

1.4. Research Questions

This research study aims to answer the following questions:

1. What are the perceived Language Testing and Assessment (LTA) training levels of English language instructors working at English preparatory programs in Turkey?
2. What are the perceived LTA training needs of English language instructors working at English preparatory programs in Turkey?
3. Is there any significant difference between ELT and non-ELT graduates in terms of their perceived training levels and training needs on LTA domains?

1.5. The Significance of the Study

The current thesis is significant due to the fact that it proposes a language testing and assessment training syllabus for English preparatory programs in Turkey by indicating the specific training areas. This training program is a guide for institutions where English language is being taught and where testing and assessment activities are continuously conducted by instructors. Since data are collected from various universities in Turkey, the findings indicate the overall training needs and can be applicable to similar units in the Turkish context.

1.6. The Scope of the Study

The scope of this research is the assessment and evaluation literacy levels and training needs of English teachers working in English preparatory programs at non-state universities in Turkey. The designed training syllabus based on the perceived training levels, training needs and the assessment practices of English language instructors sets an example for the same programs of other higher education institutions. Moreover, it could be implemented in cooperation with the Ministry of National Education and Council of Higher Education for in-service teachers who need training about language testing and assessment areas. The research study also shows whether there is a significant difference between teachers who graduated from English language teaching departments and those who graduated from a department other than English language teaching.

1.7. Limitations of the Study

The limitation of the research study was the sample size since it was less than expected. Although the data were attempted to be collected from different regions of Turkey, i.e., Marmara, Aegean, Central Anatolia, Black Sea, Mediterranean and Eastern Anatolia, only 80 English language instructors participated in the questionnaire, and 8 of whom participated in the written interview.

CHAPTER II

REVIEW OF LITERATURE

2.0. Introduction

In the field of language education, measurement, test, assessment and evaluation are frequently used to define more or less the same activities. However, there are some differences among them. In the following chapter, the definitions of the terms will be explained in detail. Various test types, i.e., proficiency, achievement, diagnostic, direct, indirect, discrete-point, integrative, norm-referenced, criterion-referenced, performance-based, subjective, objective, computer-based test will be explained. The changing practices from traditional assessment to more student-centred and dynamic assessment, the functions of assessment, and the evaluation process will then be discussed. In addition to a variety of assessment practices, the assessment of micro-linguistic aspects and skills of language will be discussed. It will be followed by English Language Portfolio (ELP), and its role in self-assessment. Moreover, the pilot studies of ELP, its effectiveness on students' self-assessment, and syllabus design studies relating to language assessment will be presented.

2.1. Basic Terms: Testing, Assessment, Evaluation

2.1.1. Testing

Testing is explained as “a method of measuring a person’s ability, knowledge, or performance in a given domain” (Brown, 2004, p.3). Oftentimes, tests are associated with paper and pen, an examination room, and a nervous atmosphere, which in turn causes anxiety and concerns. However, there is a shift from traditional to more modern and humanistic ways of language testing (McNamara, 2000). These language tests may require students to create a written or oral portfolio. Moreover, they might be observed during usual in-class language activities in groups or in pairs in terms of their language abilities. Students may also be given an assignment

to complete outside the class. The tests can be conducted via the use of digital tools as well rather than paper and pen (Alvarez, 2016).

Although there might be several reasons for testing and while the reasons may change according to the context, Fulcher (2010) states five main purposes for testing the learners, which are achievement, proficiency, diagnosis, placement, and aptitude. The aforementioned purposes have resulted in different types of tests such as proficiency, achievement, diagnostic, placement, direct, indirect, discrete point, integrative, norm referenced, criterion referenced, performance-based, objective, subjective, and computer-based tests (Brown, 2004; Hughes, 2003).

To start with, proficiency tests aim to measure the language ability of a student regarding to what extent someone is proficient in the language and as to whether they are competent enough to function successfully (Hughes, 2003). Indeed, a proficiency test is defined as a test which is not confined to “any one course, curriculum, or a single skill in the language” but that tests overall ability (Brown, 2004, p. 44). Some of the well-known proficiency tests are the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), and the Cambridge First Certificate in English (FCE).

Secondly, achievement tests are designed to measure what has been taught in courses. Its main purpose is to reveal how much a student is successful meeting these course objectives. Hughes (2003) explains that there are two different achievement tests: progress achievement and final achievement. Progress achievement tests measure the student’s development throughout the course, while final achievement test is conducted at the end of the course.

Diagnostic tests are designed to find out students’ needs, strengths, and weaknesses on specific topics (Hughes, 2003). While achievement tests measure what has been already taught, the diagnostic, as the name suggests, aims to diagnose certain aspects of language such as intonation, sentence stress, or a specific grammar topic, and in this end it “should elicit information on what students need to work on in the future” (Brown, 2004, p.47).

When it comes to placement tests, they aim to place students to a specific program or level in accordance with their language abilities (Hughes, 2003). They are more likely to be used in specific teaching contexts since they may not work well in each school or institution.

Hughes (2003) makes a distinction between direct and indirect testing. While direct testing attempts to measure directly what skill has to be measured, indirect testing is related to testing the underlying skills of the learners.

Discrete-point tests measure language components separated from its context and in discrete language parts such as the skills of reading, morphology, phonology, or syntax (Brown, 2004). On the other hand, integrative testing intends to measure more than one language skill at once.

The other category for testing is criterion versus norm-referenced tests. Criterion-referenced tests are conducted to give feedback to students about specific lesson aims, whether they are achieved or not, and to find out where weaknesses and strengths are. Student performance can be compared with learning objectives (Gómez, 1999). However, norm-referenced tests are carried out to compare students' scores with all the other test-takers. The score is presented in such a way that students can see the rank order and the percentile rank (Brown, 2004; Gómez, 1999; Hughes, 2003).

In addition to classic paper and pen tests, performance-based tests are also preferred in the last years. It includes written or oral production, interactive, and group performance tasks (Brown, 2004). Thanks to performance-based tests, students perform the language skills, and the teacher observes them on tasks immediately.

On the issue of subjective and objective testing, the method of scoring determines the test type. While the teacher scores some productive skills' tests such as paragraphs, essays, monologues, dialogues, and oral presentations with a holistic or impressionistic rubric in the subjective testing, there is a room for subjective evaluation. However, there is no need for the

teacher's judgement in objective tests such as multiple choice where there is a determined answer. Teachers just follow the answer key and do not make any decision on students' tests.

Computer-based testing has started to be adapted as it is more convenient than classic paper and pen tests in terms of its quick computer-based scoring and administration (Alvarez, 2016). One of the advantages of computer adaptive tests is that the score is given upon the end of the tests. Therefore, it has captured the attention of teachers and students. In addition to immediate results it gives, efficient administration, high quality item development, and authenticity are among the benefits of computer-based testing (Alvarez, 2016).

To conclude, tests are the instruments used to gather information about learner abilities. As mentioned above, there are several types of language tests attempting to answer different purposes in different contexts. Due to changing practices in language teaching with the development of technology, the test types are varied as well.

2.1.2. Assessment

Assessment is defined as the process of gathering information about what we are interested in (Bachman and Palmer, 2010). That is to say, in consequence of the information gathering process, there is a verbal statement or a score. Compared to testing, it is a continuous process (Brown, 2003), and is distinguished from other ways of collecting information due to its systematicity and substantive grounding (Bachman and Palmer, 2010). It is systematically designed through determined procedures, which allows other researchers or teachers to apply it in different occasions and to different students. The second feature, being substantively grounded, suggests that it must depend on some criteria such as course content, research, or theory. To summarise, in language education, assessment is the process of systematic data collection about students' language abilities. The assessment procedure can involve different testing activities such as observations, quizzes, portfolios, spoken and written products (Coombe, 2018).

Language assessment habits have been changing over years with the changing teaching practices, methodologies and perspectives. Therefore, there is a shift from more traditional assessment practices to more communicative and student-centred ones.

To start with, formal assessment is the type of assessment that follows certain formal procedures. They are planned systematically and happens at specific times to give information about the language achievement of students (Brown, 2004). It can be a part of a formal lesson. To exemplify, it can take the form of standard-based performance assessments where students' products are scored with a rubric (Coombe, 2018). On the other hand, informal assessment does not have a regular schedule, and students are not assessed according to specific formal regulations as in formal assessment. Students can be observed during class activities, and the assessment can be conducted through verbal or written comments such as 'Well done!, Good job!, etc.' (Brown, 2003).

Another distinction in the assessment is about its functions. If the assessment is conducted to amend students' language abilities in their future learning by the teacher's feedback on students' strengths and weaknesses, it becomes formative assessment (Bachman and Palmer, 2010). With the formative assessment, teachers can take some formative decisions such as correcting the response of students, changing question types in the tests, and finding out the topics needed to be practiced more. In addition to teachers, students can also take some formative decisions such as the fact that they can change their study strategies or spend more time on specific topics (Bachman and Palmer, 2010). On the contrary, summative assessment does not necessarily mean that it will shed light on students' future learning activities since it aims to measure what has been learned at the end of the unit, or a lesson. It presents the final evaluation of the learning activities of the student. As a result of this, summative assessment is generally in the form of formal assessment such as final tests, midterms exams, and proficiency tests (Brown, 2004; Coombe, 2018). According to the summative assessment, teachers can make decisions about whether students can pass the course, or about whether they are knowledgeable enough to be certified for their abilities (Bachman and Palmer, 2010).

In recent years, there has been an interest in the alternative ways of language assessment as a response to traditional paper and pen assessment practices (Douglas, 2009). Traditional assessment can be defined as ‘one shot, standardised, timed, multiple-choice format, decontextualised test items, norm referenced scores, summative, non-interactive, oriented to product, fostering extrinsic motivation’ (Brown, 2004. p.13). However, alternative assessment implies that assessment should be the extension of regular in-class activities, and it must require learners to use the language in the real-like contexts. It must be based on the process rather than products. In addition to these, the products must be assessed by teachers, but not by computers (Douglas, 2009). Therefore, this type of assessment prioritises individual students’ development. Different from traditional assessment practices, in the alternative assessment students are evaluated holistically with regards to their performances. Indeed, the alternative assessment demands a long-term and continuous assessment process. Unlike the traditional assessment, it is contextualised with communicative activities and provides interactive performances. As a result of this assessment, students are provided individual and formative feedback for their language learning (Brown, 2004). Some of the assessment tools are portfolios, self and peer assessments, essays, presentations, interviews, and projects (Coombe, 2018).

Dynamic assessment is a relatively new term which has been discussed for more than three decades (Leung, 2007). It is based on Vygotsky’s ZPD concept and sociocultural theory (Coombe, 2018; Leung, 2007; Poehner, 2005; Poehner, 2008). It is used to evaluate what students can do with their own potential and also when they are supported by their teachers and peers. The other forms of assessment are criticised due to the fact that the performance of a student on any task may not be representing the bigger picture; that is to say, what they can do in the future (Leung, 2007). To this end, social interaction is seen as an important criterion.

To conclude, it is seen in the literature that language assessment has been shaped as a result of changes in perspectives, methodologies, and teaching practices. Although in recent years, there is an inclination to more communicative, alternative and dynamic assessment types, institutions and teachers need to decide on what type of assessment to adopt according to their aims, needs and contexts since each context has distinctive features.

Assessing Grammar

Grammar, i.e., the structure of a language, is at the heart of use of language (Purpura, 2004). It lays the foundation of reading, listening, speaking and writing skills. Grammar teaching has been a major area of interest in the field of language teaching and there has been much debate about the grammar teaching and assessing (*ibid.*). Before the 20th century, it was argued that a language could be best learned deductively, which requires memorizing and reciting the rules and explanations of the target language. On the other hand, others put forward that language could be best learned inductively. In this approach, students were expected to figure out the rules and prescriptions of the language from the examples in the target language. Another traditional approach claimed that language could be mastered with translation from one to another. It was believed that languages were similar and translation could help learning the target language. However, in these approaches, the grammar knowledge was based on the set of rules. The assessment was conducted by depending on students' reciting rules, translating texts, and analysing texts to formulate the rules (*ibid.*).

In the late 20th century, there was an increasing interest in the communicative purpose of grammar teaching. The criticism for the previous traditional approaches was that grammar was not only learned but also used for communicative and linguistic purposes (*ibid.*). Students could use grammar where they could ask and answer questions, compose sentences and write paragraphs. With this perspective, the assessment of grammar shifted from reciting and formulating rules and translating to the ability to apply them to different linguistic contexts. That is to say, it was regarded as set of rules used for communication. In this way, the assessment was conducted depending on the tasks where students proved their abilities in speaking and writing (*ibid.*). To assess the grammatical knowledge, Purpura (2012) suggested that students were provided with tasks enabling them to prove their receptive and productive grammar knowledge. It was important to keep in mind that grammar must be assessed both in form and meaning in sentence and discourse levels. Moreover, Purpura (2004) regarded grammatical ability as “the combination of grammatical knowledge and strategic competence; the capacity to realise grammatical knowledge accurately and meaningfully in testing or other language-use situations” (Purpura, 2004, p. 98). While the grammar tests give information about students' grammatical performance, the main aim must be

to make judgements about the underlying ability and how students are able to use the knowledge to transfer meaning (*ibid.*).

In addition, grammatical form, and grammatical and pragmatic meaning must be considered while constructing the assessment activities. Therefore, while designing tests, grammatical knowledge, ability, and performance must be considered. Bachman and Palmer (1996) presented a framework including task characteristics and performance. This included setting characteristics, test rubrics, input, expected response and the input and response relationship (Bachman and Palmer, 1996). It was suggested while designing test tasks “to adapt a variety of them such as selected response tasks, i.e., multiple choice, true/false, matching, etc., limited-production tasks, i.e., gap-filling, cloze, short-answer, dictation, information-gap, extended-production tasks, i.e., summaries, essays, dialogues, interviews, stories, reports, problem-solving and decision-making activities” (Purpura, 2004, p. 127). These test tasks could also be categorized as subjective and objective tests. Objective tests are the ones which are evaluated according to correctness criteria without personal judgement, while subjective test tasks depend on the personal judgements according to the criteria.

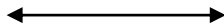
To conclude, it is of importance to construct test specifications, consider the time allotment, the setting, outcomes, evaluating criteria, and adapt a variety of test tasks in assessment activities where grammatical knowledge, ability, and performance are all assessed.

Assessing Vocabulary

Words are the foundation of sentences, paragraphs, and texts (Read, 2000). A great number of learners consider that the first step of language learning is vocabulary learning. A great deal of time is spent to learn vocabulary. Moreover, vocabulary teaching is seen as a fundamental area in language teaching, which also requires language teachers to assess vocabulary knowledge of students. As in the Figure 1, there are three different vocabulary assessment perspectives (*ibid.*). Rather than providing a model for assessing vocabulary, Read (2000) provided a variety of procedures in the use of discrete and also more integrative vocabulary assessment types.

Discrete

A measure of vocabulary knowledge or use as an independent construct

**Embedded**

A measure of vocabulary which forms part of the assessment of some other, larger construct

Selective

A measure in which specific vocabulary items are the focus of the assessment

**Comprehensive**

A measure which takes account of the whole vocabulary content of the input material (reading/listening tasks) or the test taker's response (writing/speaking tasks)

Context-independent

A vocabulary measure in which the test taker can produce the expected response without referring to any context

**Context-dependent**

A vocabulary measure which assesses the test-taker's ability to take account of contextual information in order to produce the expected response

Figure 1. Dimension of Vocabulary Assessment (Read, 2000, p. 9)

In the first dimension of vocabulary assessment, vocabulary knowledge is intended to be measured. That is to say, the teachers could make decisions about the results such as to what degree the students have learned the words in the units, or they could see whether they can infer meanings of the words from the reading passages. These discrete vocabulary tests are the most common test types which handle the words as independent constructs and which separate them from other language components (*ibid.*). On the other hand, the embedded vocabulary assessment type is

based on larger constructs. In this perspective, vocabulary assessing is embedded in other components of language such as deducting the meaning of words from a reading passage or while writing an essay. While in discrete tests the focus is on the construction of the knowledge of vocabulary, in embedded testing, the assessment does not barely depend on vocabulary items but measures the knowledge of other components of language (*ibid.*).

The second-dimension, selective vocabulary assessment, is based upon the selected set of words to assess the knowledge and use of vocabulary. On the other hand, comprehensive assessment depends on all the contents of written and spoken text. For instance, in a writing task teachers could look at all the low-level or sophisticated words and evaluate the vocabulary knowledge and use of students (*ibid.*).

As for the last dimension context is the foundation. Read (2000) stated that the whole text and more broadly discourse is of importance to present the meaning of the vocabulary. Vocabulary assessment in cloze tests and speaking or writing tests are examples of context-dependent measurement (*ibid.*). In context-dependent vocabulary assessment the students make use of the contextual information. If they do not take advantage of the contextual clues and are assessed as if the vocabulary is in isolation, the vocabulary assessment is context independent.

Assessing Reading

Although reading is a receptive skill, it requires readers to actively participate in the process by bottom-up processing, i.e., “where the reader begins with the printed word, recognises graphic stimuli, decodes them to sounds, recognises words and decodes meanings” (Alderson, 2000, p. 16) and top-down processing, i.e., using background knowledge and information depending on the different contexts to comprehend the reading. Alderson (2000) mentioned some implications for the assessment of reading. Students must be presented with the content-focused tests that are related to their interest, academic backgrounds. They should be presented longer texts rather than short pieces and these enable them to enjoy their reading experience. Moreover, the tasks should be doable and not discouraging due to the level of difficulty. Also, students are given chance to enable their background knowledge to interpret the texts. Test makers should be aware of the

potential of multiple interpretations and understandings. In addition, extensive reading should not be weakened by assessment activities. Portfolios could be a way of evaluating.

There is no best method for assessing reading, and there are plenty of techniques used for this purpose. Cloze-tests, gap-filling, multiple choice, matching, ordering tests, true/false items, short answer, summary, information-transfer tests are some of the techniques used for assessing reading.

Assessing Listening

Listening comprehension is a “very complex process” (Buck, 2001, p.1). Similar to reading, listening is also a receptive skill, and it is an active process where learners make meaning from the sounds. However, testing listening is a lot more challenging, complicated, and time consuming than reading (Buck, 2001). It was also discussed that listening comprehension required many sub-skills.

There are three approaches for assessing listening: discrete point approach, integrative approach, and communicative approach (*ibid.*). To start with, discrete point approach is used to test isolated and separated parts and bits of language. As for listening, it is testing the phonemes, intonation, stress, vocabulary etc. True/false, multiple-choice, phonemic discrimination and paraphrasing tests are some of the commonly used discrete-point approach techniques (*ibid.*). On the other hand, the integrative testing attempts to measure the language as a whole rather than bits of language as in discrete point approach. Cloze tests and gap-filling activities are the most common techniques of the integrative approach. This approach is accepted as a more semantic and literal meaning-focused approach. The third approach, the communicative approach, focuses on the meaningful target language use (*ibid.*). Rather than focusing on how much the learner knows, this approach is concerned with successful communication. Communicative test items are based on authentic texts and authentic tasks. Buck (2001) compiled some suggestions for listening assessment and stated listening tests must compromise of texts carrying the linguistic characteristics of spoken language. Students can understand the basic linguistic information in a variety of texts. Grammatical knowledge must not be an obstacle. In addition to the literal meaning,

it is also important to include the tests which requires inferring meaning. Moreover, rather than inferences depending on the common sense, inferences based on the texts should be included.

Assessing Speaking

Speaking is a significant area in language teaching and thus, assessing speaking is important and challenging at the same time (Louma, 2004). There are many factors in the speaking cycle such as interlocutors, students, raters, and scales all making the process challenging. That is, while students are being assessed in face-to-face communication, there are such factors as time constraints, the need of self-monitoring, planning of the speech, and lack of grammatical and lexical knowledge (Goh, 2016). Speaking is a unique skill and occurs in an interactive nature and it requires human scoring. Therefore, it is quite important to achieve the validity and reliability of the evaluating process (Louma, 2004). As Luoma (2004, p.174) defined “reliability relates to the consistency of the scores, and validity to their meaningfulness for the intended uses.”. To be able to conduct valid and reliable scoring, the teachers must have a deeper understanding of the nature of speaking, and accordingly design a task and arrange the scoring (Csépes and Fekete, 2018). To provide a reliable and valid scoring, rating scales could be employed (Csépes and Fekete, 2018; Luoma, 2004). With the help of rubrics, a variety of criteria including vocabulary, interactive communication, grammar and pronunciation, and task achievement could be assessed. Moreover, as speaking requires subjective evaluation, providing “intra-rater or internal consistency, which means that raters agree with themselves, over a period of a few days, about the ratings that they give” (Luoma, 2004, p. 179) is significant. In addition to reliable and valid scoring, it is significant to assess students in real-life situations where they can perform the functions of the language (Csépes and Fekete, 2018).

Speaking tasks could be topic-based, or could depend on visuals, and could be conducted as monologues, paired mode, group mode or interlocutor mode (Csépes and Fekete, 2018). There are a variety of speaking tasks: description, narrative, instruction, comparing and contrasting, explaining and predicting, decision-making, role-play and simulation, reacting in situation, and structured speaking (Luoma, 2004). However, while designing a speaking task, it is of importance to choose what to test, and this of course depends on the teaching methodology and approach.

Whether the principles of testing and teaching match must be considered while designing tasks (Luoma, 2004). In addition to valid tasks, writing task specifications are significant since it will help the test designers cover the skills which have been covered in the class. That is to say, the task specification includes the instructions, the administration plan, the materials needed for the task, skills to be assessed, timing, and structure (Luoma, 2004). The language in the task instruction must be carefully written, as it could guide the test-takers wrongly or could be vague.

Assessing Writing

Similar to the speaking skill, writing is a productive skill. While even writing in one's mother language is difficult, writing in a second language is much more difficult (Csépes and Fekete, 2018). As the written language is composed of complex structures, assessing the written language is challenging. Designing writing tasks requires meticulous preparation keeping in mind the task itself, the students who write it, and the scoring (Csépes and Fekete, 2018; Weigle, 2002).

Weigle (2002) highlighted the importance of purpose before designing the task. It is due to either drawing conclusions about the language proficiency or making decisions depending on the inferences. In writing assessment, grammatical, textual, functional, and sociolinguistic knowledge is evaluated at the same time. Although the language knowledge is mostly assessed, the other components should not be ignored while designing writing tasks (Weigle, 2002).

Writing assessment is a performance assessment in which the learner's performance of the language is being assessed (Weigle, 2002). Bachman and Palmer (1996) draw attention to the qualities of tests: construct validity, reliability, interactiveness, practicality, authenticity, and impact. In addition to the previous qualities, there is also the washback effect of the tasks. Weigle (2002, p. 54) defines, positive washback as "any effect of a testing procedure that encourages teachers to adopt practices that are in line with the current best thinking in the fields with respect to pedagogy". Moreover, Bachman and Palmer (1996) noted that the washback effect for students could depend on some factors such as students' preparation, test-taking process, and the received feedback about the performance and the decisions regarding the results. Furthermore, it is

significant to take into consideration the way the students take the test, the quality of the feedback, and reliable and valid scoring to be able to make accurate decisions about the results.

Similar to speaking, writing requires subjective assessment, and it is critical since the results are used to make decisions about the students. To conduct a reliable scoring, rating scales which are either holistic (depending on overall impression, a score is given to the whole text accordingly) or analytic (texts are evaluated according to multiple criteria) should be employed (Heaton, 2011; Weigle, 2002). Also, to maintain high reliability in writing assessment, the texts could be assessed by two raters. In addition, the interrater reliability could be cross tabulated so that the extreme scores are checked again. When it comes to the validity of scoring, whether the scale is suitable to the task to assess it properly and whether the raters are based too much on their judgements rather than the scoring scales are to be questioned. Another important issue to question is the practicality and the usefulness of the scores gathered. Teachers must also consider if the scores allow them to make proper decisions about the students (Weigle, 2002).

For writing assessment, story writing, controlled writing, letters, emails, photograph description, summary, drilling, objective tests for mechanics, essay are frequently used tasks (Heaton, 2011; Weigle, 2004).

2.1.3. Evaluation

Evaluation can be made based on the information gathered from the student. Indeed, it is the process of making decisions based on assessment, measurement, and tests (Bachman and Palmer, 2010; Douglas, 2009). Evaluation is a general term than testing and assessment (Douglas, 2009). Figure 2 represents the relationship between assessment and evaluation.

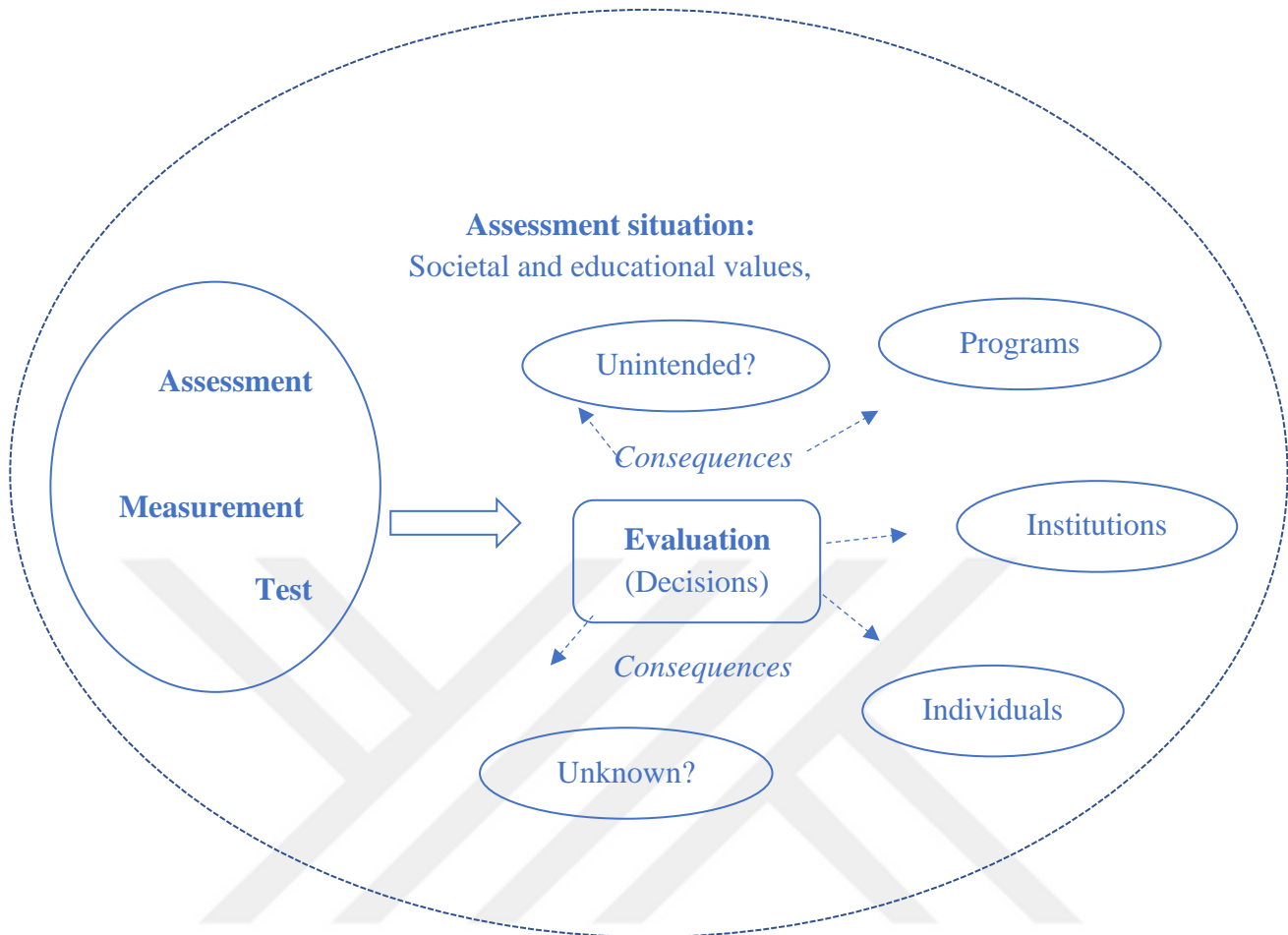


Figure 2. Relationship between assessments/measurement/tests, their use for evaluation, and the consequences of assessment use (Language Assessment in Practice, Bachman and Palmer, 2010 p.22)

It is clearly seen that there is a strong relationship between assessment and evaluation. Assessment is used to find out learners' strengths and weaknesses, to put them in the correct instructional program, or to make a decision about whether they should pass or fail. These decisions are categorised under three headings which are decisions made about individuals, programs and research (Bachman and Palmer, 2010). The decisions about individuals may be about placing the learner to a program, providing a certificate for proficiency, amending instruction, improving teaching activities, or determining passing or failing learners. When it comes to the second category, it is stated that some decisions can be made for programs such as strengthening existing programs or finding out whether there is a need for a completely new program. As for the last category, some changes can be made for the methodology you have adopted and your views about the language teaching.

2.2. European Language Portfolio and Self-assessment

In addition to the more recent assessment types, i.e., alternative assessment and performance-based assessment, self-assessment has taken place in the assessment practices lately. Whether it is used or effectively used or how much attention is paid to it in the classroom are questionable. European Language Portfolio (ELP) is one way for fostering self-assessment and is a very detailed type of documentation of the learning process.

ELP is a document that language learners reflect on their own language learning experiences and achievements (Council of Europe, 2006; Roman and Soriano, 2015). It is accepted as a project of Council of Europe in the modern languages field. This project aims to foster:

- the deepening of mutual understanding among citizens in Europe;
- respect for diversity of cultures and ways of life;
- the protection and promotion of linguistic and cultural diversity;
- the development of plurilingualism as a life-long process;
- the development of the language learner;
- the development of the capacity for independent language learning;
- transparency and coherence in language learning programmes.

(Schneider and Lenz, 2001; p.2)

According to the ELP guidelines, the project has two goals: to encourage learners by diversifying and extending their language skills and to keep a record of their acquired cultural and linguistic skills (Schneider and Lenz, 2001). It is of importance to remember that ELP is not a substitute for language textbooks, curriculum, and tests but can guide and help to design a useful and effective curriculum.

ELP consists of three parts: language passport, language biography and dossier. To start with, language passport allows for the review of a learner's proficiency in different languages. The skills are reviewed according to Common European Framework (CEFR) reference levels. In line with the reference levels, the learner's competencies, intercultural experiences and learning process are evaluated. Language passport indeed provides teacher assessment, self-assessment,

and all the other assessments conducted by institutions. This record of performance gives information about what is assessed, when, and by whom (Schneider and Lenz, 2001).

As for the language biography, it does not completely have a pedagogic function. It allows for learners to take part in their own learning adventure by planning, and assessing the process, and reflecting upon their progress, and in turn this documentation provides the planning of further learning activities and detailed personal information. It can consist of the following components:

- a) a personal and more or less detailed biography covering language learning and socio- and intercultural experiences;
- b) checklists related to the common reference levels;
- c) checklists or other forms of descriptions of skills and competencies that are not related to the common reference levels;
- d) planning instruments such as personal descriptions of objectives.

(Schneider and Lenz, 2001; p.20)

The dossier is the selection of materials indicating the experiences and achievements of the learner registered in the language passport and biography. In the dossier, the learners can keep the activities, worksheets, texts from their reflection, self-assessment, audios, checklists, certificates, etc. This trend is the same with portfolio assessment (Schneider and Lenz, 2001).

In the literature, there are several studies about the effectiveness of the ELP, results of pilot studies, and self-assessment and the ELP. Little and Perclova (2001) put forward that self-assessment is indeed at the heart of the ELP. With the descriptors originated from CEFR, the learners assess their proficiency levels in the language passport. They set their learning goals in biography by assessing their own performance. For the dossier, they decide on what to include and exclude, which in turn requires self-assessment. Considering the role of self-assessment in the ELP, it is clear that it leads to autonomous learning.

The ELP was piloted as a project in various countries at all educational levels, i.e., adults' education, university, vocational school, and secondary, and primary schools between the years of

1997 and 2000 (Little and Perclova, 2001). In the report that Little and Perclova (2001) wrote, they compiled the views of teachers. They stated that both the teachers and students benefited from the self-assessment due to the fact that it gave the students self-control and let teachers know about the learner problems. It was uttered by Czech teachers of German and English that students were more self-confident, creative, voluntary to participate and had good relationships with the teachers and classmates. In addition, they could reflect on what they did and noticed they could improve their English outside the class as well. Apart from the students, teachers were also more creative and had a chance to report the students' performance to the parents (Little and Perclova, 2001). Besides, the teachers reported that weaker students could take an advantage of the ELP since it increased their confidence.

While it had advantages, it was also found problematic. Teachers were concerned about the way students assessed themselves and highlighted the reliability problems and their less knowledge of self-assessment. However, self-assessment has three focus areas. The first focus is on the learning process. The learners should be able to assess their progress, i.e., how well they perform in the learning tasks and how much they meet the learning objectives at specific levels. This depends on the learners' own judgement, e.g., I think this activity is boring, I made an improvement at this level. Little and Perclova (2001) warned that this kind of assessment is quite subjective depending on the views of students. This corresponds with the reflective approach, and the teachers must not have an aim to test the reliability of this type of assessment. The second aim of the self-assessment in the ELP is measuring learner's communicative proficiency. Learners are able to decide on what they can or cannot do with the descriptors adopted from the CEFR. In addition to communicative proficiency, linguistic proficiency is the other focus of self-assessment. This could be more difficult than measuring communicative proficiency for learners. However, Little and Perclova (2001) suggested some ways to ease the process such as the fact that students can correct their own and one another's linguistic output. By using the scales to correct or evaluate their own and one another's product, they will get familiar with the methods of self-assessment and formal assessment.

Little (2005) mentioned the relationship between the ELP and self-assessment. Learner-centered language education, aiming to enhance learner autonomy by entailing the learner making

decisions about the content, aims and learning methods, gives a prominent role to learners for self-assessment (Little, 2005). The CEFR and its adaptation to the ELP indeed aim to facilitate the culture of self-assessment. The importance of self-assessment in language teaching depends on three factors according to Little (2005). First of all, it is required for a learner centered curriculum. It is dissatisfying when the learners are included in decisions regarding the content and methodology but excluded from evaluating their own progress and achievement. Secondly, a learner centered pedagogy brings learner autonomy. If the students are involved in the process of setting objectives and selecting learning materials and activities, then they can develop a sense of self-assessment. They can be more knowledgeable about the upcoming assessment. Thirdly, the language learned in the classroom can go beyond the class, and the learners can take an advantage from it by reflecting on their own learning thanks to self-assessment. Little (2005) concluded that self-assessment gives learners insider roles, thus enabling the interaction between assessment and curriculum. He implied that the assessment by teachers or other authorities cannot be replaced by self-assessment. On the other hand, self-assessment culture may be developed for effective language learning, and this can be provided by the adaptation of the ELP.

Stoicheva et. al (2009) compiled the results of these studies from the UK and some Baltic countries and categorised the main results by focusing on some key areas, i.e., the ELP use in the classroom, the effects of the ELP on the materials, assessment, language policy and on other projects. Generally, the views about the ELP were positive, implying that it brought the sense of a common European product and gave trainers, teachers, and heads of the departments examination committee roles such as deciding on their aims and practices according to the CEFR, and there was a shift to more student-oriented teaching. However, some concerns regarding the lack of support from the institutions and national level authorities were raised. When it comes to the views about the use of the ELP in the test and exams, there was an ambiguity about self-assessment in reference to ELP checklists, exams and tests. The ELP was considered irrelevant when a school's examination was different than a national exam system's since graduating from the school did not depend on "can-do statements". Moreover, some stated working with the ELP was not necessary when the courses had clear specifications for examination, and it was found as an overload. The overall point was the fact that the ELP checklists and national tests were not in accordance with each other. On the other hand, the ELP plays a crucial role in designing curricula, and assessing

the students. In setting objectives to non-accredited courses, the use of the ELP could be of importance.

With the introduction and the implementation of pilot studies in the ELP and with the increasing importance of self-assessment, there has been a growing body of research on the topic. Ziegler (2014) explored whether the ELP provides validity for self-regulated learning. He worked with 575 students and 19 teachers in 6 different schools. The research had a quasi-experimental design, and the researcher designed two groups: a control group and an experimental group. In the experimental group, students were using the ELP, while in the control group students were not using it. Data were collected through student and teacher questionnaires and interviews. The aim was to find out whether there was a significant difference between the control and experimental groups in term of self-regulated learning. Students in the experimental group that were using the ELP during the learning process were reported to have higher mean values in self-academic efficacy, task value, self-regulatory efficacy, and goal orientation. In addition, Ziegler (2014) reported the more the frequency of using the ELP in the classroom was, the more significant the difference was within the experimental group. It was concluded that the ELP achieved its pedagogical goal. The students who used the ELP were more self-motivated and self-regulated. Besides, the students who liked using the ELP were more autonomous than the others. The teachers in the experimental group participated in a 2-year training program about the ELP before the implementation in the classroom, and the teacher interviews indicated that to implement the ELP successfully in the classrooms training was a must. The research approved that the ELP supported the self-regulated learning via self-assessment of their abilities and proficiency levels, and it helped in setting the goals and building strategies for learning.

Hişmanoğlu and Hişmanoğlu (2010) attempted to find out to what degree the ELP supported self-assessment and reflective learning in English for Specific Purposes (ESP) lessons at tertiary level. The research was conducted in European University of Lefke. Fifty students received a 2-hour-ESP class each week for 15 weeks and were introduced with the concept of the ELP via classroom discussions about the use of the ELP, its functions, and the characteristics of a dossier. During the course, the ELP developed by the Ministry of Education was used. The students were given questionnaires both before and after the ESP course. Results revealed that the ELP

provided self-assessment by enabling them to learn to reflect on their own learning. Nonetheless, some of the learners showed pessimistic reactions when they were asked to assess themselves. Hişmanoğlu and Hişmanoğlu (2010) suggested constant support in the use of the ELP not only for teachers but also for educational authorities. Last but not least, the researchers concluded that learners' reflections were positive, and this kind of new assessment, self-assessment, could be beneficial for their long-term language learning goals.

Different from the previous literature, Pérez Cavana (2012) investigated the electronic ELP use in terms of learning styles through self-assessment. The Electronic ELP was developed for the Development of Languages at the Open University. Seventeen students participated in the research. Data were collected via semi-structured online interviews. The focus of the interviews was to shed light on the pedagogical usefulness of the ELP, the structure of the electronic ELP, learning styles and intercultural experiences of learners. Both the qualitative and the quantitative results revealed that students were more aware of what their proficiency level in the target language was and of what strategies they would love to adopt in learning a language. Moreover, thanks to the electronic ELP, they could reflect on their language learning experience. It was suggested that the ELP could be an alternative to diagnostic tests and a significant tool to assess learning styles.

Since the concept of the ELP in language teaching was recent in around 2012, Román and Soriano (2015) attempted to analyse the impact of the ELP once it was first applied in Spain. The researchers analysed the students' use of ELP and self-assessment, compared the ELP usefulness and investigated the encouragement of plurilingualism from students' and teachers' perspectives. 25 fifth-grade primary school Spanish-speaking students participated in this case study. Data were collected via various tools such as classroom observations, video recording while students were filling in the language biography, field notes about the use of the ELP, questionnaires for students and the teacher to explore the pedagogical functions of the ELP in relation to self-assessment and autonomous learning, and interviews focusing on the teacher beliefs about the students' autonomous learning, considerations about self-assessment and the usefulness of the ELP. Results of the research revealed that much more time was needed to use and adapt the ELP to make students autonomous learners. Students were not able to assess themselves properly, and Román and Soriano (2015) stated that strategies used to enhance self-assessment were ineffective.

Students did not fully understand the concept of self-assessment and needed much more time to be able to use them effectively. In this sense, the reliability of self-assessment in the ELP was not achieved. The teacher, on the other hand, was more optimistic, and she stated that the ELP was a useful tool if the required effort and time were given to it. She also uttered the importance of collaboration with the colleagues. As a result, while autonomous learning improved, self-assessment was found ineffective due to the fact that that it required high cognitive skills.

Like Pérez Cavana (2012), Bertolotti and Beseghi (2016) adopted the electronic version of the ELP depending on the principles of learner autonomy, reflective learning, and self-assessment. The Language Centre of the University of Parma in Italy developed a self-study program in 2011. In this program, students followed a personalised program according to their specific needs. The program started with an advising session in which they expressed the difficulties they had encountered in language learning, their previous experiences, and abilities. In the second session, they created their own learning plan under the guidance of teachers. Being not traditional, the plan allowed for specific focus areas according to the students and could be changed regarding the skills where practice was needed more. From the beginning of the self-study program, students were supposed to keep record of what they did and learned, and of how they felt about their competencies and progress. All these entries were supposed to be written in a learning diary with the aim of reflecting on the learning process and at the same time providing concrete evidence for students' real achievements. After a few years, the Centre decided to improve the self-study program by integrating the learning plan and the diary into the online form of the ELP (e-ELP). This was done on an online English laboratory where the students could share their own experiences with one another. e-ELP consisted of a language passport, a language biography, and a dossier. Students and teachers were first introduced with this new tool and used it during their learning. This time online diaries, exercises and files were uploaded to the online portfolio. The aim of the online pilot program was to let students self-assess themselves according to the CEFR items in line with their preset aims in the language biography. 20 students participated in the pilot study, and they all passed the exam at the end of the year. The results of the pilot study indicated that students improved their autonomy by experiencing the language learning with an online tool. They became more aware of their competencies and weaknesses thanks to the language biography, the language passport, self-assessment and reflection. The questionnaires showed that students had

positive attitudes towards adopting an online tool and that e-ELP enhanced their motivation to learn the language. Students also implied the significance of self-assessment and collaborative learning.

Şentürk (2017) explored the effect of self-assessment in the ELP, a learner style inventory and end of unit checklists to English language learning. Data were collected via a student attitude questionnaire and semi structured interviews. Thirty students and five instructors participated in the research from the School of Foreign Languages. Both the quantitative and qualitative data showed that the ELP, the learner style inventory and unit checklists improved self-assessment when they were used effectively. Students were willing to use the ELP for self-assessment. On the other hand, although the instructors stated that it promoted self-assessment and created learner-centered classrooms, they were reported to have little knowledge about the ELP since it was a relatively new concept in Turkey.

To sum up, it is evident ELP is of importance because it fosters self- assessment by enabling learners to become active agents in their language learning adventures. The review of the literature indicates that they become more aware of their strengths and weaknesses and were more into language learning activities. Moreover, it promotes collaborative learning via peer-assessment. It helps learners set goals for their own language learning by showing their progress and revealing their purposes to learn the language. However, questions have been raised about the effective use of ELP during language teaching. The literature indicates that EFL teachers had little knowledge of ELP and self-or-peer assessment and highlighted the necessity of trainings in the use and adaptation of ELP. By taking the relationship between ELP and language assessment into account, it is significant to pay attention to the definition of assessment literacy and the research literature with regards to language testing and assessment literacy of EFL instructors. In the next part, these issues will be discussed in detail.

2.3. Language Assessment Literacy

It is clear that language testing, assessment and evaluation are in a strong relationship. As mentioned in the parts above, they provide beneficial information on teaching practices, course

contents, students' success, and institutional programs. Assessment and testing are used for different purposes, namely, mainly for achievement, diagnosis, placement, and proficiency (Fulcher, 2010). Moreover, they can be used to check whether learning outcomes and the expectations of teachers and institutions are met.

There is no doubt that assessment and testing are developing and becoming more and more important every day (Taylor, 2009). Growing numbers of students take different kinds of tests in classes. Teachers find themselves preparing, designing and delivering tests, assessing and evaluating students. Some reasons for the increased number of language tests are shown as the increase in international migration (McNamara, 2008) and the idea that assessment is essential for learning (Black and William, 1998). Besides, as highlighted in the plethora of research, there is a shift from summative assessment to formative and dynamic assessment that is framed within Zone of Proximal Development, and these lay emphasis on the necessity of continuous assessment (Inbar-Lourie, 2012; Leung, 2007; Poehner, 2005). All these responsibilities are placed upon teachers since the beginning of the 21st century with the growing interest (Fulcher, 2012). It is seen that teachers have two roles at the same time: being an assessor and instructor (Inbar-Lourie, 2012). Thus, language assessment literacy gains importance and is critical for education.

There are various definitions of assessment literacy in the literature. In the early literature, it is seen that American Federation of Teachers National Council on Measurement in Education National Education Association (1990) declared standards for teachers about assessment competence, which can be relatable with assessment literacy in a way. According to the document, teachers must be skilled at choosing and developing correct and appropriate assessment methods. They should be competent in applying the methods, scoring the tests, and interpreting these methods to make decisions about syllabus, and learning outcomes. They are also expected to develop a valid and reliable scoring procedure, and deliver these results to students, their parents and other teachers. Finally, they are warned to avoid from inappropriate and unethical ways of assessment.

Assessment literacy term was first used by Stiggins (1991). It is basically knowledge about assessment procedures (Sevimel-Sahin and Subaşı, 2019), including the knowledge of how to

prepare and design assessment activities, of what assessment type to adopt, and of how to interpret the gathered information for varied aims (Inbar-Lourie, 2012; Taylor, 2009). Indeed, there are significant principles such as test specifications, test design, validity, reliability, and marking (Coombe, 2018; Douglas, 2009; Fulcher, 2010). Pill and Harding (2013) defined it as the competency of understanding, evaluating, and preparing language tests when required, and analysing the results of them. On the other hand, O'Loughlin (2013) handles the topic in a critical manner and states that assessment literacy is having a variety of skills regarding test preparation, interpretation of scores, and evaluation of the procedure along with the critical understanding of the role of assessment in the society and education. Similarly, this critical perspective on assessment literacy is found in the definition of Vogt and Tsagari (2014). They claim that it is the ability to evaluate assessment practices critically and score tests in the virtue of theoretical knowledge. It is quite important for teachers to have these qualifications to implement suitable assessment practices. Furthermore, they will be able to review whether the aims are met and to what extent the teaching and learning activities are effective.

2.3.1. Language Assessment Literacy Research Conducted Abroad

A considerable amount of literature has been published on assessment literacy knowledge and training needs of English language instructors working at universities. The main aspects of the reviewed literature are assessment literacy perceptions, training levels, training needs, and assessment practices.

To start with, Hasselgreen et al. (2004) explored the training needs of teachers throughout Europe. The aim of the survey was to find out the areas of needs and accordingly provide training in the language testing and assessment. An online survey was conducted and sent to teachers from different European countries. Teachers were categorised according to their professional roles; language teachers, trainers, and experts who were defined as item writers and test designers. There were 914 participants from 37 European and 50 non-European countries. The collected data were analysed in three main themes: classroom focused assessment, purposes of assessment and content and concepts of assessment. According to the survey results, it was concluded that teachers needed training on preparing tests, making conclusions depending on test results, using peer/self-

assessment, assessing culture related aspects, providing reliability and validity of tests, using statistics, and writing test items. Depending on the results, Hasselgreen et al. (2004) consider common training for teachers and teacher trainers necessary. However, it was suggested by the researchers that specialist training would be better. However, according to the reports revealed by the survey findings, there were three implications. First of all, general training sessions must be designed for both the teacher trainers and the teachers. Secondly, these training sessions could focus on the needs emerging from the findings of this survey; however, it could be better to analyse different contexts and design a training program accordingly. Thirdly, a different separate professional training program could be provided for the experts' needs. Hasselgreen et al. underlined the significance of formal training programs about LTA and implied that during those years it was not catered in the European education system.

Similarly, Fulcher (2012) designed an online survey to find out training needs of language teachers. Via the online survey, 278 responses were gathered around the world including New Zealand, North and South America, Middle and Far East, Australia, and Europe. Teachers were worried about reliability and validity of tests and reported that they needed training for fair scoring. In addition to these, they highlighted the importance of context in classroom assessment practices.

Similar to previous research, Vogt and Tsagari (2014) explored the training levels and also needs of the teachers from seven European countries such as Cyprus, Germany, Poland, Turkey, Greece and Italy. In this mixed method research, 853 teachers responded to the questionnaire and 63 of them were interviewed. The questionnaire comprised three main parts; classroom-focused language testing and assessment, purposes of testing, concepts and contents of language testing and assessment. It was found that teachers' assessment literacy was underdeveloped. In all three parts, teachers found themselves least trained with 42.4 % in the category of purposes of testing. For instance, giving scores, giving final certificates, and placing students into programs were the ones among purposes at which they see themselves not skilled enough. The second least developed literacy was preparing tests. It was made clear by the researchers that teachers were not able to critically evaluate the validity and reliability of the tests. As noted in previous research, the teachers were reported to need to attend both a general training program and a more advanced one.

A small-scale study in Iran revealing similar results by Jannati (2015) investigated the perceptions about assessment literacy levels and instructors' practices regarding knowledge about assessment and testing in Iran. 18 instructors from different language institutions participated in the research, and they were asked semi-structured interview questions. The participants were grouped according to their years of experience. However, no difference was found among groups. Content analysis of the interviews represented that although instructors were found quite knowledgeable in basic testing and assessment concepts and principals, not much was reflected in their practice, proving that they need training in language testing and assessment.

Similar to the previous small-scale research, Hakim (2015) investigated 30 language instructors' assessment awareness working at an English Language Institute of a university. Like in Jannati's (2015) study, the instructors were put into three different groups according to the years of teaching experience. The questionnaire developed by the researcher was analysed qualitatively, and the findings showed that participants were knowledgeable about the principles and valued them. One interesting finding from the research was that experienced instructors valued validity and reliability of tests, integrated assessment, the compliance of assessment with course objectives, self and peer-assessment more than other groups. Nevertheless, the results also revealed their assessment and testing practices were limited.

Unlike the aforementioned small-scales studies, Xu and Brown (2017) explored the assessment levels of 891 English language teachers working at Chinese universities. The focus of the research was to find out the effects of demographic characteristics on their assessment levels. Since the aim was to investigate the teachers' assessment levels, a questionnaire including testing and assessment scenarios was adapted along with questions about demographic information such as the highest level of degree, years of experience, gender, age, and so on. The results showed that the teachers had a basic level of assessment knowledge. Moreover, there was a limited effect of demographic variables on assessment levels of the teachers. The researchers highlighted the need for context dependent training programs since the assessment levels of them were found insufficient.

Giraldo-Aristizabal (2018) explored language assessment literacy beliefs and practices of 60 English language instructors working at a language institute in Colombia. Data were collected through likert type questionnaires, documents, and interviews. Since the researcher's aim was to develop assessment literacy levels of the instructors, the research was designed as an action research study. The data were analysed qualitatively, and findings indicated that the majority of the instructors believed the tests must provide basic principles: reliability, validity, authenticity and positive washback effect on teaching and learning. However, these fundamental principles were not observed in the teachers' practices. Finally, Giraldo-Aristizabal (2018) suggested that a training program explaining grammar and listening test designs, along with positive washback alternatives would be beneficial for the instructors.

In the same vein, Latif (2021) explored the language assessment knowledge at the tertiary level in the Saudi Arabian context. The researcher adopted a Classroom Assessment Literacy Questionnaire to investigate the instructors' assessment knowledge and skills. The data collected from 80 EFL instructors who were working in English Language Preparatory Program indicated that they had limited knowledge of assessment skills and understanding of recent concepts in the assessment. Pointing out the inadequacies in the assessment literacy of EFL instructors at the tertiary level, Latif (2021) put forward that both pre-service and in-service teacher development regarding the theoretical knowledge must be developed. Moreover, the importance of continuous and extensive professional training based on the recent practices and theoretical knowledge was underlined.

2.3.2. Language Assessment Literacy Research Conducted in Turkey

There are several research studies conducted in Turkey regarding assessment literacy perceptions of teachers. A review of literature in Turkey indicates that although teachers have received training or have little assessment knowledge, they do not put their knowledge into practice. The recent evidence suggests that teachers need continuous, in-service training programs and workshops to be up-to-date and improve their assessment literacy (Ballıdağ, 2020; Büyükkarcı, 2016; Mede and Atay, 2017; Ölmezer-Öztürk, 2021; Öz and Atay, 2017, Yastıbaş and Takkaç, 2018).

To start with, one of the early studies was conducted by Köksal (2014), who explored the assessment and testing knowledge of English language teachers working in different schools by collecting the test samples they prepared. 56 English tests were analysed by the researcher and concluded that there were problems such as the fact that there were more than one correct answer, unclear instructions, and uncontextualized questions hindering the meaning. Indeed, the validity and the reliability of the tests were questionable. The results indicated that teachers had insufficient training regarding assessment and testing.

Büyükkarcı (2016) explored the assessment literacy levels of in-service teachers working in the ministry of education and at universities. The researcher also investigated the effects of years of experience and level of degree on the assessment literacy level. Data were collected via an assessment literacy inventory from 30 teachers. Findings showed that both the university instructors and the teachers working in the ministry of education presented a low level of assessment literacy. The years of experience and the highest level of degree did not make any difference regarding the assessment level. Surprisingly, novice teachers presented to have more assessment knowledge than experienced teachers. Büyükkarcı (2016) concluded that there was a need for comprehensive training not only in theory and knowledge but also in principles, skills and practices.

Similarly, Öz and Atay (2017) investigated Turkish EFL instructors' in-class language assessment practices, beliefs and reflections. The relationship between instructors' perception and years of experience was explored. 12 instructors in the preparatory program were interviewed, and data were analysed qualitatively. Interview results were presented according to two categories: instructors' perceptions about classroom assessment and their classroom practices. To start with the perceptions, they were aware of the fact that classroom assessment was significant and necessary, and they knew some of the basic terminologies and the characteristics of classroom assessment. When it comes to their practices, the interview analysis indicated that although they counted validity, reliability, and authenticity as important characteristics of classroom tests, when

they were asked how they achieved these quality issues, they explained their own ways of reliability and validity. It was also revealed in the interviews that in spite of the abundance of classroom assessment tools, they used a limited number of tools. Another important finding was that nearly most of the instructors adopted formative assessment and barely used summative assessment in the classroom. Moreover, while evaluating students' performances during classroom activities, they did not use any scales or rubrics but depended on their incentive judgments. The researchers suggested that there was a need for a scale that must come from administration, indicating that they could not adopt one on their own. To sum up, nearly all the instructors presented basic classroom assessment knowledge; still, they did not put this knowledge into practice, proving that there was an inconsistency in their perceptions and practices. Besides, no significant correlation was found between years of experience and perception. The researchers pointed out the necessity to focus more on the assessment topic in pre-service education and the need for continuous professional development training for in-service teachers with regard to assessment literacy. The researchers put forward that the discrepancy between the perceptions and practices could be because most instructors were novice and newly graduated. They were familiar with the terminology, theoretical concepts, but in time they could put their knowledge into practice.

Mede and Atay (2017) explored the assessment literacy of English language instructors working at non-state universities' English preparatory programs in Turkey. The aim of the research was to find out training levels and training needs of EFL instructors. Data were collected via an online assessment literacy questionnaire and focus group interviews. 350 ELT graduates who had at least 5 years of experience participated in the research. The findings of the questionnaire showed that the knowledge of assessment was found insufficient and that the participants needed training in classroom-focused and content and concepts of LTA. Similarly, the instructors stated they were not competent enough to test productive, receptive and integrated skills. In addition, they were not familiar with the principles of testing such as validity, reliability, as well as statistics to evaluate the overall results. However, the domain in which they found themselves more comfortable was testing micro-linguistic aspects of language; grammar and vocabulary. Depending on the findings, the researchers laid stress on the context-dependent training about all LTA domains, i.e., classroom-focused LTA, purposes of testing and content and concepts of LTA.

Unlike the previous studies, Yastıbaş and Takkaç (2018) examined how Turkish EFL instructors developed their assessment knowledge. 8 instructors were interviewed via think-aloud protocols, and data were analysed qualitatively. The results proved that the instructors were very textbook- and student-dependent while preparing tests. The instructors also gave importance to the positive washback effects of the tests. However, most of them decided the question numbers depending on their beliefs.

Şahin (2019) explored language testing and evaluation course in English language teaching programs in Turkey. The data was collected via semi-structured interviews, language assessment literacy questionnaire and 36 language testing and evaluation course syllabi. Şahin (2019) concluded that only one language testing and assessment course in English language teaching program was not enough to appropriately learn the theoretical knowledge and practical skills. The findings also indicated the priority was given to summative assessment in the language testing and evaluation course in Turkey. Regarding the examined syllabi, it was presented that the focus was given to the theoretical knowledge more than the practical aspects of language testing. Moreover, evaluating the language testing and evaluation course, pre-service teachers were less satisfied about formative and alternative assessment methods.

When it comes to more recent research, there are several theses related to assessment literacy in the Turkish context. Ballıdağ (2020) explored the language assessment literacy levels of teachers working at middle, secondary, and high schools in Turkey. In this mixed method research, data were collected via a assessment literacy questionnaire and interviews. In total, 101 Turkish language teachers participated, and their responses were analysed under three domains, i.e., classroom focused assessment literacy, purposes of testing and content and concepts of LTA. Majority of the quantitative findings showed that teachers received little training about all the domains, and they were found they need further training.

Similarly, Kaya (2020) investigated language assessment literacy of instructors working in English preparatory programs in Turkey. The researcher questioned the relationship between demographic information such as educational background, years of experience, attendance to assessment and testing training and their assessment knowledge. 195 EFL instructors participated in the research and given an assessment knowledge scale, and they were also interviewed. The findings of the research indicated that the EFL instructors had a high level of assessment

knowledge, which is quite different from the reviewed literature. However, no impact on assessment literacy was observed regarding the demographic information.

Ölmezer-Öztürk (2021) examined the effect of a classroom assessment course on pre-service English language teachers' LTA literacy. The pre-service teachers took language assessment course through 14 weeks. During the lessons, the importance of assessment, the role of giving feedback, ready-made language tests, and the importance of planning process in language assessment were discussed. The researcher collected the data from 48 pre-service teachers with reflection reports and focus group interviews. At the very beginning of the course, the pre-service teachers attached negative meanings to assessment. However, through the end of the course, the findings were more positive. They stated that they were more confident in language assessment and became more knowledgeable about the purposes of testing, the role of the teacher in this process, and the techniques of classroom language assessment. Ölmezer-Öztürk (2021) suggested that language teacher educator programs could be enhanced with more language assessment related courses, both optional and compulsory, to be able to improve pre-service language teachers' assessment literacy.

To conclude, there is a plethora of research conducted on the assessment literacy of in-service teachers. The previous research contributed to the literature by discussing teachers' perceptions, training levels, training needs, practices and the improvement of assessment knowledge. Some of the research aimed to investigate the relationship between demographic information of teachers and assessment knowledge (Büyükkarcı, 2016; Hasselgreen, 2004; Jannati, 2015; Kaya, 2020; Öz and Atay; 2017; Xu and Brown, 2017). However, in the reviewed literature, it is seen that little attempt has been made to investigate ELT and non-ELT graduates' perceived training levels and training needs regarding language testing and assessment literacy. Furthermore, although some research conducted on assessment literacy perceptions of the teachers in Turkish context, conducting more research on the topic are suggested by the previous researchers due to the contextual differences (Ballıdağ, 2020; Büyükkarcı, 2016; Mede and Atay, 2017). Therefore, this paper attempts to explore the perceived training levels and training needs of EFL instructors working in an English Preparatory Program of a non-state university in Istanbul and aims to compare ELT and non-ELT graduates' perceptions.

2.4. Syllabus Design

Syllabus is one of the ignored, and confusing topics in language teaching (Krahnke, 1987). The difference between a syllabus and a curriculum is that while the former is more specific, the latter is more comprehensive that it might even include many syllabi. Krahnke (1987) stated that the curriculum might include whole year practices while the syllabus only consists of one specific teaching area, e.g., speaking lesson syllabus. The syllabus identifies the goals and content of that specific lesson. On the other hand, the curriculum indicates what the learners will be able to do at the end of the year. Yassi et. al (2018) made a distinction between curriculum and syllabus by stating the curriculum is a broader concept and includes the ways students learn specific topics, materials, syllabus design in language testing and assessment methods. However, the syllabus is just the content of a subject with the suggestions about methodology. Similar to Krahnke (1987), Yassi et. al (2018) explained that syllabus might be a part of a curriculum.

There are six different types of syllabus, i.e., structural, notional-functional, skill-based, task-based, content-based (Krahnke, 1987) (Figure 3). It was indicated that it was almost impossible to distinguish them from one another. In other words, in a language syllabus, it is most likely to see the syllabus types mixed. As shown in Figure 3, the syllabus types are in a continuous relationship.

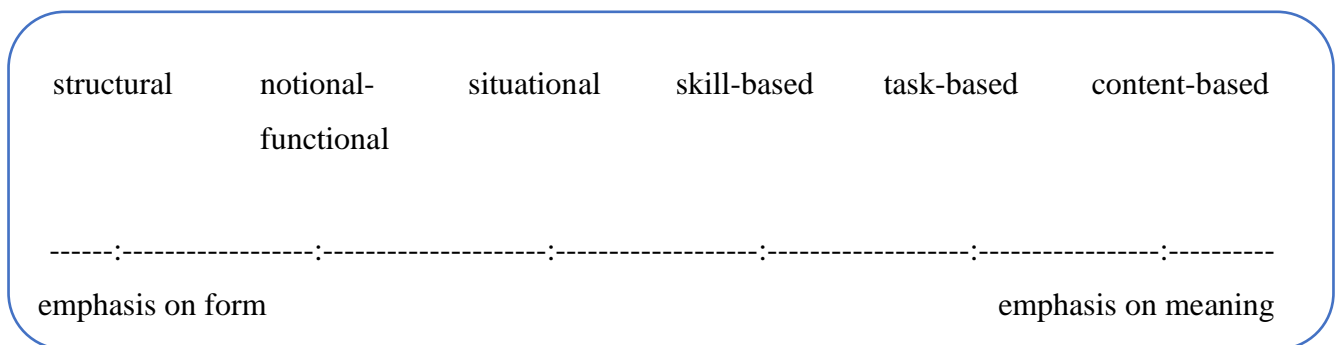


Figure 3. Continuum of syllabi (Krahnke 1987, p.18)

To start with, a structural syllabus, in its other name a formal syllabus is based on forms of language such as grammatical rules, adverbs, adjectives, clauses and also the morphology and pronunciation of the language.

Secondly, a notional-functional syllabus as the name suggests, focuses on the functions of the language. Making a polite request, agreeing, and disagreeing are the examples of this type of syllabus. It is aimed to teach the communicative functions of the language (Thayniath, 2017). It adopts realistic tasks, and these lead the learners to communicate with one another.

Thirdly, a situational syllabus is based on the real and imaginary situations. The learners are supposed to use the language in these specific situations. This type of syllabus indeed highlights the importance of discourse in which the language is spoken in specific situations. For instance, buying a book from the bookstore, asking for the directions in a new city, and going to the airport are the situations where the learners practice the specific language.

A skill-based syllabus mainly depends on improving the skills of the language. To be able to give oral presentations, to write well-developed essays, to listen to the details of a speech, to read texts for the main ideas can be the examples of this type of syllabus.

A task-based syllabus aims to teach the language via the performance of tasks. The main purpose is to perform the tasks rather than to learn the language. Therefore, the tasks have specific aims, e.g., applying for a job, renting a house, completing bureaucratic forms, reading leaflets to find out a proper language course. The language is learned if the task requires it. In other words, the language is the medium of performing the targeted tasks.

Lastly, a content-based syllabus is not actually used for language teaching but for content teaching. The subject is taught by means of the language that the learners want to learn; however, the aim is to master the content and get information on that content e.g., a science lesson in English.

2.4.1. Syllabus Design Studies

There are a few language testing syllabus design studies conducted in the literature. The studies reveal insights into what pre-service teachers expected from language testing courses and what topics in-service teachers needed to focus more to improve their language testing literacy.

To start with, Huang et al. (2019) aimed to improve the English language teachers' assessment literacy via designing a syllabus for Massive Open Online Course (MOOC). To achieve this, Huang et. al (2019) conducted three-phase research. In the first phase, they investigated 468 English language teachers' perceptions of assessment literacy and their opinions about what to include in an assessment course by means of an online open-ended questionnaire. In this questionnaire, the teachers were asked what skills and knowledge they must have to assess students, to what extent they had these and what they anticipated to see in this training course. Upon the results of the first data collection phase, the researchers designed an online questionnaire which consisted of three parts. The first part revealed their demographic information. As for the second part, the aim was to reveal their views about different assessment literacy competencies, e.g., theories and concepts, test construction, rubric design, adoption of information technology in assessment, the administration of assessment activities, analysing the results of the assessment. And in the last part of the questionnaire, the teachers were requested to write other assessment literacy components different from the ones in the second part of the questionnaire. In the third phase of the study, five experts were asked about the syllabus appropriateness to be able to validate the main findings in the previous phases. According to the first phase of the data findings, the teachers had some basic knowledge about assessment: functions of assessment, the ways to design assessment, the concept of validity and reliability. However, they reported they had little knowledge about the purposes of assessment, the analysis of the results, assessment types. Designing various assessment activities was the area where the training was most needed. According to the second phase of the data findings, an inventory was developed including six assessment literacy items. Considering the data driven from these results, an assessment MOOC syllabus was designed consisting of five modules. The name of the course was "How to Assess English Learning". Basic concepts, test specifications, scales, procedures of testing, item analysis,

the assessment of skills, grammar and vocabulary were included in the syllabus. The researchers followed a top-down approach while designing the syllabus. In other words, the syllabus started with broader topics of assessment and continued with specific areas. The syllabus was context-specific and focused the practices in Chinese context. This designed syllabus started to be used in 2018, and it was reported to be systematic and scientific. On the other hand, Huang et. al (2019) put forward that the syllabus could be enriched with the teachers' reflection on their assessment practices.

Different from the previous research, Rohadi (2017) aimed to design a syllabus for the language teaching course and discuss its effectiveness by providing insights into the challenges. To design a syllabus for the course, Rohadi (2017) collected information via a needs analysis survey which was given to both students and the teachers. The purpose was to find out the current problems and detect the expectation from language testing course. After the analysis, a task-based and notional syllabus was designed. It consisted of the theories and the concepts of language testing and included tasks and assignments. In addition, Rohadi (2017) stated the syllabus featured critical thinking where students could solve problems and make a decision when needed in real-world tasks, which also fostered collaborative learning. Moreover, including writing a reflective journal in the syllabus was found effective in terms of reflective thinking, self-discovery, active learning, and critical thinking. The researcher concluded that thanks to the task-based and notional syllabus of language testing course which was based upon problem-solving, multi-tasking, experiencing, collaborative and critical thinking students were able to put the theory into practice, improved their assessment knowledge, and increased their motivation.

Similar to Rohadi (2017), Hatipoğlu (2015) investigated the pre-service teachers' current knowledge about language testing and their expectations from an English Language Testing and Evaluation course. The data were collected from 124 pre-service English language teachers at Middle East Technical University during the years of 2009-2012. To collect the data, need analysis questionnaire and interviews were employed. The needs analysis survey and interviews aimed to reveal students' background information, e.g., age, gender, and education background. Besides, to gather information on students' previous language testing and assessment training level and to understand their views about the current language testing and evaluation course, some open-ended

questions were asked to the pre-service teachers. As for the last phase of data collection, Hatipoğlu (2015) requested them to write down five things that have to be taught in the course. In addition to the questionnaire, focus group interviews with selected pre-service teachers were conducted to let them express their thoughts and suggestions about the course. The findings of the research indicated that pre-service teachers expected to be able to compose language tests and prepare their students for the national exams. The topics emerging from the students' expectations of the language testing and assessment course for the suggested syllabus were the ways to evaluate the learners, the characteristics of the test-takers, assessment-specific topics such as idioms, and test relativization, testing the skills of language, different test types and techniques, the administration of a test, the theoretical knowledge of testing, the interpretation of test results, and alternative test types. Hatipoğlu (2015) put forward the findings of the research could guide the teachers in the same departments and professionals in designing a course syllabus.

To conclude, it is evident in the literature that to design a syllabus, it is significant to get the teachers' and students' opinions. It could be inferred from the syllabus studies that the needs analysis for designing a comprehensive language testing and assessment syllabus could provide insightful knowledge into the areas necessary for training teachers and could shape the contents of the syllabus.

CHAPTER III

METHODOLOGY

3.0. Introduction

This chapter introduces the methodology of the current research. It first starts with the explanation of the research design by presenting the rationale for the adopted mixed-method design. It is followed by the research questions which framed the current research. The setting and participants are then accompanied. Data collection tools and procedures are clarified in detail, and the rationale for employing the language assessment literacy questionnaire and written interviews are explained. Finally, the way the data is analysed and the steps followed during the analysis are indicated.

3.1. Research Design

The current research draws on both quantitative and qualitative approaches to research since mixed- methods research designs provide more in-depth and thorough evaluation on the topic, and it helps the researcher gain a multidimensional and more accurate view (Ivankova et al., 2006). The strength of employing a mixed-methods research design is the fact that qualitative and quantitative data complement each other, enriching the overall understanding of the existing case. In this mixed-methods research, an explanatory design has been adapted to better understand the participants' perceptions. There are different mixed-method research designs; however, this research adopts a sequential explanatory design, which was firstly initiated with the analysis of quantitative data and followed by the analysis of qualitative data (Creswell, 2013; Ivankova et al., 2006). The reason why to start with the quantitative data is that it provides a general understanding about the research problem. Qualitative data, on the other hand, further elaborate on the numerical findings and help the researcher explore the participants' views (Creswell, 2013; Ivankova et al., 2006). It is of great importance to make a decision about the priority of the data, either qualitative or quantitative, or sometimes both (Creswell, 2013; Ivankova et al., 2006). Although both

approaches were addressed, the priority was given to the quantitative data in this research. Since the data collection procedure started with the quantitative data and as the major data was collected in this phase, the priority was given to quantitative approach. Besides, the third research question, which explored whether there were any significant differences between the ELT and NON-ELT graduates, required some statistical calculations because it would be more reliable to answer this question with statistical analyses rather than qualitative. Moreover, the collected qualitative data was smaller than the quantitative, which does not fulfil the scope of the research. The aim of the second phase of the research, the qualitative phase, was to gain a deeper understanding and strengthen the depth of the statistical findings.

3.2. Research Questions

The current research seeks to address the following questions:

1. What are the perceived Language Testing and Assessment (LTA) training levels of English language instructors working at English preparatory programs in Turkey?
2. What are the perceived LTA training needs of English language instructors working at English preparatory programs in Turkey?
3. Is there any significant difference between ELT and non-ELT graduates in terms of their perceived training levels and training needs on LTA domains?

3.3. Setting and Participants

The research takes place in the School of Foreign Languages of non-state universities in Turkey. The School of Foreign Languages generally have an intensive English program for academic studies. In the school, there might be different units helping the management and organisation. For instance, the Assessment and Evaluation Unit (AEU) is responsible for testing activities of the English Preparatory Program. To briefly explain, the AEU prepares the test specifications in accordance with the curriculum outcomes, prepare quizzes, progress achievement, proficiency and general evaluation tests. In addition to exams, the AEU analyse the

results of the tests and tracks the learner achievement so as to give feedback to the instructors and vice principals. In addition to formal testing, informal and dynamic testing activities might be adapted in the schools, such as speaking and writing portfolios, role plays, and group projects. The unit might also organise standardisation workshops for speaking and writing grading before progress achievement tests since both require a subjective assessment.

These programs also have Planning and Materials Development Units that are responsible for the planning of lessons and the materials that will be followed throughout the year. Instructors are mostly provided with these materials and only allowed to use them in their teaching.

The participants of the research were chosen according to two criteria. Since the data collection tools were a questionnaire and a semi structured interview, two strategies were adopted. That is, the participants for the questionnaire were selected with convenience sampling involving respondents who are available and accessible to the researcher (Cohen, Manion and Morrison, 2007; Creswell et al., 2012). To reach the English language instructors, non-state universities which were affiliated with the Council of Higher Education were explored. The researcher asked for the approval of the school of foreign languages directors to gather research data. To communicate with the directors and instructors, the researcher searched for the schools' websites and sent email to those who volunteered to participate in the questionnaire. The foreign languages schools were chosen from different regions of Turkey, i.e., Marmara, Aegean, Mediterranean and Eastern regions to portray the views of the instructors from a variety of cities around the country. However, some universities did not have an updated school website that the email addresses of the instructors or sometimes even the directors were not found. The table below presents the participants' demographic information in detail.

Table 1.

Participants' Demographic Information

		N
Gender	Male	24
	Female	56
Age	22-24	10
	25-30	27

	31-35	22
	36-40	11
	41-45	3
	46-50	3
	51-55	1
	56-60	2
	60+	1
Years of Experience	less than a year	6
	1-3 years	12
	4-7 years	29
	8-11 years	13
	12-15 years	7
	16-19 years	7
	20+ years	6
	60+	1
Major	ELT	40
	ELL	25
	AL	3
	TI	7
	Other	5
The Highest Level of Degree	BA	43
	M.A	35
	PhD	2
Total		80

The research was carried out with 80 EFL instructors working at non-state universities' English Preparatory Programs in Istanbul. Majority of the participants were female instructors (N=56). The participants' ages varied between 22 and 60. The participants aged between 25 and 30 constituted the majority (N=27), which were followed by 31- and 35-year-olds (N=22). Most of them had 4 to 7 years of teaching experience (N=29). It could be inferred from the table that

many of them were experienced teachers and had 4 to 60+ years of experience (N=63). A small quantity of them were experienced less than 3 years (N=18). As for the majority, half of the instructors graduated from an English Language Teaching program. 35 of the rest of the instructors were reported to study English related majors, e.g., English Language and Literature, American Literature, and Translation and Interpretation Studies. On the other hand, the minority of them graduated from other departments not relating to English language or teaching majors (N=5). With respect to the graduated program, it is clearly seen from Table 1 that ELT graduates and non-ELT graduates were equal in terms of number ($N_{ELT}=40$, $N_{NON-ELT}=40$). Regarding the highest level of degree, only 2 of the participants held a PhD, but nearly half of them had an M.A degree (N=35), and over half of them had a B.A degree (N=43).

For the follow-up interviews, 8 instructors were chosen with the convenience sampling (Cohen, Manion and Morrison, 2007; Creswell et al., 2012). The instructors were participated according to their availability and convenience. The follow-up interviews were conducted through Microsoft-Word as in the written format. Conducted as in written format, interviews provided an ease for the analysis and the interviewed instructors as well. When the interviews are conducted face-to-face some may abstain from them and do not want to participate. Via written interviews, they might have felt comfortable while expressing their views. Table 2 indicates the further information about the interviewed instructors' demographic information.

Table 2.
Interviewee's Demographic Information

Interviewees	Gender	Graduated Program	Years of Experience	Duty
Instructor 1	Female	ELT	4-7 years	Assessment and Evaluation Unit Coordinator, Instructor
Instructor 2	Female	ELT	4-7 years	Instructor, Material Development and

				Curriculum Design Unit Member
Instructor 3	Male	Non-ELT	8-11 years	Head of the Program, Instructor, Material Development and Curriculum Design Unit Member
Instructor 4	Male	Non-ELT	4-7 years	Instructor
Instructor 5	Male	ELT	1-3 years	Instructor
Instructor 6	Female	ELT	4-7 years	Instructor
Instructor 7	Male	Non-ELT	16-19 years	Instructor
Instructor 8	Female	ELT	16-19 years	Instructor, An Old Assessment and Evaluation Unit Member

In total, eight instructors were interviewed, five of whom were ELT graduates, and the rest of whom were non-ELT graduates. Instructor 1, who was an ELT-graduate, was working as an Assessment and Evaluation Unit (AEU) coordinator, prepared exams, and she was familiar with all testing practices besides the instruction of lessons. She has 4-7 years of teaching experience. Instructor 2, the other ELT- graduate, was responsible for material development and curriculum design, and checked the test specifications whether they are in line with the curriculum, i.e., whether the tests provide validity or not and she had 4-7 years of teaching experience. Instructor 3 who was a non-ELT graduate, was the head of the department, responsible for the day-to day coordination and implementation of preparatory programs, and 8-11 years of experience. Instructor 4 was a non-ELT graduate, and he had no other responsibility than instructing. He had 4-7 years of teaching experience. Instructor 5 was an ELT graduate instructor and less experienced than the other interviewed instructors. Instructor 6 was a female instructor and was only responsible for teaching the lessons. On the other hand, Instructor 7, was more experienced than the others with 16-19 years of teaching experience and his responsibility was teaching the lessons. As for the last

interviewee, Interviewee 8, she worked as an Assessment and Evaluation Unit member and had hands-on experience. She was quite experienced with 16-19 years of teaching experience.

3.4. Data Collection Tools and Procedures

In order to explore the perceptions on LTA training levels and needs of EFL instructors working at non-state universities' English preparatory programs, both quantitative and qualitative data were collected. Prior to the data collection, research ethics committee approval was received from Çanakkale Onsekiz Mart University, and informed consents were collected from the participants. They were informed that they had the right to withdraw from the research at any time. They were ensured that their demographic information was not used to reveal their identities. To better investigate the participants' perceptions, the researcher gathered demographic information such as gender, age, years of experience, graduated program, and the highest level of degree.

In the first phase of the data collection procedure, to collect the quantitative data, a 3-point likert-type "Teachers' Questionnaire" developed by Vogt and Tsagari (2014) was adapted (See Appendix A). The questionnaire had three main parts: classroom focused LTA, purposes of testing, content and concepts of LTA. Each part included two subsections as perceived training levels and training needs. Vogt and Tsagari (2014) ran internal reliability tests for the questionnaire and found a high reliability for the individual scales indicating an internal consistency (ranging from $\alpha=.80$ to $\alpha=.93$). To investigate the first and second research questions, descriptive statistics regarding perceived training levels and needs were calculated with the Statistical Package for the Social Sciences (SPSS) version 25. The minimum, maximum and mean scores, and standard deviations were computed for the total of six parts.

For the second phase of data collection, written interview questions were prepared related to each part of the questionnaire (See Appendix B) and sent to 8 instructors via email addresses. The collected written interview responses were transferred into an electronic spreadsheet by dividing them on sentence levels. After an initial coding was done, the overall interview was reviewed for a better wording. The themes were already set; that is to say, each question referred to one part of the questionnaire (e.g., classroom focused LTA, purposes of testing and content and concepts of LTA). Therefore, the interviewees' responses were checked according to whether they matched with the target theme, and according to whether they drifted away from the topic or not.

Therefore, the written interviews were analysed through thematic analysis since it is an advantageous method for investigating the perspectives of the participants by pointing out the similarities and differences, as well as helping produce a well-organized report (Nowell, et al., 2017).

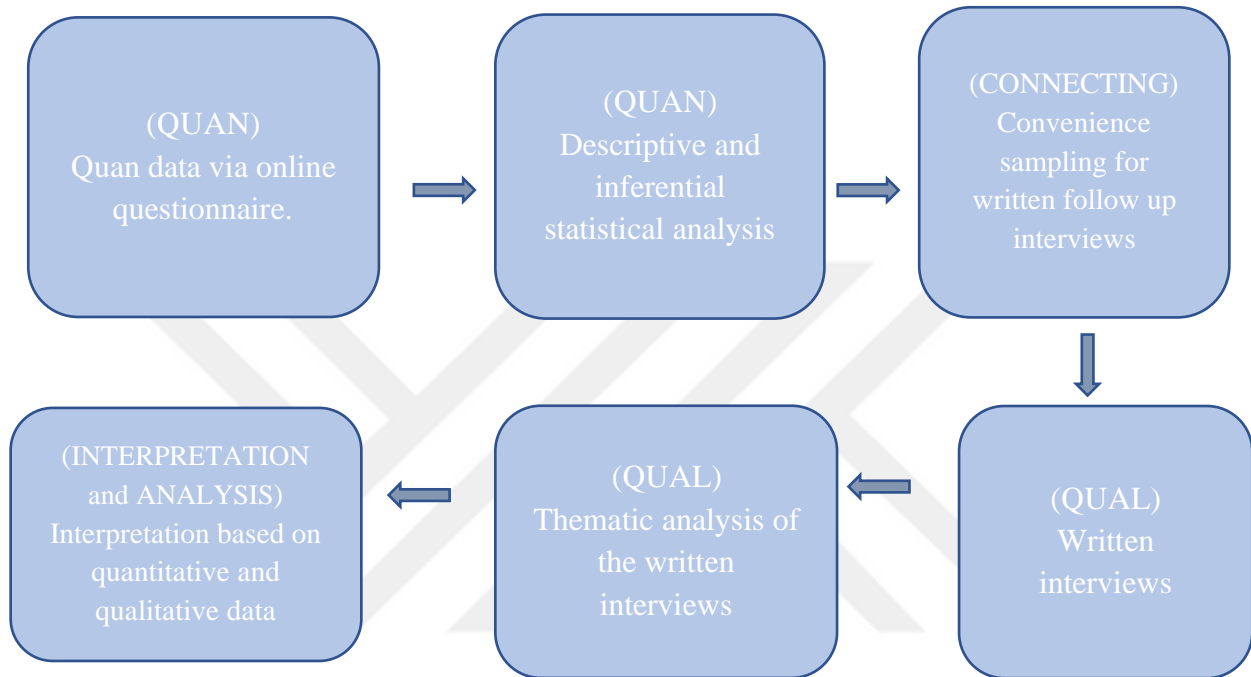


Figure 4. Data Collection Procedures and The Research Design

3.5. Data Analysis

Before running some tests on the SPSS, the distribution of the data was checked to apply the suitable tests for the last research question. Checking the distribution of the data helps the researcher to conduct the appropriate tests: either parametric or non-parametric. By the help of this test, the researcher finds out whether there is normally distributed data or not. Normality is “a symmetrical, bell-shaped curve, which has the greatest frequency of scores in the middle with smaller frequencies towards the extremes” (Pallant, 2016, p. 147). If the result is insignificant ($p > .05$), it means the data is normally distributed (Pallant, 2016).

Table 3.

The Normality Test Results

Questionnaire Parts	Kolmogorov-Smirnov		Shapiro-Wilk
	<i>df</i>	<i>p</i>	<i>p</i>
Classroom Focused LTA (Training Levels)	80	.01	.01
Classroom Focused LTA (Training Needs)	80	.16	.03
Purposes of Testing (Training Levels)	80	.00	.01
Purposes of Testing (Training Needs)	80	.00	.00
Content and Concepts of LTA (Training Levels)	80	.00	.02
Content and Concepts of LTA (Training Needs)	80	.00	.00

In the current research, the normal distribution was not observed, and it was found out that Shapiro-Wilk statistics of Classroom Focused LTA perceived training levels ($p = .01$), purposes of testing perceived training levels ($p = .01$), and content and concepts of LTA perceived training levels ($p = .02$) were significant. Moreover, Kolmogorov-Smirnov statistics of purposes of testing training needs ($p = .00$), and content and concepts of training needs ($p = .00$) were found significant. Since the results of the tests were significant, it was inferred the data were not normally distributed. Due to non-normally distributed findings, the researcher conducted non-parametric tests. With regards to the last research question, Mann Whitney u -test was computed to compare two groups that are ELT and non-ELT graduates. That is, for perceived training levels and training needs under three domains, 6 Mann Whitney U-test in total were computed.

CHAPTER IV

FINDINGS

4.0. Introduction

The findings of the three research questions mentioned previously are presented in this chapter under two categories as quantitative and qualitative findings. Quantitative findings of the three domains of the questionnaire which are classroom-focused LTA, purposes of testing, and content and concepts of LTA are demonstrated with the tables showing statistical values and explained accordingly. In addition, the results of the analysis of qualitative findings are demonstrated via tables and quotations.

4.1. Quantitative Findings

4.1.1. Results of Data Analysis for the First Research Question

The first research question aimed to explore the perceived training levels of EFL instructors working at English Preparatory Programs in Turkey regarding LTA. Descriptive statistics along with mean score and standard deviation were used to present the findings for the first and second research questions.

Table 4.

LTA Training Background of Instructors

		<i>N</i>
Pre-service LTA training	No	26
	Yes	54
Inservice LTA Training	No	16
	Yes	64

($N_{Total} = 80$)

As the table 4 above displays, more than half of the instructors reported that they took Language Assessment and Testing course in their pre-service training ($N=54$). Similarly, a great number of the instructors took part in Language Testing and Assessment training sessions during their in-service years ($N=64$).

Table 5.

Descriptive Statistics of Perceived Training Levels on Classroom-focused Language Testing and Assessment (LTA)

Classroom-focused LTA	Min.	Max.	<i>M</i>	<i>SD</i>
Giving feedback to students based on information from tests/assessment	0	2	1.43	0.71
Using self or peer-assessment	0	2	1.40	0.63
Using ready-made tests from textbook packages	0	2	1.20	0.70
Using informal, continuous, non-test type of assessment	0	2	1.15	0.64
Preparing classroom test	0	2	1.15	0.64
Using the European Language Portfolio, an adaptation of it or some other portfolio	0	2	1.06	0.74

(0=Not at all, 1= A little, 2= More advanced $N=80$)

EFL instructors were asked to choose the best option that described their training levels on LTA. Table 5 represents the findings of the first part of the questionnaire, which is about classroom-focused LTA. According to the findings, in none of the areas, the instructors believed they had more advanced training. “Giving feedback to students based on information from the tests and assessment” ($M=1.43$, $SD=0.71$) had the highest mean value. The instructors believed they had little training. It was followed by “Using self or peer assessment” ($M=1.40$, $SD=0.63$). “Using ready-made tests from textbook packages” ($M=1.20$, $SD=0.70$) was the third item about which the instructors believed they had little training. Similarly, they had little training about “Using informal, continuous, non-test type of assessment” ($M=1.15$, $SD=0.64$) and “preparing classroom test” ($M=1.15$, $SD=0.64$). As shown in the table above, “Using the European Language Portfolio (ELP), an adaptation of it or some other portfolio” had the least mean value ($M=1.06$, $SD=0.74$).

Table 6.

Descriptive Statistics of Perceived Training Levels on Purposes of Testing

Purposes of Testing	Min.	Max.	<i>M</i>	<i>SD</i>
Finding out what needs to be taught/ learned	0	2	1.56	0.61
Giving grades	0	2	1.37	0.66

Placing students onto courses, programs, etc.	0	2	1.01	0.72
Awarding final certificates (from school/program; local, regional or national level)	0	2	0.61	0.70

(0=Not at all, 1= A little, 2= More advanced $N=80$)

Table 6 illustrates the EFL instructors' perceived training levels on purposes of testing. The highest mean value belongs to "Finding out what needs to be taught and learned" ($M=1.56$, $SD=0.61$). It is apparent that the second most trained area is "Giving grades" ($M=1.37$, $SD=0.66$). They have received almost more advanced training about the aforementioned areas. In addition, they stated they had little training about "Placing students onto courses, programs, etc." ($M=1.01$, $SD=0.72$). On the other hand, "Awarding final certificates from school or program; local, regional or national level" was found the least trained area ($M=0.51$, $SD=0.66$). They received almost no training about the last item.

Table 7.

Descriptive Statistics of Perceived Training Levels on Content and Concepts of LTA

Content and Concepts of LTA	Min.	Max.	<i>M</i>	<i>SD</i>
Testing and assessing micro-linguistic aspects (grammar and vocabulary)	0	2	1.46	0.62
Testing and assessing productive skills (speaking and writing)	0	2	1.43	0.63
Testing and assessing receptive skills (Reading and Listening)	0	2	1.39	0.67
Establishing reliability of tests/assessment	0	2	1.28	0.64
Testing and assessing integrated language skills	0	2	1.28	0.66
Establishing validity of tests/assessment	0	2	1.25	0.63
Using statistics to study the quality of tests/assessment	0	2	1.02	0.69
Testing and assessing aspects of culture	0	2	0.91	0.78

(0=Not at all, 1= A little, 2= More advanced $N=80$)

Table 7 is quite revealing in terms of the perceived training levels on content and concepts of LTA of EFL instructors. Overall, the instructors stated they had a little training about the content and concepts of LTA. The highest mean score shows that the most trained area is “Testing and assessing micro-linguistic aspects of the language (grammar and vocabulary)” ($M=1.46, SD=0.62$). This is followed by “Testing and assessing productive skills (speaking and writing)” ($M=1.43, SD=0.63$) and “Testing and assessing receptive skills (reading and listening)” ($M=1.39, SD=0.67$). Similar to “Establishing reliability of tests/assessment” ($M=1.28, SD=0.64$), “Testing and assessing integrated language skills” ($M=1.28, SD=0.66$) was stated as one of the little trained areas. These were followed by “Establishing validity of tests/assessment” ($M=1.25, SD=0.63$). On the other hand, “Using statistics to study the quality of tests/assessment” had a bit lower mean value than the previous items ($M=1.02, SD=0.69$), indicating almost no training. From the table above, it is quite revealing that the least trained area was “Testing and assessing aspects of culture” ($M=0.91, SD=0.78$), revealing that instructors received less than basic training in this area.

Table 8.
Perceived Training Levels in LTA Domains

	<i>M</i>	<i>SD</i>
Content and Concepts of LTA	1.25	0.67
Classroom-focused LTA	1.23	0.68
Purpose of Testing	1.14	0.67

(0=Not at all, 1= A little, 2= More advanced)

As Table 8 presents, for all parts of the questionnaire, instructors stated that they received a little training. “Contents and concepts of LTA” was found the mostly trained domain ($M=1.25, SD=0.67$). The second most trained domain was “Classroom-focused LTA” with a very close mean value to the first domain ($M=1.23, SD=0.68$). However, “Purposes of testing” had the least mean value of all ($M=1.14, SD=0.67$).

4.1.2. Results of Data Analysis for the Second Research Question

The second research question aimed to explore the perceived training needs of EFL instructors working at English Preparatory Programs in Turkey regarding LTA. The tables below provide an in-depth analysis.

Table 9.

Descriptive Statistics of Perceived Training Needs on Classroom-focused Language Testing and Assessment (LTA)

Classroom-focused Language Testing and Assessment	Min.	Max.	<i>M</i>	<i>SD</i>
Using the European Language Portfolio, an adaptation of it or some other portfolio	0	2	1.30	0.70
Using informal, continuous, non-test type of assessment	0	2	1.19	0.75
Using self or peer-assessment	0	2	1.05	0.76
Preparing classroom tests	0	2	1.00	0.75
Giving feedback to students based on information from tests/assessment	0	2	0.90	0.81
Using ready-made tests from textbook packages or from other sources	0	2	0.79	0.74

(0=None, 1=Yes, basic training, 2=More advanced training $N= 80$)

Table 9 illustrates the perceived training needs of EFL instructors on classroom-focused LTA. The instructors stated for most items they needed basic training, and for a few items they needed no training. As Table 9 shows, training is needed the most in “Using the European Language Portfolio, adapting it or some other portfolios” ($M=1.30$, $SD=0.70$) and “Using informal, continuous, non-test type of assessment” ($M=1.31$, $SD=0.63$). These two areas are followed by “Using self or peer-assessment” ($M=1.05$, $SD=0.76$) and “Preparing classroom tests” ($M=1.00$,

$SD=0.75$). Indeed, they claimed they needed basic training for the aforementioned areas. In contrast, no training was needed about “Giving feedback to students based on information from tests and assessment” ($M=0.90$, $SD=0.81$), and “Using ready-made tests from textbook packages or from other sources” ($M=0.79$, $SD= 0.74$).

Table 10.

Descriptive Statistics of Perceived Training Needs on Purposes of Testing

Purposes of Testing	Min.	Max.	<i>M</i>	<i>SD</i>
Placing students onto courses, programs, etc.	0	2	1.06	0.79
Awarding final certificates (from school/program; local, regional, or national level)	0	2	1.04	0.80
Finding out what needs to be taught/ learned	0	2	0.96	0.82
Giving grades	0	2	0.79	0.84

(0=None, 1=Yes, basic training, 2=More advanced training $N= 80$)

In Table 10, perceived training needs on purposes of testing are illustrated. It is seen that EFL instructors needed training the most about “Placing students onto courses and programs” ($M=1,06$, $SD=0.79$), and about “Awarding final certificates from school and program; local, regional or national level” ($M=1.04$, $SD=0.80$). Regarding these two areas, they stated they needed basic training. No training was needed about “Finding out what needs to be taught and learned” ($M=0.96$, $SD= 0.82$), and about “Giving grades” ($M=0.79$, $SD=0.84$) though.

Table 11.

Descriptive Statistics of Perceived Training Needs on Content and Concepts of LTA

Content and Concepts of LTA	Min.	Max.	<i>M</i>	<i>SD</i>
Establishing reliability of tests/assessment	0	2	1.24	0.73
Establishing validity of tests/assessment	0	2	1.24	0.73
Testing and assessing aspects of culture	0	2	1.21	0.72

Using statistics to study the quality of tests/assessment	0	2	1.18	0.76
Testing and assessing integrated language skills	0	2	1.14	0.81
Testing and assessing receptive skills (Reading and Listening)	0	2	1.01	0.79
Testing and assessing micro-linguistic aspects (grammar and vocabulary)	0	2	0.95	0.84
Testing and assessing productive skills (speaking and writing)	0	2	0.95	0.86

(0=None, 1=Yes, basic training, 2=More advanced training $N= 80$)

As seen in Table 11, instructors needed training the most about ‘establishing reliability of tests and assessment’ ($M=1.24$, $SD=0.73$) and “Establishing validity of tests and assessment” ($M=1.24$ $SD=0.73$). The second highest mean value belongs to “Testing and assessing aspects of culture” ($M=1.21$, $SD=0.72$). In addition, instructors stated they needed little training about “Using statistics to study the quality of tests and assessment” ($M=1.18$, $SD=0.76$), “Testing and assessing integrated language skills” ($M=1.14$, $SD=0.81$), and “Testing and assessing receptive skills (Reading and Listening)” ($M=1.01$, $SD=0.79$). However, no training was perceived as needed about “Testing and assessing micro-linguistic aspects (grammar and vocabulary)” ($M=0.95$, $SD=0.84$) and “Testing and assessing productive skills (speaking and writing)” ($M=0.95$, $SD=0.86$).

Table 12.

Perceived Training Needs in LTA Domains

Questionnaire Parts	<i>M</i>	<i>SD</i>
Content and Concepts of LTA	1.11	0.78
Classroom-focused LTA	1.03	0.75
Purposes of Testing	0.96	0.81

(0=None, 1=Yes, basic training, 2=More advanced training)

As illustrated in Table 12, instructors stated that they needed basic training in “Contents and concepts of LTA” ($M=1.11$, $SD=0.78$) and “Classroom focused LTA” ($M=1.03$, $SD=0.75$). However, it appears that they did not need any training about the “Purposes of testing” ($M=0.96$, $SD=0.81$).

4.1.3. Results of Data Analysis for the Third Research Question

The last research question was to compare perceived training levels and training needs of ELT and non-ELT graduates. Since the data were not normally distributed, non-parametric tests were computed. Mann-Whitney U-test, which is an alternative of the t-test, was applied to compare two independent groups for continuous values (Pallant, 2016). Therefore, six Mann Whitney U-tests were conducted to analyse ELT and non-ELT graduates in terms of their perceived training levels and training needs.

To start with, perceived training levels for classroom-focused LTA of ELT graduates ($Mdn=1.50$, $N=40$) and non-ELT graduates ($Mdn=1.08$, $N=40$) significantly differed ($U=463.50$, $p = .00$) with a medium effect size $r = -0.37$. Similarly, perceived training levels for purposes of testing of ELT graduates ($Mdn=1.25$, $N=40$) and non-ELT graduates ($Mdn=1.00$, $N=40$) were significant ($U=511$, $p = .01$) with a medium effect size $r = -0.32$. Lastly, perceived training levels for content and concepts of LTA of ELT ($Mdn=1.50$, $N=40$) and non-ELT graduates ($Mdn=1.06$, $N=40$) significantly differed ($U=438.50$, $p < .001$) with a medium effect size $r = -0.4$.

However, as for the perceived training needs of instructors, no significant difference was found for all domains. Perceived training needs for classroom focused LTA of ELT graduates ($Mdn=1.08$, $N=40$) and non-ELT ($Mdn=1.00$, $N=40$) graduates ($U=743.00$, $p>.05$) was insignificant. Similarly, there was no significant difference between perceived training needs of ELT graduates ($Mdn=1.00$, $N=40$) and non-ELT ($Mdn=0.88$, $N=40$) graduates regarding purposes of testing ($U=739.00$, $p>.05$). When it comes to the last subcategory, there was no significant difference between ELT ($Mdn=1.12$, $N=40$) and non-ELT graduates ($Mdn=1.00$, $N=40$) regarding perceived training needs for content and concepts of LTA ($U=721$, $p>.05$).

4.2. Qualitative Findings

4.2.1. Interview Findings regarding Classroom-focused LTA practices of EFL instructors

The interview results regarding classroom-focused LTA practices of instructors indicated that they adopted a variety of techniques to assess students.

Table 13.

Interview Findings-Classroom Focused LTA Practices

Instructors' Classroom LTA Practices
Ready-made tests for grammar and vocabulary: containing cloze items, open ended items, short-answer items, long-answer items
The (adapted) use of provided instructional materials for assessing Teacher-student conversations
Oral interviews/dialogues
Whole class discussions
Student presentation
Checking student products
Standardized tests
Online interactive tests
Portfolio assessment
Performance evaluation by scales

According to Table 13, instructors' classroom language testing and assessment practices varied. One of the most common techniques was ready-made tests for grammar and vocabulary. A variety of assessment types were used such as cloze, open-ended, short-answer, long-answer items. In addition, they stated they made use of the provided instructional materials, as Instructor 3 said "For the most part, I make use of the instructional materials provided, utilizing them for checking their understanding of the topics studied in class." However, Ins.8 preferred to modify the assessment activities if she had to use the provided instructional material by claiming:

“I rarely used the assessment materials from the books or if I did, I modified them heavily because they weren’t very good), except for the recordings from the listening unit tests because there wasn’t time to prepare my own.”

Most of the interviewed instructors also mentioned the use of formal and informal dialogues, teacher-student conversation to assess their knowledge in the classroom:

I also ask my students to summarize what we have talked and create conversation. (Ins.2)

By talking to the students. If they reply in an advanced way, it means they know it well. If their answer is just yes/no, then I’m not sure about their answer. (Ins.4)

In addition to the formal/informal dialogues, instructors also stated they checked the classroom products, and some said they used scales to evaluate the products:

I check their writing. If they write complex sentences, I guess they have a good command of English. If not, then they may need to improve a bit. (Ins.4)

For speaking section, the students are given a topic that is related to the unit. They record their speech online and send it to the system. I can listen to it and give grades/feedback.

The same applies for writing. They send it and then I check it. (Ins.5)

I use a writing rubric to evaluate students’ writing tasks. (Ins.6)

Ins. 5 also mentioned he used an online platform to check students’ progress and give feedback accordingly:

For receptive skills we are using an online website that is called MyELT which is backed up by National Geographic. The students do online homework that includes reading and listening, I can keep a track of the records such as how much time they have spent, how many questions they got it right etc.

Interviewees were also asked to report whether they were familiar with more recent LTA methods and used them. Responses indicated that all the instructors somehow used self and peer-assessment while instructing the productive skills: speaking and writing. However, it was noted that there were concerns about their effectiveness such as the question of whether they had any meaningful effects on students’ language learning. For instance, Ins. 1 mentioned “I have tried portfolio, self/peer assessment. All were terrible due to the students’ readiness and proficiency levels.”

In addition to Ins.1, Ins.8 stated that the real portfolio trend was not adapted properly in many preparatory schools in Turkey and there was no such time to train the students for how to arrange portfolios and all these caused a failure:

In practice in Turkey, I've found university prep departments refer to certain assignments as "portfolio" work, but there's never an actual portfolio (as in a folder with paper in it or online equivalent) involved. Unfortunately, there's just never time (or time set aside) to focus on and train students how to do effective self-reflection/assessment or feedback.

Likewise, Ins.3 mentioned that peer-assessment was unsuccessful due to the failure anxiety:

Peer-assessment was not effective as learners tended to provide overly-positive attitude towards their classmates when assessing their work, which, I believe was based on their end-focused attitude towards language learning as they were anxious about failing –which would mean that they would have to repeat the whole year; and pay extra money.

On the other hand, Ins. 6 explained although self/peer/portfolio assessment were frequently used in writing classes, she preferred using scales for feedback and scores since she believed it prepared the learners better for standardized tests:

Portfolio assessment and self-or peer-assessment are used in common in writing class due to providing an interim/benchmark process. However, coding for feedback can be more useful to get the students ready for a proficiency exam at the end of the term.

Ins. 7, on the other hand, reported that although he had some information about self/peer/portfolio assessment, he hadn't used it before.

4.2.2. Interview Findings regarding Purposes of Testing

The instructors were asked about the purposes of language testing and assessment, and the findings were categorized under three headings: instructional, student-centred, and administrative purposes.

Table 14.

Interview Findings-Purposes of Testing

Categories	Purposes
-------------------	-----------------

Instructional purposes	To adjust my classroom language or my pace To plan my own teaching To plan the curriculum according to students' needs
Student-centred purposes	To obtain information on students' progress To provide feedback to my students as they progress through the course To make them plan their study plans To motivate my students to learn the language To diagnose the strengths and weaknesses in my students To improve their language To motivate them to make them work harder
Administrative purposes	To provide information to the university

The categorization was adopted from the categories of Cheng and Fox (2017). To start with, many instructors stated they conducted assessment to adjust their classroom language or pace, to plan their teaching activities, and curriculum according to the students' needs.

For instance, the following instructors stated assessing English knowledge of the students was crucial:

Without assessment, it is not possible to plan the instruction and curriculum according to our students' level and needs and know if there is improvement in our students' knowledge or not. (Ins.2)

To check understanding, and to have a general understanding of their level of English so as to structure my next lessons based upon that information. (Ins.3)

To understand how I should adjust my classroom language or my pace and to decide on the difficulty level of the things I teach/use in the classroom (Ins.1)

When it comes to student-centred purposes, they stated they assessed the students to obtain information on student's progress, to provide feedback to see their progress, to make them design

their study plans, to motivate them, to diagnose their strengths and weaknesses, to improve their language proficiency and to motivate them to work harder as the extracts from their accounts show:

While giving feedback, we aim to improve their language in teaching a foreign a language process. (Ins.6)

To see how they're progressing and which areas they need to improve. (Ins.7)

I think the first main reason for doing assessment is to get a snapshot of where students are at and to catch any weaknesses that might not be obvious, so I know what to spend a little more time on. The other main reason is so that students can get feedback on their own strengths and weaknesses. And though I dislike this last reason, students kind of need to be assessed regularly to keep them focused, on-track, and studying regularly. It feels a bit like punishment assessment, but they really do start to lose motivation without regular assessments, and attendance drops, or they don't pay attention or do homework. (Ins.8)

It was also striking in the interviews that instructors conducted assessments because of administrative purposes. Since most of the preparatory programs administer end-of-year exams, proficiency tests, and strictly follow an assessment schedule throughout the year with standardized tests, quizzes, and assignments, the instructors are obliged to assess the learners for administrative purposes, to decide on pass or fail situation as they stated:

Where I work, I don't have any control over the summative assessment administered. (Ins.3)

At my current job, it's because we have to, and everything is prepared and administered on the same day. Personally, I think they're assessed too much. (Ins.8)

4.2.3. Interview Findings regarding Contents and Concepts of LTA

Table 15.

Interview Findings-Assessment of different language skills

Micro-Linguistic Aspects of Language (Grammar-Vocabulary)	Multiple choice questions Completing sentences Matching Ordering
--------------------------------------------------------------	---------------------------------------------------------------------------

	Informal dialogues
	Ready-made tests
Receptive Skills (Listening-Reading)	Cloze items
	Comprehension questions
	Textbook activities
Productive Skills (Speaking-Writing)	Oral interviews
	Teacher student conversations
	Peer activities
	Role play
	Decision making
	Dictation
	Formal/ informal email
	Story writing
	Blog posts writing
	Making comment

As for the contents and concepts of LTA, the instructors mentioned they used a variety of assessment techniques for different language skills. The table above indicates different assessment types that the instructors used according to different skills. To exemplify, in the receptive skills Inst. 2 stated “To assess the receptive skills, we generally use multiple choice questions, matching or ordering, summarizing, and inserting sentences and words activities.”

As for the reading skill, Ins. 4 mentioned he assigned extensive reading and talked about it in the classroom:

For reading, I give them books to read. I mean, not the curriculum books but the short stories, etc. I give them books and assign them time to finish. After they’ve finished reading it, we discuss the book. I ask them different questions about the book. We do brainstorm, or discussion.

Ins. 8 highlighted the importance of using scales to evaluate productive skills and told that she tried to focus on all students’ can-dos to encourage them in the learning process:

Unless the rubric tells me otherwise, when grading speaking and writing, I try to focus on what students can do rather than what they can't do, and I try to use materials/questions that make everyone feel successful, regardless of their level within the group.

Similar to Ins. 8, Ins. 2 mentioned the use of evaluation scales as follows:

While assessing writing skills, since we teach general English at the institution I work, we generally give a situation to our students and ask them to write a formal/informal email about it. We sometimes ask them to write stories, blog posts or comments. We give feedback on the students' products weekly, but they are not graded. We use rubrics while grading the students' spoken or written products.

4.2.4. Interview Findings regarding Self-Efficacy Beliefs of EFL Instructors

The interviewees were lastly asked to report their self-efficacy beliefs related to testing and assessing students' English language knowledge. Most of the respondents reported that they were moderately skilled to assess the students and that they needed some training:

To some extent, not fully competent or knowledgeable. (Ins. 1).

I have a basic understanding of language assessment, but I don't think I am completely capable of assessing language competency in all four skills. There are multiple reasons for that, one of them being my educational background, since I only have a TEFL certification, and my BA is in Literature. (Ins.3)

I have received some kind of education at the university and also, I have a certificate for teaching IELTS too. So, I can say that I have enough skills and knowledge, but there are still many things that I need to learn and develop. (Ins. 5)

However, ELT graduate interviewees were more optimistic about themselves, and they highlighted they were educated for testing and assessing, but they still believed they needed training on testing and evaluation. For instance, Ins. 2 explained her reasons for training needs as follows:

I studied ELT at university and learned how to assess my students' knowledge, so I believe that I am skilled to assess my students' English language knowledge, but I never thought of myself to be 'a testing master'. Creating test items is the most difficult thing for me since you must be sure that the level of a question suitable for the students' level and it matches

to the curriculum. You need to review the questions repeatedly and it takes a lot of time. That is why I have always thought that I need more, intensive training on testing and evaluation.

4.2.5. Interview Findings regarding Suggested Topics for LTA Training

Table 16.

Instructors' Suggestions about LTA Training Topics

Basic Concepts of Assessment
Assessing Skills
Online Assessment
Self-Peer Assessment
Portfolio Assessment
Scale Development
Reliability and Validity in Assessment

To uncover instructors' expectations about a possible LTA training program, they were asked on which topics they would like to receive basic or further training. The common themes, as in Table 16, which emerged in the interviews were basic concepts of assessment, assessing different skills of language effectively in the adult language teaching, online assessment with the recent changes in the mode of language instruction, getting familiar with alternative assessment such as self-peer-portfolio assessment, developing scales to assess productive skills, and providing the reliability and validity in assessment.

For instance, Ins. 3 said "Besides foundational concepts, I would look for practical, workshop-style sessions as a part of such training –including rubric development, item writing, assessing validity and reliability, curriculum alignment of tests, and item analysis." In addition to these topics, Ins, 5 mentioned he would like to learn "What to assess, when to assess, how to cross-check, how to ensure reliability and validity of the tools, and how to assess receptive and productive skills."

Ins. 8 highlighted:

“I’d like to learn more ways of assessing listening, as this seems to be the skill the least is known about, and the assessment methods (listening to recordings and answering questions) seem not have changed in 40 years; I’d also like to know more about using writing/speaking to assess grammar, especially for higher level students.

In addition to assessing skills, the semi-structured interview revealed that the instructors needed to learn more about alternative assessment tools and procedures. For instance, Ins. 6 pointed out “Alternative assessment tools and procedures can be shared in a kind of training program on Language Testing and Assessment.” Similarly, Ins. 8. mentioned “I’d also like to learn ways to do more effective self- and peer feedback/assessment; I’d like to be able to do a proper portfolio with students”.

CHAPTER V

DISCUSSION, CONCLUSION, AND SUGGESTIONS

5.0. Introduction

This chapter starts with the discussion of research findings with references to the literature in Chapter II. It is followed by the conclusion of the research. Finally, pedagogical implications, limitations and suggestions for future research are presented.

5.1. Discussion of the Findings

5.1.1. Discussion of the Findings of the First Research Question

This research study set out to explore the perceived training levels and training needs of EFL instructors who were working at English language preparatory programs of non-state universities in Turkey under three LTA domains: classroom-focused LTA, purposes of testing, content and concepts of LTA. The last question in this research study sought to determine whether there were any significant differences between ELT-graduate and non-ELT graduate-instructors in terms of perceived training levels and needs.

The first research question sought to explore the perceived training levels of EFL instructors. The results obtained from the preliminary analysis indicated that in all the domains of LTA, EFL instructors had basic training, proving that there was a need for extensive training. These results were consistent with the previous research (Ballıdağ, 2020; Büyükkarcı, 2016; Hasselgreen et al., 2004; Mede and Atay, 2017; Öz and Atay, 2017; Vogt and Tsagari, 2014). The mean score for overall training levels for all three domains, i.e., classroom-focused LTA, purposes of testing and content and concepts of testing were computed as 1,21. This mean score refers to ‘little training’ since the highest score is 2, which means ‘more advanced training’. Although the present findings are consistent with previous research findings, there are also little differences in the domains of LTA.

The most trained domain was stated as content and concepts of LTA ($M=1.25$, $SD=0.67$). This also accords with the previous findings in the research of Tamerer (2019) and Ballıdağ (2020). In the current research, the most trained area in the content and concepts of LTA was found testing and assessing micro-linguistic aspects of language, i.e., grammar and vocabulary ($M=1.46$, $SD=0.62$). The result is consistent with the findings of Vogt and Tsagari (2014). The researchers stated that only some aspects of LTA were found to be trained and it was the micro-linguistic aspects of language, while the other aspects of language were found to be underdeveloped, which is also consistent with the current research. Similar to Vogt and Tsagari (2014), Mede and Atay (2017) found that the most trained area where in-service teachers felt comfortable much was testing the micro-linguistic aspects of language. However, the most trained area in the content and concepts of LTA in the research of Tamerer (2019) and Ballıdağ (2020) were found to be the receptive and productive skills of language.

In the current study, the least trained area in the content and concepts of LTA was testing and assessing the aspects of culture ($M=0.91$, $SD=0.78$) and using statistics to study the quality of tests ($M=1.02$, $SD=0.69$). Assessing the aspects of culture had the lowest mean value which means 'no training'. This result aligns with the findings of Tamerer (2019) and Ballıdağ (2020). In their research study, assessing the aspects of culture was found to be the least trained area, however, the mean scores were not as low as in the current study. Although in the current study the instructors stated they received almost no or very little training, Tamerer (2019) and Ballıdağ (2020) found the instructors were little trained in the area. Moreover, Mede and Atay (2017) reported that while the instructors perceived themselves as moderately trained in assessing the aspects of culture, they highlighted they needed basic training.

The second most trained domain was classroom focused LTA ($M=1.23$, $SD=0.68$). The result is in agreement with Tamerer's (2019) and Ballıdağ's (2020) findings which showed that the second most trained area was classroom focused LTA in their research. The area in which the instructor felt most comfortable was found to be 'giving feedback to students based on information from tests/assessment' ($M=1.43$, $SD=0.71$). This is also in accordance with the research study conducted by Ballıdağ (2020). 'using the European Language Portfolio, an adaptation of it or some other portfolio' was found the least trained area of classroom focused LTA ($M=1.06$, $SD=0.74$),

indicating a very little training level. This finding observed in this study mirrored those of the previous studies by Tamerer (2019), Ballıdağ (2020), Vogt and Tsagari (2014), who examined classroom focused training levels of instructors. Especially Vogt and Tsagari (2014) pointed out that alternative and innovative assessment types such as self-peer-portfolio assessment and the European Language Portfolio (ELP) were critical and urgent areas to be mastered. Vogt and Tsagari (2014) added that while the teachers reported they were aware of the terms ‘self-peer-portfolio assessment’, they did not include them during their assessing practices. It is evident in the literature that the ELP promotes self-assessment and provides a multidimensional assessment (Hismanoğlu and Hismanoğlu, 2010; Little, 2005; Little and Perclova, 2001; Schneider and Lenz, 2001; Soriano, 2015; Stoicheva et. al 2009; Ziegler, 2014). Also, students were found to be more confident in their learning process and could take an advantage from their peers. However, there were concerns about the use of the ELP since they did not feel competent enough to adapt it to their assessment procedures. (Stoicheva et. al 2009). Stoicheva et. al (2009) noted that teachers were left without support from the institutions and national authorities, and some found the ELP useless when the school adopted the strategy of pass/fail according to the summative assessment. In addition, in the current research study, the data collected from the interviews indicated the participants needed support to learn more about the peer and portfolio assessment. Moreover, some instructors were hopeless to use the ELP since they believed first students must interiorise the ELP and understand the functions and benefits of it well. Some instructors also mentioned that because students got scores from end-of-year exams, analytic scoring was more useful and that self-peer or portfolio assessment did not reflect the real language competency of students.

Depending on the interview findings, the reason why some instructors did not adopt the ELP or self-peer assessment could be the fact that most preparatory programs followed more or less similar assessment procedures, that they adopted mostly summative assessment where students were evaluated with mid-year, end-of year tests, and that these designated their achievement. Some instructors felt uncomfortable spending time for self-peer or portfolio assessment, putting forward that they lost time with unrealistic purposes and that sometimes even students were unaware of what they were doing and why they were doing it as one of the instructors mentioned in the interviews. He stated that students were overly positive not to upset their friends while giving feedback. One also pointed out that can-do statements did not represent their mistakes

and did not help them at all for the summative assessment where their fate was determined, pass or fail. It is striking from the interviews that most instructors were not knowledgeable and free enough to adopt the ELP or self/peer assessment.

According to the instructors' perceptions, the least trained domain was purposes of testing ($M=1.14$, $SD=0.67$), which corroborates the findings of the studies by Tamerer (2019) and Ballıdağ (2020). Although it was the least trained area, it was obvious that the instructors stated they received little training. The most trained area in the purposes of testing was 'finding out what needs to be taught or learned' ($M=1.56$, $SD=0.61$), indicating that they received almost more advanced training as the mean value was close to 2 (more advanced training). The least trained area was found 'awarding students final certificates from school/program: local, regional or national level' ($M=0.61$, $SD=0.70$), which was again consistent with the findings of the studies by Tamerer (2019) and Ballıdağ (2020). Although the least trained domain of LTA was found the purposes of testing in this research study, Mede and Atay (2017) revealed that purposes of testing was the area where instructors felt most comfortable as they believed they were trained most in this area.

5.1.2. Discussion of the Findings of the Second Research Question

As the second research question, perceived training needs of instructors who worked at English preparatory programs in Turkey were explored under three domains: classroom focused LTA, purposes of testing and content and concepts of testing. When the mean value of all domains were computed regarding their perceived needs, it was found that they needed basic training ($M=1.03$, $SD=0.78$), which seems to be consistent with the interview findings as well. Moreover, the recent literature supports that although the teachers seemed they received some kind of training in LTA during the pre-service and in-service practice, they were still in need of a basic training program to broaden their current knowledge (Ballıdağ, 2020; Büyükkarcı, 2016; Giraldo-Aristizabal, 2018; Hakim, 2015; Hasselgreen et al., 2004; Jannati, 2015; Latif, 2021; Mede and Atay, 2017; Ölmezer-Öztürk, 2021; Öz and Atay, 2017, Vogt and Tsagari, 2014; Xu and Brown, 2017; Yastıbaş and Takkaç, 2018).

The findings of the research indicated that in two of the domains of LTA which were content and concepts of LTA ($M=1.11$, $SD=0.78$), and classroom-focused LTA ($M=1.03$, $SD=0.75$), instructors stated that they needed a little training. The most training was needed in ‘establishing the reliability and validity of tests/assessment’ ($M=1.24$, $SD=0.73$) area in the content and concepts of LTA domain. The results are consistent with the previous research of Mede and Atay (2017), Tamerer (2019) and Ballıdağ (2020). When the results were compared with perceived training levels, it was striking that the least trained area was testing and assessing culture with the lowest mean value. However, as for the perceived training needs, instructors reported they would like to receive the most training in establishing the reliability and validity of testing and then assessing the aspects of culture ($M=1.21$, $SD=0.72$). The reason why they stated so was the fact that reliability and validity are the basic principles that must be achieved in tests (Bachman and Palmer, 1996). This was also found significant in some research studies in the literature. Fulcher (2012) found that teachers were worried about the validity and reliability issues of the tests. Similarly, Jannati (2015) concluded that instructors gave importance to reliability and validity. According to the findings of the research by Giraldo-Aristizabal (2018), instructors considered that tests had to achieve validity and reliability and that they had a positive washback effect on students’ learning. Vogt and Tsagari (2014), on the other hand, highlighted that instructors’ assessment literacy was underdeveloped and that they could not evaluate the validity and reliability of the tests critically.

The least training was needed in ‘testing and assessing micro-linguistic aspects (grammar and vocabulary)’ ($M=0.95$, $SD=0.84$) and ‘testing and assessing productive skills (speaking and writing)’ ($M=0.95$, $SD=0.86$), which accords with the previous research (Ballıdağ, 2020; Mede and Atay, 2017; Tamerer, 2019). Comparing with the perceived training levels, instructors already perceived themselves very competent in testing the micro-linguistic aspects and productive skills of language, which also shows that the collected data about their training levels and needs are consistent in the present study. In the interviews, the instructors also indicated they used a variety of assessment procedures to assess grammar and vocabulary, i.e., multiple choice questions, completing sentences, matching, ordering, informal dialogues, and ready-made tests. Although they used a variety of techniques, they were provided with more alternative ways of assessment. Moreover, while the instructors perceived themselves as the most trained in content and concepts

and classroom-focused LTA, it is striking they still reported they needed the most training in these areas. The reason why they would like to receive training in these domains might be the fact that contents and concepts include very basic and significant foundations for assessment such as assessing different skills of language, providing reliability and validity, and using statistics to evaluate the test scores. Moreover, classroom assessment is so significant that they evaluate the students in a multidimensional way through self-peer assessment, ready-made tests, informal and continuous assessment, and the ELP and that they are also able to give feedback depending on all these techniques.

The second domain where instructors needed training was classroom focused LTA ($M=1.03$, $SD=0.75$). The most training was needed in ‘using the European Language Portfolio, an adaptation of it or some other portfolio’ ($M=1.30$, $SD=0.70$), which supports the previous research (Ballidağ, 2020; Tamerer, 2019; Vogt and Tsagari, 2017) and ‘using informal, continuous, non-test type of assessment’ ($M=1.19$, $SD=0.75$). It is also noticeable in the results that instructors needed almost no training, or very little training, in ‘giving feedback to students based on information from tests/assessment’ ($M=0.90$, $SD=0.81$) and ‘using ready-made tests from textbook packages or from other sources’ ($M=0.79$, $SD=0.74$). Since most English preparatory programs had a separate department for assessment and evaluation, the instructors did not deal with the preparation of assessment related activities. Most of the formal and informal assessment procedures and guidelines were provided by separate units and there were no or little rooms for preparing their own assessment activities. Even if subjective evaluation was required in speaking and writing skills, how to score, what steps are followed while scoring, or what scales are used were explained thoroughly by the assessment and evaluation units, and these would not be left to instructors’ own practices or judgements. Similarly, the materials that would be used during teaching were arranged by material development units. The certificates depending on the results of the students were given by the principal of the school, and again it reveals instructors were not alone in all these testing and assessment procedures. Therefore, it is not surprising to find out the instructors were confident in ‘using ready-made tests from textbook packages or from other sources’ as they were mostly doing these.

As for the last domain which is the purposes of testing ($M=0.96$, $SD=0.81$) the instructors stated they needed no training, but the mean value is very close to 1, which could be interpreted as very little training. It seems they needed less training than basic training. There are similarities between the findings of Ballıdağ (2020) and of this research. The participants in Ballıdağ's (2020) research also stated they needed less training than basic training in the purposes of testing domain. In the current research, the highest mean value belonged to 'placing students onto courses, programs, etc.' ($M=1.06$, $SD=0.79$), indicating the need of basic training. However, they stated they needed less training than basic training in 'giving grades' ($M=0.79$, $SD=0.84$) and 'finding out what needs to be taught/ learned' ($M=0.96$, $SD=0.82$). The reason why they felt quite comfortable in giving grades might be the situation that in most of the preparatory schools, answer keys for receptive skills and micro-linguistic aspects of language, and scoring scales for productive skills were provided beforehand, and the fact that this could made them feel confident and let them think this was a reliable process. Similarly, finding out what needs to be taught and learned was the second area where they needed less than basic training. This could be confirmed with the interview findings. Most interviewed instructors explained why they assessed English knowledge of their students and stated they would like to plan their way of teaching and curriculum according to students' needs, get information about their progress, and accordingly provide feedback. Thus, they constantly checked the students and knew their needs. It also seems possible that they felt confident in this area as they were always provided with the materials that would be used, as they followed the syllabus given by the authorities, and because they were mainly dependent on the strict schedule so that they could find out the needs easily.

5.1.3. Discussion of the Findings of Third Research Question

In order to analyse the difference between ELT and non-ELT graduates' perceived training levels and training needs under three domains of LTA which are classroom-focused LTA, purposes of testing and content and concepts of LTA, six Mann-Whitney U-tests were computed. The results for perceived training levels of ELT ($Mdn=1.50$, $N=40$) and non-ELT instructors ($Mdn=1.08$, $N=40$) were found significant for classroom-focused LTA ($U=463.50$, $p=00$) with a medium effect size $r = -0,37$. In addition, there was a significant difference between ELT ($Mdn=1.25$, $N=40$) and non-ELT graduates' ($Mdn=1.00$, $N=40$) perceived training levels in the purposes of testing

($U=511$, $p = .01$) with a medium effect size $r = -0.32$. Similar to previous domains, perceived training levels of ELT ($Mdn=1.50$, $N=40$) and non-ELT graduates ($Mdn=1.06$, $N=40$) significantly differed ($U=438.50$, $p < .001$) with a medium effect size $r = -0.4$ for the content and concepts of LTA. It seems that ELT graduate instructors received more training compared to non-ELT graduate instructors, and it was evident in their perceptions as well. Since the ELT graduate instructors received a testing and assessment related course during their pre-service years, this finding is not surprising. However, some of the non-ELT graduates received formal training to become English language teachers after graduating from English related departments such as English Language and Literature, American Literature, Translation and Interpretation Studies. Overall, 26 of the participants stated they did not receive any testing and assessment related course in their pre-service trainings, while 54 of them mentioned they received training. Besides, 64 of the participants highlighted they had learnt something about language testing and assessment in theory or practice during their in-service years. Considering these, the training programs they attended during the pre-service and in-service years were not effective or sufficient enough for non-ELT graduates. It was also obvious in the interviews that ELT graduates were comfortable with their language assessment knowledge and practices although they were eager to learn more about LTA.

As for the perceived training needs of ELT and non-ELT graduates, no significant difference was detected for all of the domains. Both the ELT and non-ELT graduates' perceived training needs were similar. This is quite expected since both groups worked as English language instructors, and they shared similar responsibilities. Therefore, their needs might have been found similar.

In the literature, there were some studies exploring the demographic background's effect on assessment literacy. For example, Hakim (2015) and Jannati (2015) explored the effect of the years of experience. While Jannati (2015) found no significant difference, Hakim (2015) found out that experienced instructors gave importance to the reliability and validity more than less-experienced instructors did. Similarly, Xu and Brown (2017) investigated the effect of gender, age, experience and found out these variables had limited effect on assessment literacy, and the researchers put forward the need of context-dependent training sessions. Although there were

studies about some demographic features, no recent research explored the impact of the graduated program. Therefore, no comparisons could be made about this issue.

To conclude, the EFL instructors working in an English preparatory school of a non-state university had an unsatisfactory level of training on LTA literacy, and they needed extensive training about language assessment and testing both in theory and practice. This result corroborates with the literature as well (Ballıdağ, 2020; Büyükkarcı, 2016; Giraldo-Aristizabal, 2018; Hakim, 2015; Hasselgreen et al., 2004; Jannati, 2015; Latif, 2021; Mede and Atay, 2017; Ölmezer-Öztürk, 2021; Öz and Atay, 2017, Vogt and Tsagari, 2014; Xu and Brown, 2017; Yastıbaş and Takkaç, 2018).

5.2. Conclusion

The research attempted to explore perceived training levels and training needs of EFL instructors working in English preparatory programs in Turkey. One of the most significant findings to emerge from the research was that instructors had limited knowledge about LTA. The most trained domain was found to be ‘Content and Concepts of LTA’ ($M=1.25$, $SD=0.67$). The most trained area in the content and concepts of LTA was ‘testing and assessing micro-linguistics aspects’ with the highest mean value ($M=1.46$, $SD=0.62$), and it was followed by testing and assessing productive skills ($M=1.43$, $SD=0.63$). It was also evident in the interviews that instructors were quite confident in assessing grammar and vocabulary. However, the least trained area in the content and concepts of LTA was the testing and assessing aspects of culture with a lower mean value ($M=0.91$, $SD=0.78$). Classroom focused LTA ($M=1.23$, $SD=0.68$) was the second most trained domain. In this domain, the most trained area was ‘Giving feedback to students based on information from tests/assessment’ ($M=1.43$, $SD=0.71$), while the least trained area was ‘Using the European Language Portfolio, an adaptation of it or some other portfolio’ ($M=1.06$, $SD=0.74$). Similarly, in the interviews, the instructors mentioned portfolio assessment was not successful in their classrooms, and some did not favour it since they believed it did not provide the real assessment. Some also highlighted they were not knowledge enough; and they were concerned with its effectiveness. Finally, the least trained domain was the ‘Purposes of Testing’ ($M=1.14$, $SD=0.67$).

The most trained area in this domain was ‘Finding out what needs to be taught and learned ($M=1.56$, $SD=0.61$), while the least trained area was ‘Awarding final certificates (from school/program; local, regional or national level)’ ($M=0.61$, $SD=0.70$). The interviewed instructors mentioned they tested the students for instructional purposes such as planning their teaching and curriculum and adjusting their pace accordingly. However, none of the instructors told they conducted assessment for awarding final certificates, while they were describing the purposes of assessment. All in all, the instructors were found to be little trained, and there was insufficiency in their training.

The second aim of the research was to find out the training needs of EFL instructors. In the ‘Content and concepts of LTA’ ($M=1.11$, $SD=0.78$) and ‘Classroom-focused LTA’ ($M=1.03$, $SD=0.75$), they stated they needed basic training. Although the least trained domain was perceived as the ‘purposes of testing’ in the questionnaire, they stated they needed the most training in the ‘Content and Concepts of LTA’. The reason for this might be because basic concepts such as reliability, validity, assessment of receptive, productive skills and micro-linguistic aspects of language underpin the language testing and assessment, and they give much more importance to this domain. However, the least training was needed in the purposes of testing ($M=0.96$, $SD=0.81$), although they perceived this domain as the least trained domain of LTA. The mean value indicated they needed no training, but as the mean value is very close to 1, it could be interpreted that they almost needed basic training. Not surprisingly, the most training was needed in ‘Establishing reliability and validity of tests/assessment’ ($M=1.24$, $SD=0.73$), aspects of culture ($M=1.21$, $SD=0.72$) in the domain of content and concepts. Considering the instructors’ assessment practices, it was evident they did not construct classroom tests and formal tests, and thus they did not experience and practice these. As also mentioned in the instructors’ perceived training levels, ‘Using the European Language Portfolio, an adaptation of it or some other portfolio’ was less trained, and it was found they needed the most training in this area in the domain of classroom-focused LTA ($M=1.30$, $SD=0.70$). Besides, they highlighted they needed basic training in ‘Using informal, continuous, non-test type of assessment’ ($M=1.19$, $SD=0.75$) and ‘Using self or peer-assessment’ ($M=1.19$, $SD=0.75$). The instructors also mentioned in the interviews that there was inefficacy in self and peer-assessment, which corroborated with the quantitative findings. In

consistent with the perceived training levels, the most training needed in ‘Placing students onto courses, programs, etc.’ ($M=1.06$, $SD=0.79$). and ‘Awarding final certificates (from school/program; local, regional, or national level)’ ($M=1.04$, $SD=0.80$).

On the question of difference between ELT and non-ELT graduates in terms of perceived training levels and training needs, this research found that ELT and non-ELT graduates’ perceived training levels were significantly different in all domains of LTA. However, no significant result was observed in ELT and non-ELT graduates’ training needs, indicating both groups needed the same level of training.

5.3. Pedagogical Implications

The research has a number of implications. One of the issues emerging from the research was that EFL instructors did not consider themselves as sufficiently trained in LTA, and they highlighted they needed further training. Pre-service training might not have been satisfactory. More research in exploring the efficiency of pre-service language testing and assessment courses should be conducted; thus, the weaknesses should be detected so that the course could be enriched with activities where they could practice the theoretical knowledge they have learned during the sessions. Many ELT departments have just one language testing course throughout the program. More courses relating to language testing should be presented to the pre-service teachers so that they would feel competent enough. In addition, hands-on experience could contribute to their teaching and language assessment literacy. It is always more effective to learn by experiencing than just reading the theories, which makes the practice important.

Beside the pre-service training, in-service training courses play a significant role. Instructors should be supported with continuous work-shop style practical in-service training sessions, especially including topics as alternative assessment, validity and reliability, assessment of different language skills and aspects of culture, placing students onto courses and programs, and awarding certificates to learners. Although there were no significant differences in the training levels of ELT and non-ELT graduates, more support could be given to non-ELT graduate

instructors with in-service training courses since significance differences were observed in the perceived training levels. Similar to the investigation of effectiveness of pre-service trainings, the effectiveness of these in-service training programs should be questioned at the end of every education year, and weak areas must be found out and strengthened accordingly.

Moreover, it was evident in the interviews that the EFL instructors did not effectively use self-peer-portfolio assessment and gave more importance to summative assessment. Instructors must be informed that formative assessment is significant for improving students' language achievement. They should equip themselves with these recent and more alternative assessment methods.

This research outlined the training topics for the EFL instructors currently working in English preparatory programs of Foreign Languages Schools and suggested a syllabus for LTA training. The prepared extensive training syllabus of LTA could be used in all preparatory programs in Turkey, and it aims to improve EFL instructors' language assessment literacy. However, further research could be conducted to investigate the strengths and weaknesses of the training syllabus.

5.4. Limitations and Suggestions

A number of important limitations need to be considered. First of all, the sample size was small. Since the survey participation was requested via email, the participation was lower than expected. The sample size could be larger if the data collection could be conducted face-to-face. The setting of the research study was non-state universities in Turkey. Although all the non-state universities were aimed to be included in the research, some of them could not be reached due to the fact that some university websites were not up to date and did not include instructors' email addresses. It could have been better to include state universities from various cities to gain much more understanding of the views of EFL instructors working at English Preparatory Programs in Turkey. Thirdly, since the focus of the research was to shed light on the perceived training levels and the needs of EFL instructors and accordingly suggest a training syllabus, the language testing and assessment practices were of importance to better understand their needs and design a training

syllabus. In other words, classroom practices could provide information whether the perceived language testing and assessment trainings were implemented or to what degree it was implemented during teaching. Thus, in addition to survey and written interviews, conducting classroom observations and making field notes about instructors' language testing and assessment practices could have provided great amount of data to interpret. Moreover, some of the answers given to the written interviews were not sufficient to analyse and interpret. The reason for this could be there was no guiding interviewer as in face-to-face interviews. Last but not least, investigating students' perceptions and views about the instructors' LTA practices could have offered variety to the current situation. However, it should be noted that each context is distinctive, and the training needs of instructors may vary regarding schools, teaching practices, aims, and students.

It is also important to mention that most instructors were not responsible for the design and preparation of tests in this context. They were provided with testing tools, the syllabus, and materials by the separate units in the school. This might have caused them to feel as if they did not need extensive training or advanced knowledge about testing and assessing.

To conclude, although the scope of the research was relatively limited, it contributed to the literature by revealing the perceived training levels and needs of EFL instructors on LTA domains and examined the educational background's effect on perceived training needs and levels. According to the findings of the research, extensive training is of great importance to support reliable and valid testing and assessing practices, and to equip instructors with tools for preparing alternative and more dynamic assessing activities rather than traditional methods.

REFERENCES

- Alderson, J. C. (2000). *Assessing reading*. Ernst Klett Sprachen.
- Álvarez, M. F. (2016). Language testing in the digital era. In E. Martin-Monje, I. Elorza,, and, B. Garcia Riaza (Eds.) *Technology-enhanced language learning for specialized domains. Practical applications and mobility* Routledge: New York
- American Federation of Teachers. (1990). Standards for teacher competence in educational assessment of students. <https://buros.org/standards-teacher-competence-educational-assessment-students>
- Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bachman, L. & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80.
- Ballıdağ, S. (2020) Exploring the language assessment literacy of Turkish in-service EFL teacher. Unpublished Master Thesis, Kocaeli University, Kocaeli.
- Bertolotti, G. & Beseghi, M. (2016). From the learning diary to the ELP: An e-portfolio for autonomous language learning. *Language Learning in Higher Education*, 6(2). doi:10.1515/cercles-2016-0023
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Longman Education.
- Buck, G. (2001). *Assessing listening*. Ernst Klett Sprachen
- Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Eğitim ve Öğretim Dergisi*, 6(3), 333- 346.

- Cheng, L., & Fox, J. (2017). *Assessment in the language classroom: Teachers supporting student learning*. Palgrave.
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. 4th ed. Thousand Oaks, CA: Sage.
- Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education*. Routledge. New York.
- Coombe, C. (2018). *An A to Z of second language assessment: How language teachers understand assessment concepts*. London, UK: British Council.
- Council of Europe, (2006). European Language Portfolio: key reference documents. Strasbourg: Council of Europe. Available at www.coe.int/portfolio, section PROCEDURES FOR VALIDATION.
- Csépes, I. & Fekete, A. (2018b). Assessing speaking skills. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis and S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 98-128). Nicosia, Cyprus.
- Douglas, D. (2009). *Understanding language testing*. Routledge.
- Fulcher, G. (2010). *Practical language testing*. Oxford University Press.
- Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly*, 9(2), 113–132. doi:10.1080/15434303.2011.642041
- Giraldo Aristizabal, F. G. (2018). A Diagnostic study on teachers' beliefs and practices in foreign language assessment. *Íkala: Revista de Lenguaje y Cultura*, 23(1), 25-44. doi:10.17533/udea.ikala.v23n01a04
- Goh, C. C. (2016). Teaching speaking. In *English Language Teaching Today* (pp. 143-159). Springer, Cham.

- Gómez, E. L. (1999). Assessment Portfolios and English Language Learners: Frequently Asked Questions and a Case Study of Brooklyn International High School. LAB, Northeast and Islands Regional Educational Laboratory, Brown University.
- Heaton, J. B. (2011). *Writing English language tests* (3rd ed.). UK: Longman Group.
- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education and Literacy Studies*.
- Hasselgreen, A., Carlsen, C., Helness, H. (2004). European survey of language testing and assessment needs. Report: Part one – general findings.
- Hatipoglu, C. (2015). English language testing and evaluation (ELTE) training in Turkey: expectations and needs of pre-service English language teachers. *ELT Research Journal*. 2015. 111-128.
- Hismanoglu, M., & Hismanoglu, S. (2010). The European language portfolio in ESP classes: a case study of learner reflection and self-assessment. *European Journal of Social Sciences*, 12(4), 671-684.
- Huang, L., Xu, M., Chen, Z., & Liu, F. (2019, March). Syllabus design for teacher education MOOCs (massive open online courses): A mixed methods approach. In *International Conference on Technology in Education* (pp. 156-168). Springer, Singapore.
- Hughes, A. (2003) *Testing for language teachers*. Cambridge University Press
- Inbar-Lourie, O. (2012). *Language Assessment Literacy*. *The Encyclopedia of Applied Linguistics*.doi:10.1002/9781405198431.wbeal0605
- Ivankova, N. V., Creswell, J. W., and Stick, S. L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods*, 18(1), 3-20.doi:10.1177/1525822x05282260
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26-37.

- Kaya, T. (2020). *Exploring language assessment literacy of EFL instructors in language preparatory programs*. Unpublished Master Thesis, Bahçeşehir University, Istanbul.
- Koksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the Net. *Turkish Online Journal of Distance Education (TOJDE)*, 5(1), 1-11.
- Krahnke, K. (1987). *Approaches to syllabus design for foreign language teaching*. Prentice Hall, New Jersey.
- Leung, C. (2007). Dynamic Assessment: Assessment for and as teaching? *Language Assessment Quarterly*, 4(3), 257–278. doi:10.1080/15434300701481127
- Little, D. & R. Perclová (2001). *European Language Portfolio: guide for teachers and teacher trainers*. Strasbourg, France: Council of Europe.
- Little, D. (2005). The Common European Framework of Reference for Languages and the European Language Portfolio: Involving learners and their judgments in the assessment process. *Language Testing* 22(3), 321–336
- Luoma, S. (2004). *Assessing speaking*. Ernst Klett Sprachen.
- Maddalena Taras (2005) Assessment – summative and formative – some theoretical reflections, *British Journal of Educational Studies*, 53:4, 466-478, DOI: 10.1111/j.1467-8527.2005.00307.x
- McNamara, T. (2008). The socio-political and power dimensions of tests. In E. Shohamy and N. Hornberger (Eds.) *Encyclopedia of language and education; Volume 7: Language testing and assessment* (pp. 415–427). New York, NY: Springer.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi-Ankara Universitesi TOMER [Language Journal- Ankara University TOMER]*, 168(1), 43-60.

- Nowell, L. S., Norris, J. M., White, D. E., and Moules, N. J. (2017). Thematic Analysis. *International Journal of Qualitative Methods*, 16(1), 160940691773384. doi:10.1177/1609406917733847
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380. doi:10.1177/0265532213480336
- Ölmezer Öztürk, E. (2021). Developing language assessment literacy of EFL pre-service teachers through classroom assessment course. *International Journal of Contemporary Educational Research*, 8(3), 13-22. DOI: <https://doi.org/10.33200/ijcer.932721>
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Maidenhead: Open University Press.
- Peréz Cavana, M. L. (2012). Autonomy and self-assessment of individual learning styles using the European Language Portfolio (ELP). *Language Learning in Higher Education*, 1(1) .doi:10.1515/cercles-2011-0014
- 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
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233–265. doi:10.1191/1362168805lr166oa
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer.
- Purpura, J. E. (2004). *Assessing grammar* (Vol. 8). Cambridge University Press.
- Purpura, J. E. (2012). Assessment of grammar. *The encyclopedia of applied linguistics*.
- Read, J. (2000). *Assessing vocabulary*. Cambridge university press.

Rohadi, T. (2017). A Breakthrough Syllabus in Teaching Language Testing

Course. *Loquen*, 10(2), 15-30.

Román, S. N., & Soriano, S. A. (2015). Autonomous learning and self-assessment through the

European language portfolio (ELP): A pilot study on primary education. *The Journal of Language Teaching and Learning*, 2, 37–53.

Schneider, G. & Lenz, P. (2001). European Language Portfolio: guide for developers. Strasbourg, France: Council of Europe.

Şahin, S. (2019). *An Analysis of English language testing and evaluation course in English language teacher education programs in Turkey: Developing language assessment literacy of pre-service efl teachers*. Middle East Technical University, Ankara.

Sevimel-Sahin, A., & Subasi, G. (2019). An overview of language assessment literacy research within English language education context. *Kuramsal Eğitim Bilim Dergisi [Journal of Theoretical Educational Science]*, 12(4), 1340-1364.

Stiggins, R. J. (1991). Assessment Literacy. *Phi Delta Kappa* 72(7), 534-539.

Stoicheva, M., Hughes, G. & Speitz (2009, 29 September-1 October). *The European language portfolio: An impact study*. 8th International Seminar on the European Language Portfolio, Graz.

Şentürk, B. (2017). The effect of different self-assessment tools on students' attitudes towards learning English. Unpublished Master Thesis, Hacettepe University, Ankara.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. doi:10.1017/s0267190509090035

Thayniath, S. (2017). An introduction to syllabus design. *Language in India*, 268-277.

Vogt, K. & Tsagari, D. (2014). Assessment Literacy of Foreign Language Teachers:

Findings of a European Study. *Language Assessment Quarterly* 11(4), 374-402. DOI: 10.1080/15434303.2014.960046.

Weigle, S. C. (2002). *Assessing writing*. Ernst Klett Sprachen.

Xu, Y. & Brown, G. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment* 6, 133-158.

Yassi, A. H., (2018). *Syllabus design of English language teaching*. Prenada Media.

Yastibas, A. E., & Takkac, M. (2018). Understanding the development of language assessment literacy. *Bingol University Journal of Social Sciences Institute*, 8(15), 89-106. doi:10.29029/busbed.364195

Ziegler, N. A. (2014). *Fostering Self-Regulated Learning Through the European Language Portfolio: An Embedded Mixed Methods Study*. *The Modern Language Journal*, 98(4), 921-936. doi:10.1111/modl.12147

APPENDICES
APPENDIX A
THE QUESTIONNAIRE

Part I: Demographic Information

Gender

Male

Female

Age

22-24

25-30

31-35

36-40

41-45

46-50

51-55

56-60

60+

How long have you been teaching?

less than a year

1-3 years

4-7 years

8-11 years

12-15 years

16-19 years

20+

What is your nationality?

What subjects/ skills of English language are you teaching?

Listening

Reading

Speaking

Writing

Grammar

Vocabulary

Your responsibilities at school:

Instructor

Unit coordinator

Head of department at school

Principal

Course coordinator

Did you study language testing and assessment course when you were at university?

Yes

No

During your in-service teacher training, have you learned something about testing and assessment (in theory and practice)?

Yes

No

What is your major?

English Language Teaching

English Language and Literature

American Language and Literature

Translation and Interpretation

Other: ---

What is the highest level of degree you have received?

MA

BA

PhD

Which M.A / PhD program did you graduate from?

Part II. Questionnaire

Training in Language Testing and Assessment (LTA)

Part 1. Classroom Focused LTA

Part 1.1. Please specify to what degree you WERE TRAINED in the following statements.

Not at all (0)/ A little (1) / More advanced (2)

- a) Preparing classroom tests
- b) Using ready-made tests from textbook packages or from other sources
- c) Giving feedback to students based on information from tests/assessment
- d) Using self- or peer-assessment
- e) Using informal, continuous, non-test type of assessment
- f) Using the European Language Portfolio, an adaptation of it or some other portfolio

Part 1.2. Please specify to what degree you NEED TRAINING in the following statements.

None (0) / Yes, basic training (1) / Yes, more advance training (2)

- a) Preparing classroom tests
- b) Using ready-made tests from textbook packages or from other sources
- c) Giving feedback to students based on information from tests/assessment
- d) Using self- or peer-assessment
- e) Using informal, continuous, non-test type of assessment
- f) Using the European Language Portfolio, an adaptation of it or some other portfolio

Part 2. Purposes of Testing

Part 2.1 Please specify to what degree you WERE TRAINED in the following statements.

Not at all (0)/ A little (1) / More advanced (2)

- a) Giving grades
- b) Finding out what needs to be taught/ learned
- c) Placing students onto courses, programs, etc.
- d) Awarding final certificates (from school/program; local, regional or national level)

2.2. Please specify to what degree you NEED TRAINING in the following statements.

None (0) / Yes, basic training (1) / Yes, more advance training (2)

- a) Giving grades
- b) Finding out what needs to be taught/ learned
- c) Placing students onto courses, programs, etc.
- d) Awarding final certificates (from school/program; local, regional or national level)

Part 3. Content and Concepts of LTA

3.1. *Please specify to what degree you WERE TRAINED in the following statements.*

Not at all (0)/ A little (1) / More advanced (2)

- 1. Testing/Assessing:
 - a) Receptive skills (reading/listening)
 - b) Productive skills (speaking/writing)
 - c) Micro-linguistic aspects (grammar/vocabulary)
 - d) Integrated language skills
 - e) Aspects of culture
- 2. Establishing reliability of tests/assessment
- 3. Establishing validity of tests/assessment
- 4. Using statistics to study the quality of tests/assessment

3.2. *Please specify to what degree you NEED TRAINING in the following statements.*

None (0) / Yes, basic training (1) / Yes, more advance training (2)

- 1. Testing/Assessing:
 - a) Receptive skills (reading/listening)
 - b) Productive skills (speaking/writing)
 - c) Micro-linguistic aspects (grammar/vocabulary)
 - d) Integrated language skills
 - e) Aspects of culture
- 2. Establishing reliability of tests/assessment
- 3. Establishing validity of tests/assessment
- 4. Using statistics to study the quality of tests/assessment

APPENDIX B
WRITTEN INTERVIEW QUESTIONS

Written Interview Questions
<i>Please write your answers below.</i>
1. How do you assess English language knowledge of your students in the classroom? What kind of instruments do you use? Can you attach any of those instruments please?
2. Why do you assess English knowledge of your students?
3. How can you assess different English language skills? Please give details about your practices.
3.1. Do you know about more recent LTA methods (e.g., portfolio assessment, self-or peer-assessment, dynamic assessment)? Have you ever tried them? How was it?
4. Do you feel you are knowledgeable and skilled enough to assess English language knowledge of your students? Why / Why not?
5. If there was a training about Language Testing and Assessment, what topics would you like to be included?

APPENDIX C
THE SUGGESTED TRAINING SYLLABUS

Module	Main Headings	Sub-headings	Notes
Module 1	Introduction to Assessment - Basic Concepts	1. What is assessment? 2. Teaching and Assessment 3. Assessment Types (Formative-Summative) 4. Assessment and Evaluation 5. Purposes of Assessment (Instructional, Student Centred and Administrative) 6. Qualities of good assessment (Alignment, Validity, Reliability, Fairness, Washback, Practicality) 7. What is the best assessment method? (Via project and discussion) 8. Online Assessment Tools (LMS, Google Classroom, Edmodo, Kahoot, etc.)	<i>Suggested Task 1:</i> Instructors should keep a learning journal reflecting their experiences, newly learned topics and their self-assessment. <i>Suggested Task 2:</i> Teaching-Assessment Reflection Questionnaire ¹ . Instructors will be able to reflect upon their teaching practices and thus considering assessment practices.
	Assessment Procedures	1. Test Specification (Discussion and Pair work)	<i>Suggested Task 3 (Group Work):</i> Instructors choose one skill and level. Write test specifications accordingly.

¹ Questionnaire: What is most important in teaching a language? (Appendix E) (Cheng and Fox, 2017, p. 23)

Module 2	Classroom Focused Language Assessment	<p>1. Classroom Assessment Procedures and Tools (Essay questions, Multiple-choice questions, True or false questions, Oral presentation, Writing Portfolio, Self-assessment, Peer Assessment, Portfolio Assessment)</p> <p>2. Alternative Assessment: Portfolio Assessment</p>	<p><i>Suggested Task 4 (Pair Work):</i> Instructors prepare classroom tests in pairs. Instructors should give feedback for one another's classroom tests and discuss how to improve them.</p>
Module 3	Assessing Micro-linguistic and Cultural Aspects of Language	<p>Why to assess vocabulary? Vocabulary Assessment Methods and Tools</p> <p>Why to assess grammar? Grammar Assessment Methods and Tools</p> <p>Cultural Aspects in Assessment Tasks</p> <p>Conducting Needs Analysis for Intercultural Topics</p> <p>Integration of Multi-Cultural Aspects into Tests</p>	<p><i>Suggested Task 5:</i> Evaluation of assessment materials in terms of cultural aspects.</p> <p><i>Suggested Task 6:</i> Needs Analysis for Intercultural Topics. Instructors can conduct a needs analysis to find out what cultural topics students are into so that they can be aware of those cultural aspects and include them in their assessment practices.</p>
Module 4	Receptive Skills Assessment (Reading and Listening)	<p>Why to Assess Reading? Reading Assessment Methods and Tools</p> <p>Why to Assess Listening? Listening Assessment Methods and Tools</p>	

	Reading Test	Improving a ready-made reading test	<i>Suggested Task 7:</i> Instructors improve a ready-made reading test which already exists and discuss the weak points and strengths of these tests.
	Listening Test	Improving a ready-made listening test	<i>Suggested Task 8:</i> Instructors improve a ready-made listening test which already exists and discuss the weak points and strengths of these tests.
Module 5	Productive Skills Assessment (Speaking and Writing)	Why to assess speaking? Speaking Assessment methods and tools Speaking Scales Validity and Reliability in Speaking Assessment Giving Feedback in Speaking Assessment	
		Why to assess writing? Writing Assessment methods and tools Scoring Procedures for Writing Assessment Writing Scales Validity and Reliability in Writing Assessment Portfolio Assessment	<i>Suggested Task 9:</i> Instructors can evaluate different level writing tasks via different scales. They can compare their final scores and discuss the reasons for discrepancy.

		Giving Feedback in Writing Assessment	
	Speaking/Writing Tasks	Improving ready-made Speaking/Writing Tasks	<i>Suggested Task 10:</i> Instructors improve ready-made speaking and writing tasks which already exist and discuss the weak points and strengths of these tasks.
Module 6	Statistics for assessment	Describing Test Scores Analysing Test Tasks Making Statistical Inferences	<i>Suggested Task 11:</i> Instructors can analyse test scores of a selected skill in one of their classroom and make statistical inferences accordingly.
Module 7	Feedback	Presenting Reflections Feedback about the Training (Suggestions and Improvements)	

APPENDIX D
THE ETHICS COMMITTEE APPROVAL



T.C.
ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ REKTÖRLÜĞÜ
Lisansüstü Eğitim Enstitüsü
Bilimsel Araştırma Etik Kurulu



Sayı : E-84026528-050.01.04-2200018233
Konu : Başvuru İncelenmesi

24.01.2022

Sayın Seda AÇIKPORTALI

Yürütücülüğünüzü yapmış olduğunuz 2022-YÖNP-0017 nolu projeniz ile ilgili Bilimsel Araştırmalar Etik Kurulu'nun almış olduğu 20.01.2022 tarih ve 02/40 sayılı kararı aşağıdadır.

Bilgilerinize rica ederim.

KARAR:40- Seda AÇIKPORTALI'nın sorumlu yürütücülüğünü yaptığı “Investigating Perceived Language Testing and Assessment Literacy Levels of English Language Instructors: A Suggested Training Model (İngilizce Öğretim Görevlilerinin Algılanan Dil Ölçme ve Değerlendirme Okuryazarlık Düzeylerinin Araştırılması: Önerilen Eğitim Modeli)” başlıklı araştırmasının, ilgili **kurumun izninin alınması** ve Bilimsel Araştırmalar Etik Kuruluna sunulması koşulu ile Etik Kurul ilkelerine **uygun** olduğuna oy birliği ile karar verilmiştir.

Prof. Dr. Salih Zeki GENÇ
Kurul Başkanı

APPENDIX E

THE QUESTIONNAIRE: WHAT IS MOST IMPORTANT IN TEACHING A LANGUAGE? (CHENG AND FOX, 2017, P. 23)

Disagree ----- Agree

	0	1	2	3	4	5
1. When I teach a language, my primary focus is on the rules of grammar.						
2. Memorizing vocabulary is essential in learning a new language.						
3. First and foremost, the needs of my students determine what I teach.						
4. Day-to-day interaction with my students guides what we will learn next, and how we will learn it.						
5. Understanding of the literature, culture, and history of a language is the most important reason for teaching a language.						
6. I plan all of my lessons in advance in relation to predetermined goals for learning.						
7. It is important to teach what the syllabus or textbook tells me to teach, to ensure the aims of the course are systematically met.						

8. Before teaching a new course, I start by defining what my students should know and be able to do by the end of the course.						
9. I prefer to negotiate the content we will cover during a course directly with my students.						
10. The quality of learning increases when it is meaningful to a learner.						
11. Students learn the most when they are actively engaged.						
12. It is impossible to predict what an individual student will learn, but I will help each student to learn as much as possible.						
13. My overall plan is to let language learning just happen freely and naturally in my classroom.						
14. How and what I teach depends directly on what seems to work with the greatest number of students.						
15. I always begin a course by identifying my students' needs and interests.						

16. The ongoing process of learning is most important – not the content we cover, the products, or the outcomes.						
17. I'll try any approach that I think will work to support my students' learning.						
18. It is important to correct any and all student mistakes.						
19. Unplanned and spontaneous student interaction promotes the most meaningful language learning.						
20. My main focus is on the long-term goals of my course.						
21. Who I am working with determines what and how I teach a language.						
22. I need to constantly reflect on my teaching with a view to modifying my goals in relation to my students' day-to-day development.						
23. I do not expect all of my students to achieve the same outcomes, because every student is different.						

24. The best and brightest students in my class deserve the most attention.						
25. Who is in my class is not as important as what I am teaching.						
26. I continuously evaluate my students' learning in relation to the learning outcomes defined for my course.						
27. Students only value what is graded and marked.						
28. Students' achievement increases when classrooms are highly competitive.						
29. All language in the class must be monitored and corrected by the teacher, because if students are exposed to errors, they will learn and reproduce them.						
30. Teachers must plan and then follow their plans in teaching a language class, because digressions waste everyone's time.						
31. Incidental or unexpected learning is not as important as predetermined course outcomes.						

<p>32. Spontaneous and free-flowing interaction is essential in learning to use a new language.</p>						
<p>33. I don't know exactly what I'm going to teach until I'm actually teaching it.</p>						
<p>34. Every language task, activity, or experience should be undertaken with a clear purpose or goal in mind.</p>						
<p>35. I define the learning outcomes of my course first, and then I design the tasks and activities that will help my students achieve them.</p>						
<p>36. Students learn the most when something is personally interesting or useful.</p>						