

**LANGUAGE ASSESSMENT TRAINING LEVEL AND PERCEIVED  
TRAINING NEEDS OF ENGLISH LANGUAGE INSTRUCTORS: A MIXED  
METHODS STUDY**



**Mehmet TÜRK**

**JUNE 2018**

**LANGUAGE ASSESSMENT TRAINING LEVEL AND PERCEIVED  
TRAINING NEEDS OF ENGLISH LANGUAGE INSTRUCTORS: A MIXED  
METHODS STUDY**

**A THESIS SUBMITTED TO THE  
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES  
OF  
BAHÇEŞEHİR UNIVERSITY**

**BY**

**Mehmet TÜRK**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF MASTER OF ARTS  
IN THE DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**JUNE 2018**

Approval of the Graduate School of Educational Sciences



Assist. Prof. Enisa Mede

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Art.



Assist. Prof. Hatime Çiftçi

Coordinator

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Arts.



Assist. Prof. Hatime Çiftçi

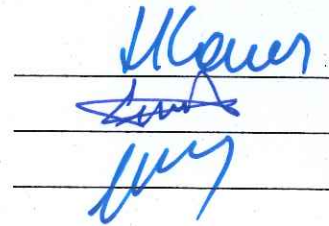
Supervisor

**Examining Committee Members**

Assist. Prof. Hatime ÇİFTÇİ (BAU, ELT)

Assist. Prof. Gürsu AŞIK (BAU, ECE)

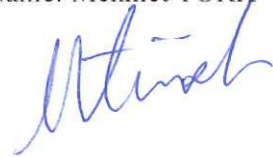
Dr. Hatice ÇELEBİ (ACU, FLD)



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Mehmet TÜRK

Signature :



## **ABSTRACT**

### **LANGUAGE ASSESSMENT TRAINING LEVEL AND PERCEIVED TRAINING NEEDS OF ENGLISH LANGUAGE INSTRUCTORS: A MIXED METHODS STUDY**

Türk, Mehmet

Master's Thesis, Master's Program in English Language Education

Supervisor: Assist. Prof. Hatime Çiftçi

June 2018, 89 pages

The purpose of this study is to investigate the reported training level and perceived training needs of English language instructors in various areas of language testing and assessment (LTA). More specifically, this study examines if the training levels and training needs of instructors demonstrate a significant difference regarding their major, highest degree in education, LTA responsibilities in testing units, and teaching experience. Adopting a mixed methods research design, the study used a questionnaire and a semi-structured interview to collect data from 246 English language instructors, who worked at English Preparatory Programmes at Turkish universities. The findings of the study suggest that although the participants report a good level of training in various areas of LTA, they still perceive a strong need for training in these areas especially in formative types of assessment and in content and concepts of LTA. Moreover, the participants who studied ELT and those who did not significantly differ regarding their perceived training needs in the areas of LTA. While some differences are observed in the reported training level of the participants in relation to their teaching experience, instructors with more teaching experience perceive less need for professional development training.

**Keywords:** Language Testing and Assessment, Assessment Literacy, Professional Development, English Language Teaching

## ÖZ

# İNGİLİZCE ÖĞRETİM GÖREVLİLERİNİN DİL ÖLÇME VE DEĞERLENDİRME EĞİTİM DÜZEYİ VE EĞİTİM İHTİYAÇLARI ALGILARI: KARMA YÖNTEMLİ BİR ÇALIŞMA

Türk, Mehmet

Yüksek Lisans, İngiliz Dili Eğitimi Yüksek Lisans Programı

Tez Yöneticisi: Dr. Öğr. Üyesi Hatime Çiftçi

Haziran 2018, 89 sayfa

Bu çalışmanın amacı, İngilizce öğretim görevlilerinin dil ölçme ve değerlendirme alanında aldıkları eğitim düzeyi ve hizmet içi eğitim ihtiyaçlarını araştırmaktır. Çalışma özellikle katılımcıların bu alanda algıladıkları eğitim düzeyi ve hizmet içi ihtiyaçlarının; mezun oldukları bölümleri, eğitim seviyeleri, ölçme ve değerlendirme birimlerindeki görevleri ve öğretim deneyimleri bakımından önemli bir farklılık gösterip göstermediğini incelemektedir. Karma yöntemli bir araştırma tasarımı benimseyen çalışmada bir anket ve yarı yapılandırılmış görüşmeler veri toplama aracı olarak kullanılmıştır. Üniversitelerin İngilizce Hazırlık Programlarında çalışan 246 İngilizce öğretim görevlisinden veri toplanmıştır. Bulgular, katılımcıların dil ölçme ve değerlendirme alanında iyi düzeyde eğitim aldığını düşündüklerini göstermesine rağmen, özellikle süreç değerlendirme ve dil ölçme değerlendirmenin içerik ve olgu alanlarında hala eğitime ihtiyaç duyduklarını önermektedir. Ayrıca, lisans eğitiminde İngiliz Dili Eğitimi alan ve almayan öğretim görevlileri arasında dil ölçme değerlendirme alanında aldıkları eğitim hakkındaki algıları bakımından istatistiksel önemde farklılıklar gözlemlenmiştir. Katılımcıların dil ölçme ve değerlendirme alanında aldıkları eğitim düzeyinin, öğretim deneyimi bakımından farklılıklar gösterdiğinin gözlemlenmesinin yanında daha fazla deneyime sahip öğretim görevlilerinin daha az hizmet içi eğitim ihtiyacı algıladıkları görülmüştür.

Anahtar kelimeler: Dil Ölçme ve Değerlendirmesi, Değerlendirme Okur-yazarlığı, Mesleki Gelişim, İngiliz Dili Eğitimi



To My Grandfather

## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to those who participated in this study. I would like to thank them for their time for the data collection process.

I would like to thank my supervisor, Assist. Prof. Hatime ifti, for her constant guidance and support during the writing and development of this study. Her contributions made it possible for me to complete this thesis. In addition, I would like to thank my thesis committee members; Assist. Prof. Grsu Aık and Dr. Hatice elebi, for their constructive feedback.

I especially would like to thank my high school English teacher, Mrs. Gzin Keskin, who taught me English and helped me become a better person.

I owe special thanks to my dear wife and friends who have encouraged and helped me with their positive comments and never-ending support.

Last but not least, I would like to express my deepest gratitude to my family who have always been there for me.



## TABLE OF CONTENTS

ETHICAL CONDUCT.....	iii
ABSTRACT.....	iv
ÖZ.....	v
DEDICATION.....	vi
ACKNOWLEDGEMENTS.....	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES.....	xi
Chapter 1: Introduction.....	1
1.1 Overview.....	1
1.2 Theoretical Framework.....	3
1.3 Statement of the Problem.....	7
1.4 Purpose of the Study.....	8
1.5 Research Questions.....	8
1.6 Significance of the Study.....	9
1.7 Definitions.....	10
Chapter 2: Literature Review.....	11
2.1 Overview.....	11
2.2 Types of Assessment.....	12
2.2.1 Formal & informal assessment.....	12
2.2.2 Formative & summative assessment.....	13
2.2.3 Standardized assessment & classroom assessment.....	13
2.2.4 Norm-referenced & criterion-referenced tests.....	14
2.2.5 Placement & proficiency tests.....	14
2.2.6 Traditional & alternative assessment.....	14
2.3 Purposes of Assessment.....	15
2.3.1 Assessment for learning.....	15
2.3.2 Assessment as learning.....	15
2.3.3 Assessment of learning.....	15
2.4 The Role / Importance of Assessment.....	16
2.5 Research on Assessment Literacy.....	16
2.5.1 Research on teachers' assessment literacy.....	16
2.5.2 Research on language teachers' assessment literacy.....	19

2.5.3 Research on language assessment courses .....	25
2.5.4 Research on language teachers' assessment literacy in Turkey .....	28
Chapter 3: Methodology.....	30
3.1 Overview.....	30
3.2 Philosophical Paradigm.....	30
3.3 Research Design.....	31
3.4 Setting and Participants.....	32
3.4.1 Setting.....	32
3.4.2 Participants.....	33
3.5 Procedures.....	35
3.5.1 Sampling.....	35
3.5.2 Data Collection Instruments.....	36
3.5.2.1 The questionnaire.....	36
3.5.2.2 The semi-structured interview.....	38
3.5.3 Data Collection Procedures.....	38
3.5.3.1 The questionnaire.....	38
3.5.3.2 The semi-structured interview.....	39
3.5.4 Data Analysis Procedures.....	39
3.5.5 Reliability.....	41
3.5.6 Trustworthiness.....	41
3.6 Limitations.....	42
Chapter 4: Results.....	43
4.1 Overview.....	43
4.2 RQ 1: Reported Training Level and Perceived Training Needs in LT.....	43
4.3 RQ 2: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their Major?.....	45
4.4 RQ 3: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their Highest Educational Degree?.....	50
4.5 RQ 4: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their LTA Responsibilities in Testing Units?.....	53

4.6 RQ 5: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their Teaching Experience?.....	56
4.7 RQ 6: What are English Language Instructors' Perception of LTA Training in Pre-service and In-service Education?.....	60
Chapter 5: Discussion and Conclusions.....	66
5.1 Overview.....	66
5.2 Discussion of Results for the First Research Question.....	66
5.3 Discussion of Results for the Second Research Question .....	68
5.4 Discussion of Results for the Third Research Question.....	68
5.5 Discussion of Results for the Fourth Research Question.....	69
5.6 Discussion of Results for the Fifth Research Question.....	70
5.7 Discussion of Results for the Sixth Research Question.....	72
5.8 Pedagogical Implications.....	73
5.9 Conclusions.....	74
5.10 Recommendations for Future Research.....	74
REFERENCES.....	76
APPENDICES	
A. The Questionnaire.....	82
B. Consent Form & Interview Questions.....	86
C. Curriculum Vitae.....	88

## LIST OF TABLES

### TABLES

Table 1 Common Types of Assessment.....	12
Table 2 Overview of Participants .....	34
Table 3 Cronbach's Alpha results for the Subsections of the questionnaire.....	41
Table 4 Previous Training in LTA.....	43
Table 5 Mean Scores of Reported Training Level and Perceived Training Needs in LTA.....	44
Table 6 Mean Scores of Average Reported Training Level and Perceived Training Needs in LTA.....	45
Table 7 Mean and Standard Deviations of Training Received and Needed by English Language Instructors Who Studied ELT and Did Not Study ELT (non-ELT).....	46
Table 8 Skewness, Kurtosis, and Z-score Values of Distribution.....	47
Table 9 Independent Samples T-test Results for the Difference in Perceived Training Needs of English Language Instructors Who Studied ELT and Those Who Did not Study ELT.....	48
Table 10 The Results of Mann Whitney U Test with Respect to Training Received in Classroom-focused LTA.....	49
Table 11 The Results of Mann Whitney U Test with Respect to Training Received in Purposes of Testing.....	49
Table 12 The Results of Mann Whitney U Test with Respect to Training Received in Content and Concepts of LTA.....	50
Table 13 Mean and Standard Deviations of Training Received and Needed by English Language Instructors Who Hold a BA Degree and Who Hold an MA or PhD Degree.....	51
Table 14 Independent Samples T-test Results for the Difference between Training Needed by English Language Instructors Who Hold a BA Degree and Who Hold an MA or PhD Degree.....	52
Table 15 Mean and Standard Deviations of Training Received and Needed by Language Testers and Regular Instructors.....	54

Table 16 Independent Samples T-test Results for the Difference between Training Needed by Language Testers and Regular Instructors.....	55
Table 17 Means and Standard Deviations of Training Received and Needed by the Three Subgroups Regarding Teaching Experience.....	57
Table 18 One-Way Analysis of Variance of Perceived Training Need for Classroom-focused LTA regarding Teaching Experience.....	58
Table 19 One-Way Analysis of Variance of Perceived Training Need for Purposes of Testing regarding Teaching Experience.....	58
Table 20 The Results of Kruskal-Wallis Test with Respect to Training Received in Purposes of Testing .....	59
Table 21 The Results of Kruskal-Wallis Test with Respect to Training Received in Content and Concepts of LTA.....	60

## **Chapter 1**

### **Introduction**

This chapter starts with an overview of the place of testing and assessment in English language teaching and learning. Then, the importance of testing and assessment and thus the assessment literacy of teachers are discussed. The chapter continues with the theoretical framework followed by statement of the problem, purpose of the study, research questions, and significance of the study. At the end of the chapter, definitions for some key terms used in this study are explained.

#### **1.1 Overview**

English is a widely spoken language in today's world. It is not only used to communicate with those whose mother tongue is English, but is also and more often used among those who have different first languages. It is the predominant language for business, trade, education, and more specifically in the Internet. As its popularity increases, more and more people learn English. Nowadays, many job seekers face the question if they speak English or not in job interviews. If yes, their level of proficiency is also inquired. While many students are required to take a proficiency exam when they want to study at a university where English is the medium of instruction, others see such exams as a gateway to continue their education abroad (Sheehan & Munro, n. d.). Now that learning English has become so valuable, deciding and determining whether a person can speak English has gained importance too. This has increased the eminence of testing and assessment in English Language Teaching (ELT) as well.

Assessment is not only essential for the examples mentioned above, but more importantly, is a *sine qua non* in the process of teaching and learning. In today's world, assessment plays a crucial role in education (Popham, 2006). Teachers assess students to collect information about their knowledge and abilities in order to reach decisions for further steps in learning and teaching (Cizek, 1997). Stiggins (2007) states that "historically, a major role of assessment has been to detect and highlight differences in student learning in order to rank students according to their achievement" (p. 22). Today, we are all well aware that it is much more than that.

Malone (2011) points out that teaching and assessment are two factors connected to and supportive of each other. Assessment aids teachers to see what students have learnt and in which areas they need improvement. Teachers can make judgements about their own teaching or modify their course programmes via assessment. Besides, assessment is essential to place students at a level in a programme, decide whether a student passes or fails, or measure a student's proficiency at a course (Coombe, Folse, & Hubley, 2007).

Given all the functions of assessment and how vital consequences it may have on learners' lives, it is a must for educators to be trained and be literate in assessment. Assessment literacy refers to "the knowledge of how to assess what students know and can do, interpret the results from these assessments, and apply these results to improve student learning and program effectiveness" (Webb, 2002, p. 1). If a teacher is bereft of assessment literacy, the progress students make may be measured incorrectly, or the decisions and judgements made by teachers, students and even parents may be misinformed, which may have adverse effects on learning and teaching (Stiggins, 2002).

However, many teachers have stayed away from assessment related issues due to lack of their assessment literacy in spite of the great role assessment plays in education. This is mainly caused by the fact that pre-service teachers do not receive enough training in testing and assessment. Taylor (2009) points out that courses on assessment theory and practice are given a small amount of time or attention at graduate level language education programmes. Hatipoğlu (2010) also notes that English language teacher candidates at Turkish universities take only one must course on English language testing and assessment during their training. According to Stiggins (2002), the situation is not very different in the USA. Only about ten of the states in the USA require their teachers to have competence in assessment so as to be rewarded a degree as a teacher. For this reason, many of the teacher training programmes mostly ignore competence in assessment, causing teachers to lack essential skills to obtain assessment literacy (Stiggins, 2002, p. 761). Moreover, English language instructors employed at English Preparatory Programmes do not necessarily have to acquire an ELT degree. Graduates of English Literature, American Literature, Linguistics, or Translation and Interpreting departments are also allowed to be hired at English Preparatory Programmes according to Higher

Education Council (YÖK) regulations. In other words, some English language instructors have not at all received any pre-service education in LTA. Recent research studies also suggest that despite being familiar with assessment related terms, many in-service teachers lack assessment literacy (Büyükkaracı, 2016; Sheehan & Munro, n. d.).

Considering the importance of assessment in learning and teaching, and the current dispute many teachers face regarding assessment literacy has led the researcher to carry out a research study on examining the level of training English language instructors report in various areas of language testing and assessment (LTA) and their needs for in-service training in these areas. More specifically, this study look into the differences in English language instructors' reported LTA training and perceived training needs regarding their major, highest education degree, LTA responsibilities in testing units and teaching experience. Teachers who are competent in assessment literacy, without a doubt, will make better decisions to enhance student achievement.

## **1.2 Theoretical Framework**

Despite being relatively new, assessment literacy is now an extensively accepted term in teacher education (Pill & Harding, 2013). The term is first coined by Stiggins (1991) simply as “having a basic understanding of high- and low-quality assessment and being able to apply knowledge to various measures of student achievement” (p. 535). In a later article, he adds that educators who have acquired assessment literacy may differentiate between “sound and unsound assessments” (Stiggins, 1995, p. 240). Also, Webb (2002) defines assessment literacy as “the knowledge of how to assess what students know and can do, interpret the results from these assessments, and apply these results to improve student learning and program effectiveness” (p. 1). Those who have assessment literacy are aware of what, why, and how they are assessing; of any probable problems and their solutions (Stiggins, 1995). Assessment literate teachers are acquainted with necessary measurement basics to understand what occurs in a classroom (Popham, 2009).

Inbar-Lourie (2008) states that the presence of language assessment courses in language education programmes is evidence that one will need further competencies



to be language assessment literate. Therefore, literacy in language assessment includes not only assessment literacy skills, but also expertise in language related areas, which may be referred to as *language assessment literacy* (Inbar-Lourie, 2008).

Davies (2008) identifies language assessment literacy as a term consisting of three factors: skills, knowledge, and principles. While skills refer to acquiring adequate methods such as item writing and test analysis, knowledge involves information pertaining to measurement and language, and principles consist of concepts such as ethics, fairness, validity, and reliability (Davies, 2008).

Moreover, adding on to the description by Davies (2008), the concept of language assessment literacy could be broadened to involve not only skills, knowledge and principles in accordance with assessment, but also an understanding of assessment practices by stakeholders ranging from educational advisors, government officials, policy planners, decision makers, to the media and general public depending on their needs and context (Taylor, 2009).

Fulcher (2012) mentions a detailed definition of language assessment literacy following a study that investigated the assessment training needs of language teachers:

The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (p. 125)

Fulcher's (2012) definition is a very broad yet comprehensive and inclusive definition. Fulcher first emphasizes that an assessment literate teacher should possess a knowledge base pertaining to assessment principles, then should be skilful and be able to design and evaluate tests keeping in mind ethical issues. He or she should also be aware of the background and culture of the society where the assessment procedures are implemented.

In 1990, *The Standards for Teacher Competence in Educational Assessment of Students* was jointly published by American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), and National Education Association (NEA). This work became a fundamental and influential knowledge base and an instructional framework for the establishment of assessment literacy. Assessment is defined here as “the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy” (AFT, NCME, & NEA, 1990). In accordance with the definition of assessment, 7 standards were set to enhance teachers’ competence in student assessment. These standards are as follows:

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3. The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
5. Teachers should be skilled in developing valid pupil grading procedures, which use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (AFT et al. 1990).

The standards above portray the skills teachers should acquire so as to become competent in assessment; in other words, assessment literate. They also guide teachers with the steps to follow before, during and after assessment procedures. The standards emphasize the significance of selecting, developing, and implementing assessment that is adequate for instructional purposes. Teachers are expected to have the ability to utilize assessment results so as to make decisions regarding students,

instruction, curriculum, and school improvement considering validity issues. An assessment literate teacher should also be adept at discussing assessment results with students, parents and other stakeholders. Lastly, the standards affirm the importance of being aware of immoral and unlawful assessment practices.

Stiggins (1995) suggests that educators should follow certain principles so as to develop literacy in assessment. With these principles it is pointed out that educators should have a clear purpose for assessment, focus on achievement targets, choose appropriate assessment methods, sample student achievement properly, and avoid bias or distortion.

Assessment literacy is not only being aware of assessment tools and making judgements using them, but also utilising those instruments in order to develop the progress of learners, improve instruction, and maximize the efficiency of courses or programmes which provide learners with more opportunities to succeed. Educators with assessment literacy know that assessment may be motivating as well as demotivating for all those involved in learning environment – not only students and teachers, but also parents, principals, and administrators (Webb, 2002).

Brookhart (2011), on the other hand, argues that the standards proposed by AFT et al. (1990) had been valuable input to the field of assessment since they were published in 1990. Though, they need, she contends, some update due to (1) lack of current conceptions of formative assessment knowledge and skills and (2) lack of issues pertinent to standards-based assessment, and offers an updated list of educational assessment knowledge and skills for teachers to become competent in assessment. In this list, Brookhart (2011) explains in detail what skills teachers should have to be competent in assessment literacy. She points out that teachers should be familiar with different assessment tools and methods, use them effectively to interpret the results and make reasonable educational decisions, and give effective feedback to students to improve their learning.

Popham (2009) asserts that the assessment literacy necessary for teachers could be provided through professional development practices, and therefore, proposes a list of contents that should be included in a professional development programme for teachers which focuses on assessment literacy. In this list, Popham (2009) not only focuses on technical issues in assessment practices such as reliability, validity,

construction of test items in various forms, and forming rubrics, but also highlights the use of formative assessment procedures such as portfolio and performance assessment, and that learners with special needs should be taken into consideration.

To sum up, language assessment literacy is composed of many layers including background knowledge of principles in LTA, and necessary skills to practise sound assessment procedures. Competency in language assessment literacy is a must for English language teachers to make the right decisions regarding learning and instruction. In the light of the framework mentioned above, this study is conducted to find out the training needs of English language instructors employed at English Preparatory Programmes at Turkish universities.

### **1.3 Statement of the Problem**

Assessment is an indispensable part of teaching and learning. Teachers use assessment in order to gather efficient information from students about their learning, make judgements and decisions for further steps in students' learning, and make modifications to course programmes so as to improve instruction and learning. Teachers who are assessment illiterate may make wrong decisions about their students, which may cause undesired effects and even a hindrance in student learning or about students' future. However, as mentioned earlier, pre-service teachers do not get enough training in language testing and assessment (Stiggins, 2002; Taylor, 2009; Hatipoğlu, 2010). Many teacher candidates graduate from university lacking the necessary knowledge and skills to be assessment literate. Therefore, during their active teaching, many teachers evade from assessment related issues due to lack of their assessment literacy. Moreover, there are some English language instructors working at English Preparatory Programmes who have not at all received any pre-service education in LTA. Therefore, it is crucial that English language instructors' background knowledge in LTA and their training needs in this area should be investigated. Given all these problems, this study investigates the reported training English language instructors received in LTA, and their perceived training needs in this area. More specifically, the study investigates the differences in English language instructors' reported LTA training and perceived training needs regarding

their major, highest education degree, LTA responsibilities in testing units, and teaching experience.

#### **1.4 Purpose of the Study**

The main purpose of this study is to examine if there is a statistically significant difference in English language instructors' reported LTA training and their perceived training needs in accordance with their educational background, teaching experience, and LTA responsibilities in testing units. The study also explores the levels of training English language instructors at English Preparatory Programmes of Turkish universities report in LTA areas and to what extent they perceive a need for in-service training in these areas. The study also aims to investigate English language instructors' professional learning priorities and preferences in approaches to classroom assessment.

#### **1.5 Research Questions**

This study aims to find the answers for the following research questions:

1. What level of training do English language instructors report in various areas of LTA and to what extent do instructors perceive a need for in-service training in these areas?
2. Is there a statistically significant difference in English language instructors' reported LTA training and perceived training needs regarding their major?
3. Is there a statistically significant difference in English language instructors' reported LTA training and perceived training needs regarding their highest educational degree?
4. Is there a statistically significant difference in English language instructors' reported LTA training and perceived training needs regarding their LTA responsibilities in testing units?
5. Is there a statistically significant difference in English language instructors' reported LTA training and perceived training needs regarding their teaching experience?

6. What are English language instructors' perceptions of LTA training in pre-service and in-service education?

### **1.6 Significance of the Study**

Determining the training levels and training needs of English language instructors in LTA is of critical priority for various reasons. First, English language education is very important at Turkish universities. While the medium of instruction is English in some departments of universities, some others offer all the courses in all their faculties in English. Students who want to study at these universities or departments must pass a proficiency test in English. Those who cannot are required to study at English Preparatory Programmes so that they are able to follow their courses in their departments in the next four years. Therefore, the English language education the students receive throughout the English Preparatory programme is crucial for them. It is the English language instructors' job to teach English to those students at preparatory programmes. At this point, it is rightful to mention that students in Turkey start learning English in their second year of primary school education. In other words, English language instructors at Preparatory programmes try to teach English to students who have been learning English for nearly 10 years, but unfortunately have not acquired the necessary skills for university education yet. Therefore, English language instructors should be well-equipped and qualified not only in their teaching skills but also in their assessment knowledge and skills to help their students learn English.

Second, English language instructors bear a lot of assessment responsibilities. Most preparatory programmes offer an intensive course of English, which means, too much input is given in a short period of time. This undesired situation causes numerous quizzes and exams that they have to deal with. Moreover, a teacher with assessment literacy should be aware that assessment is not only about testing or grading students. Using informal, continuous, non-test type of assessment such as peer-assessment, self-assessment or giving continuous feedback to students is also vital to enhance the quality of learning and instruction.

Last but not least, language assessment literacy is gaining importance in language teaching because of the increasing number of tests and assessments applied

throughout the world and in Turkey. However, although some studies have been conducted to examine the assessment literacy of teachers in Turkey (Hatipoğlu, 2010, Yetkin, 2015; Büyükkarcı, 2016), a research study pertaining to the training needs of English language instructors in LTA has not been carried out in the context of Turkey before, as far as the researcher is concerned. Therefore, this study aims to shed light on how English language instructors perceive the training they have received in LTA and in what areas of LTA they perceive a need for training.

For all the reasons mentioned above, it is crucial that the training levels and training needs of English language instructors is investigated, the differences in their reported training and perceived training needs in LTA are determined and thus, necessary precautions involving all stakeholders are put forward with further suggestions from the findings of the current study.

## **1.7 Definitions**

**ELT:** English Language Teaching

**In-service education:** In-service education refers to any training received during the period while a teacher has been employed.

**Language tester:** In this study, a language tester refers to an English instructor who has additional assessment roles in Testing and Assessment Office at an English Preparatory Programme.

**LTA:** Language Testing and Assessment

**Non-ELT:** In this study, non-ELT refers to any BA degree other than English Language Teaching (ELT), e.g. English Language and Literature, Translation and Interpreting or Linguistics.

**Pre-service education:** Pre-service education refers to any training received during the period before a teacher has undertaken any teaching.

**Regular instructor:** In this study, a regular instructor refers to an English instructor who does not have any additional duties regarding testing and assessment.

## **Chapter 2**

### **Literature Review**

This chapter starts with the definitions of certain concepts in LTA area. Then, types, purposes, role / importance of assessment are presented. Later, studies on assessment literacy in general, studies on language assessment literacy, studies on language assessment courses, and studies in on language assessment literacy in Turkey are reviewed.

#### **2.1 Overview**

Test, assessment, and evaluation are certain basic concepts in the field of assessment literacy, which are commonly used, but often misunderstood and used interchangeably in incorrect ways causing confusion among teachers. It would be wise to start with the definition of these concepts, clarify their meanings and analyse how they are different from each other.

Tests are tools used to assess students. Thus, tests may be defined as a subsection of assessment (Coombe, Folsie, & Hubley, 2007). Cizek (2009) states that “a test is any systematic sample of a person’s knowledge, skill, or ability” (p. 64). When a student takes a test, teachers may only observe a sample of what the student knows or is able to do. Therefore, tests aid teachers only to a certain extent to see what students are capable of, and teachers need to make inferences from the test results to reach conclusions about the student, which leads to assessment (Cizek, 2009).

Coombe et al. (2007) defines assessment as using different techniques to gather information about learners’ abilities and achievements. It is stated as “an umbrella term for all types of measures used to evaluate student progress” (Coombe et. al, 2007, p. xv). Brown (2004) also points out that “assessment is an ongoing process” in the classroom. Whenever a learner produces some form of output, teachers tend to make judgments based on them (Brown, 2004). Cizek (1997) states that assessment is not only gathering information but also synthesizing it in order to find out and label learners’ strong and weak sides; and to plan and improve teaching. Therefore,



teachers need assessment to make decisions about students' learning and their own teaching, which makes it a sine qua non in language teaching and learning process.

Evaluation, on the other hand, is a much broader term including both testing and assessment. Brindley (1989) describes evaluation as “conceptualized as broader in scope, and concerned with the overall program” (p. 3). Genesee and Upshur (1996) point out that second language evaluation is not only about coming to a conclusion whether a student should pass or fail. It is more about making decisions considering many steps such as placement of students in levels or courses, continuing instruction, making plans about future instruction, revising the curriculum, and many other issues pertinent to learning and teaching (Genesee & Upshur, 1996). In short, evaluation requires gathering information from all possible sources including students and all the course or programme components and making decisions to enhance learning (Coombe et al. 2007).

## 2.2 Types of Assessment

Assessment in the classroom may take many forms depending on why to assess, how to assess, who to assess, etc. Coombe et al. (2007) summarize common types of language assessment in Table 1.

Table 1

*Common Types of Assessment*

Informal	Formal
Classroom	Standardized
Criterion-referenced	Norm-referenced
Placement	Proficiency
Direct	Indirect
Subjective	Objective
Formative	Summative
Alternative, authentic	Traditional tests

**2.2.1 Formal and informal assessment.** Since formal and informal are not psychometric terms, they do not have a systematically accepted definition (Navarete,

Wilde, Nelson, Martinez, Hargett, 1990). Informal assessment may be seen as observing and evaluating learners in the continuum of learning. Informal assessment techniques can easily be applied to daily classroom activities. Examples of informal assessment include unplanned, on-the-spot feedback, advice about how to pronounce a word better, and saying “Good job!” to a student as a response (Brown, 2004).

Formal assessment, on the other hand, is the official way of investigating learning achievement. They are planned and have systematic grading procedures. Brown (2004) noted that not only all tests are kinds of formal assessment, but students’ journals and portfolios are also examples of formal assessment.

**2.2.2 Formative and summative assessment.** Formative assessments are conducted during a course in order to use the results to develop instruction (Coombe et al., 2007). Stiggins (2005) states that formative assessment is implemented not only to mainly make judgements or give marks, but also to enhance learning during learning. Popham (2008) also points out that formative assessment is a planned process and during this process both teachers and students utilize evidence from assessment so as to modify what they are doing (p. 6). Winger and Norman (2005) identify three main purpose of formative assessment as (a) informing teachers about students learning during instruction for the purpose of guiding and modifying instruction; (b) providing corrective feedback to students about learning progress for the purpose of guiding and improving learning; and (c) enhancing student motivation.

On the other hand, the main purpose of summative assessment is to determine how much a student has learned and succeeded in the objectives of the course (Brown, 2004). Harlen and James (1997) state that summative assessment “describes learning achieved at a certain time for the purposes of reporting to parents, other teachers, the pupils themselves and, in summary form, to other interested parties such as school governors or school boards” (p.370). Therefore, final exams of a course or general proficiency exams may be among the examples of summative assessment.

**2.2.3 Standardized assessment and classroom assessment.** Tests which happen to have significant consequences on learners’ lives are called standardized tests. In the context of Turkey, university entrance exam may be a good example for high-stakes tests.

Classroom assessment, on the other hand, is more associated with formative assessment, and is regarded as more of a tool to inform learning and teaching (Rea-Dickens, 2007). Formative types of assessment such as self- or peer- assessment, portfolios, and performance assessment are some tools used for classroom assessment as well as giving continuous feedback to improve learning (McMillan, 2003).

**2.2.4 Norm-referenced and criterion-referenced tests.** Norm-referenced tests are used to rank test takers with the aim of comparing them to each other. Test takers' scores are aligned so as to be put in a rank order. Test of English as a Foreign Language (TOEFL) is an example for norm-referenced tests (Brown, 2004).

Criterion-referenced tests aim to measure how well the objectives of a course or programme have been learned (Coombe et. al., 2007). According to Hughes (1996), criterion-referenced tests sort learners in accordance with their performance on pre-set tasks. In other words, the performance of learners is assessed according to their ability in how well the pre-set tasks are performed. The main purpose of these tests is to give feedback to learners (Brown, 2004).

**2.2.5 Placement and proficiency tests.** While placement tests are used to identify the language skills and abilities of learners so as to be placed in a suitable class or course, proficiency tests are designed to assess general language ability of test takers. Proficiency tests mainly portray how much a test taker is capable of the language (Coombe et al., 2007).

**2.2.6 Traditional and alternative assessment.** Coombe et al. (2007) note that an effective way of explaining alternative assessment is to differ it from traditional assessment. Unlike traditional assessment, which is mainly made up of decontextualized multiple-choice test items, alternative assessment uses various techniques which are more authentic to gather information from learners. These techniques may be formed through contextualised, meaningful and authentic communicative tasks which require learners to utilize and show various language skills and which provide assessors with multidimensional information about the learners. Portfolios, projects, and presentations are among common types of alternative assessment.

## **2.3 Purposes of Assessment**

Assessment in language learning is used for various reasons preceding, during, and after learning. These reasons include placing learners into appropriate levels, motivating learning, creating learning opportunities, giving feedback, grading, etc. Undoubtedly, it is essential for assessment to set clear purposes and that it should be designed to fit to those purposes. Thus, the purposes of assessment are divided into three subsections: assessment for learning, assessment as learning, and assessment of learning.

**2.3.1 Assessment for learning.** Assessment for learning is defined as “part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning” (Klenowski, 2009, p. 264) in the Third International Conference on Assessment for Learning in Dunedin, New Zealand.

Stiggins (2005) labels assessment for learning as “to use many different assessment methods to provide students, teachers, and parents with a continuing stream of evidence of student progress in mastering the knowledge and skills that underpin or lead up to state standards.” (p. 327). Assessment for learning is an ongoing process. Therefore, it is implemented all along the learning process (Earl & Katz, 2006, p. 29). Moreover, learners are involved in the assessment process as they see how much they have learned and succeeded in the targets that have been set for them. Assessment for learning keeps learners aware of “where they are in relation to where they want to be” (Stiggins, 2005, p. 328).

**2.3.2 Assessment as learning.** Fletcher (2016) notes that assessment as learning is a term the definition of which has varieties and not agreed upon. Dann (2002) states that assessment as learning is “most notably promoted through the process of self-assessment” emphasizing the involvement of learners in the process of assessment (p. 67). Gupta (2016) points out that assessment as learning aims to help learners to observe their own learning through self-correction. As learners monitor their learning, they become critical assessors of it, making judgements about and reflecting on what and how they have learned (Earl & Katz, 2006).

**2.3.3 Assessment of learning.** Assessment of learning aims to find out whether learners have achieved the outcomes of a curriculum or the goals of the programme

they have taken part in. It also provides information and feedback about learners' abilities so that decisions or judgements can be made about learners' future. Since the results of assessment of learning may affect learners' future to a great extent, it is paramount that assessment of learning be undoubtedly reliable (Earl & Katz, 2006).

## **2.4 The Role / Importance of Assessment**

Popham (2006) points out that assessment plays a crucial role in education. All teachers are aware that, unfortunately, "if work does not have marks attached, many students will either not do it at all or only do it in a perfunctory way" (Rust, 2002, p. 146). Many students would probably avoid classes and learning if they weren't to be assessed. Therefore, it may well be said that assessment has a great impact on triggering learning.

According to Stiggins (2008), assessment has two important roles in education. While one is to enhance learning as mentioned above, the other is to collect data to make instructional decisions. When implemented effectively, assessment supplies teachers, students, and other stakeholders with valuable information regarding performance of learners and how much of the objectives in the classroom have been achieved (Malone, 2013).

Besides, assessment may and should work together with instruction, basing a reciprocal relationship where teaching and assessment are informed by and enhanced from each other (Malone, 2013).

## **2.5 Research on Assessment Literacy**

**2.5.1 Research on teachers' assessment literacy.** Plake, Impara, and Fager (1993) gauged the assessment literacy of 555 teachers from all around the USA using The Teacher Assessment Literacy Questionnaire. The questionnaire included all assessment related competencies identified by the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT et al. 1990). It consisted of 35 items in total and 5 items were allocated to each standard. The results indicated that nearly more than 23 out of 35 items were correctly answered on average. While the participants' highest performance was on standard 3

(administering, scoring, and interpreting the results of assessments), their lowest performance was on standard 6 (communicating assessment results to stakeholders). 10 of the 35 items pertinent to choosing appropriate assessments, adequate test taking behaviour in standardized tests, communicating assessment results to parents, and recognizing unethical practices during standardized test administration were answered correctly by at least 90% of the participants. However, 2 of the items regarding standard 5 (developing valid grading procedures) were answered correctly by 30% or less of the participants. Furthermore, 1 item concerning reliability of a test score was correctly answered by only 13% of the participants.

Lomax (1996) carried out a study focusing on pre-service teachers. A wide range of data collection tools such as pre-instruction assessment journals, self-reflection assessment journals, copies of assessments used by the preservice teachers during their in-class experiences, and individualized semi-structured interviews were utilized to examine the assessment knowledge of 5 undergraduate elementary pre-service teachers before and after they participated an assessment course at a state university in the USA. The findings were analysed qualitatively. The results indicated that grading formed the main concern of the participants. Besides, the participants experienced the difficulty of having to teach to the test due to the standardised tests and the pressure of tests for teachers. It was also found out that the participants used text-supplied tests and worksheets to a greater extent than teacher-made tests. Nevertheless, it was noted that the course provided the participants with a substantial increase in their assessment literacy, which points out the significance of an assessment course in a teaching programme.

Mertler (2005) investigated the assessment literacy levels of both 67 pre-service and 101 in-service teachers at the secondary level and compared the results for these two groups. The *Classroom Assessment Literacy Inventory* (CALI), which was designed to align with the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT et al. 1990), was used to survey the participants. The CALI consisted of 35 items and five sets of questions for each standard. Pre-service teachers answered nearly 19 out of 35 items correctly overall. Items related to standard 1 (choosing appropriate assessment methods) had the most correct answers and standard 5 (developing valid grading procedures) the least. In-service teachers, on the other hand, answered almost 22 out of 35 items correctly on average, standard

3 (administering, scoring, and interpreting the results of assessments) with the most correct answers and standard 5 with the least correct answers. In all cases the in-service teachers scored higher than the pre-service teachers, showing that the in-service teachers participating in the study appeared to be more assessment literate than the pre-service teachers.

Volante and Fazio's (2007) research study aimed to investigate the assessment literacy of pre-service teachers. 69 primary / junior teacher candidates taking part in the study completed a survey with open- and close-ended questions related to level of self-efficacy, purposes of assessment, use of various assessment methods, and improving assessment literacy. The findings revealed that the candidates rated their own assessment literacy as relatively low regardless of their year in the programme. Nearly 75% of the candidates used assessment for summative purposes. Pertaining to the assessment methods used and need for further training, the candidates rated authentic assessment methods such as portfolio and performance assessment the highest. Moreover, the candidates favoured a specific course on classroom assessment and evaluation. Volante and Fazio (2007) noted that the programme the candidates studied at was not fulfilling enough for them to present a thorough understanding of diverse approaches to assessment and evaluation, which aligned with teacher education research literature.

In another study, Mertler (2009) investigated the effect of a two-week classroom assessment workshop on 7 in-service teachers. The participants were all elementary level teachers and were from five different school districts and they had never received any formal education on classroom assessment. The workshop was built upon the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT et al. 1990), and was based on practical activities on performance assessment tasks. The participants were tested before and after the workshop utilizing the 35-item *Assessment Literacy Inventory* (ALI) by Mertler and Campbell (as cited in Mertler, 2009, p. 105), which was formed to align with the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT et al. 1990). Then, the mean scores of the pre-test and post-test were compared. Moreover, the participants were asked to write daily journals on their experiences during the workshop, which were also used as data source. Regarding the results of the ALI, the mean score of the post-test was 28.29 while the pre-test's was 19.57. The comparison

of the pre- and post-test mean scores revealed that the participants' assessment literacy developed dramatically. The two greatest improvements were observed on standard 5 (developing valid pupil grading procedures which use pupil assessments.) and standard 2 (developing assessment methods appropriate for instructional decisions.). Furthermore, each one of the participants' individual score also increased from pre-test to post-test. The results from the analysis of teachers' daily journals showed that the participants all had a positive experience pertinent to the workshop and built confidence on their assessment skills.

**2.5.2 Research on language teachers' assessment literacy.** Hasselgreen, Carlsen, and Helness (2004) conducted a survey called *European Survey of Language Testing and Assessment Needs* across Europe to find out the LTA needs of language teachers, language teacher trainers and experts. The survey started with questions to gather background information of the participants and the rest of the questions were designed to address the needs of each of the three stakeholders and were grouped in three sections as classroom-based activities in LTA, purposes of testing, and content and concepts of LTA. Moreover, each question in these sections included three parts regarding if the respondents had been involved in the related type of LTA area, if they received any formal training in that area, and if they perceived a need for training in that area. 864 responds from 37 European countries and 50 responds from non-European countries were received adding up to 914 responds in total. The findings revealed that most of the activities mentioned in the survey were performed by teachers who had little or no training. Most teachers perceived a need for training as responses in the survey were majorly weighted towards 'yes'. Regarding classroom-focused assessment, teachers generally perceived a need for training in using portfolios, preparing own tests, peer/self-assessment, interpreting results, using continuous informal assessment, and giving feedback. Regarding content and concepts teachers perceived a need in assessing aspects of culture, assessing integrated skills, establishing reliability and validity, and statistics. Moreover, defining assessment criteria, reviewing and writing items, rating and interviewing were also among the areas which teachers perceived a need for training.

A mixed methods research study carried out by Vogt and Tsagari (2014), which also inspired this study, focused on foreign language teachers' assessment



literacy levels. The study investigated the background of foreign language teachers and their training needs and perceptions in various areas of LTA under three sections: classroom-focused LTA, purposes of testing, and content and concepts of LTA using an adapted version of *European Survey of Language Testing and Assessment Needs*, which was originally designed by Hasselgreen et al. (2004). 853 survey responses from teachers across 7 European countries and 63 semi-structured teacher interviews conducted in selected countries were used as data collection instruments. The findings from the survey revealed that nearly half of the respondents had no training in the area of purposes of testing such as giving grades and placing students. Nearly one third of the respondents received either no or little training in classroom-focused LTA practices, especially in alternative assessment methods. More than half of the respondents received either no or little training regarding content and concepts of LTA such as testing four skills, and validity and reliability issues. Moreover, a need for advanced training for this section was observed as nearly half of the respondents said they needed advanced training in this area. On the other hand, the findings from the interviews indicated that the majority of the interviewees received limited training regarding LTA during both their pre-service and in-service training. Most of the teachers also reported that they learned about concepts of assessment from colleagues or mentors on the job.

Kvasova & Kavytska (2014) investigated the university foreign language teachers' LTA literacy in Ukraine using the replication of the survey by Vogt and Tsagari (2014). The respondents to the survey were 35 foreign language teachers across Ukraine who participated in a TESOL Ukraine summer institute. First, the survey was completed by 25 foreign language teachers following a semi-structured interview with 10 young teachers out of 25. Later, additional 10 responses from expert teachers to the survey were collected to achieve more objective data. The findings of the first survey indicated that nearly 70% of the participants acknowledged themselves to be trained in LTA while 83% perceived a need to receive more training in order to improve their assessment literacy. It was pointed out that despite perceiving themselves as trained in LTA, perceiving a need for more training could be attributed to that the participants were highly motivated for professional development.

Zhang and Burry-Stock (2003) looked into teachers' assessment practices across teaching levels and content areas, and their self-perceived assessment skills regarding teaching experience and measurement training they received. Data from 297 teachers, who taught in various levels and content areas, were collected via Assessment Practices Inventory (as cited in Zhang & Burry-Stock, 2003, p. 326). The inventory included 67 items covering a great variety of assessment activities. The participants were asked to mark their answers as to how often they used the activities, and how skilled they perceived themselves on these activities. The findings revealed that teachers who taught higher levels depended more on objective tests in classroom assessment and were more concerned about the quality in assessment. Those who received training in measurement reported a higher level of self-perceived assessment skills in factors such as using performance measures, standardised testing, test revision, instructional improvement, and communicating assessment results compared to those who did not receive training in measurement.

Cheng, Rogers, and Hu (2004) carried out a study to document and compare the assessment practices, methods, and procedures utilised by the university instructors in the Canadian, Chinese, and Hong Kong contexts. A survey consisting of five parts regarding major constructs in classroom assessment was used to collect data from the participants. 95, 44, and 124 ESL/EFL instructors participated in the study from Canada, Hong Kong, and China, respectively. According to the results, five most common purposes of assessment as being student centred were identified as gathering information on students' progress, providing feedback to students, analysing strengths and weaknesses of students, grading, and motivating students. While the most common instructional purposes of assessment were for planning instruction and determining strengths and weaknesses in instruction; providing information to the central administration was the major administrative purpose of assessment. Regarding the methods used for skills assessment, the prominent methods used for reading skill were short answer items and student summaries of what was read. On the other hand, while using short essays was the most common method for the assessment of writing skill among three contexts, student journals stood out as the second most common method in Canada. As for the assessment of speaking skill oral presentation was the most common type of assessment method. Moreover, regarding the procedures in assessment, it was found out that most of the

instructors in all contexts developed their own assessments whereas instructors from China depended much more on printed sources than the instructors in Canada and Hong Kong. Furthermore, the average time spent on assessment was found to be nearly a quarter of the total time of working hours.

Lopez Mendoza and Bernal Arandia (2009) carried out a research study to explore English language teachers' views and perceptions pertinent to language assessment and also how they utilize language assessments in their classrooms. An online qualitative survey was used to collect data. The survey was designed to obtain information regarding teachers' feelings on and usage of assessment, and procedures about giving feedback to students. Eighty-two English language teachers across Colombia who taught in different levels responded to the survey. The responses were analysed and emerging themes related to the study were coded. Similarities and differences between views of teachers with training in language assessment and teachers without training were also investigated. The findings revealed that the participants were more inclined to use traditional assessment rather than alternative, and the feedback provided for students mostly occurred at the end of learning process in the form of grades. Therefore, it was interpreted that assessment in language classrooms in Colombia was more summative than formative. Moreover, a correlation between language assessment training and perceptions about language assessment was observed. That is, trained teachers had a positive perception of assessment and were more aware of the role of assessment in enhancing teaching and learning process. Teachers without training in assessment were more inclined to possess a negative perspective on assessment. Assessment was mostly used to give a grade or make judgements about students but not to improve learning.

A study carried out by Kiomrs, Abdolmehdi and Rashidi (2011) focused on test washback and teacher assessment literacy among EFL secondary school teachers. Kiomrs et al. (2011) implemented an assessment literacy test to 53 EFL secondary school teachers. The maximum possible score from the test was 35 whereas the mean score of the participants turned out to be 10, which reflected that the participants' level of assessment literacy was rather low.

Hakim (2015) conducted a study to find out the ideology of English language teachers on ELT assessment literacy. The researcher examined the perceptions and practices of 30 female teachers on English language assessment using a questionnaire

developed by the researcher. The results showed that overall the participants acquired an efficient understanding of English language assessment regardless of their teaching experience.

Jannati (2015) conducted a study on 18 Iranian EFL teachers to find out how adequately the participants perceive, apply and make use of assessment; whether there exists a difference between high- and low-experienced teachers' perception and use of assessment; and to what extent language assessment literacy is observed in their practices. The findings revealed that the participants were aware of the main concepts and terminologies in language assessment, and teaching experience did not have a distinctive importance on how teachers perceived assessment. Jannati (2015) attributed this result to the fact that the participants were studying for a Master's degree and had a course on language testing when the interviews took place. Nevertheless, Jannati (2015) asserted that despite being somehow assessment literate, the participants of the study failed to transfer their knowledge into practice.

Tsagari (2016) explored classroom based language assessment practices, knowledge, and skills of Greek and Cypriot primary school EFL teachers. 4 Greek and 4 Cypriot state primary school EFL teachers were interviewed with open-ended questions to collect data. The findings showed that tests were majorly applied as a source to measure students' learning and to assess the efficiency of instruction, and the content of the tests were limited to vocabulary and grammar assessment with lack of creativity in terms of task types used. It was attributed to low language assessment literacy levels that alternative forms of assessment were not used effectively. Overall, the participants in the study depended heavily on traditional assessment procedures. Furthermore, although teachers wished to receive training on professional development, they had difficulty in identifying their needs. Finally, the evidence from the data indicated that the participants didn't see themselves to be literate in classroom based language assessment.

A study by Crusan, Plakans and Gebril (2016) focused on writing teachers. A survey exploring writing teachers' knowledge, beliefs and practices on the assessment of writing skill was completed by 702 participants from 41 countries on 5 continents, which revealed intriguing results. While 80% of the participants reported having had training in writing assessment, 20% reported they received no training. However, 63% of all the participants contended they learned about assessment in

general and 57% learned about writing assessment as part of a course. 40% of the participants who wrote a comment to the related question in the survey stated that their training was mostly acquired on the job. It was also found out that novice teachers feel more confident about the writing assessment knowledge they possess than experienced teachers do.

In their recent study, Xu and Brown (2017) analysed the psychometric properties of the adapted *Teacher Assessment Literacy Inventory* (TALQ), and investigated university English language teachers' assessment literacy in China, and whether demographic characteristics affected their assessment literacy performance. For the adaptation of the questionnaire to the Chinese context, the number of 35 items was reduced to 24 items with shorter and more personalised prompts. Some items directly related to the US context were omitted. The questionnaire was filled out by 891 university English language teachers across China. The findings of the psychometric analysis of the questionnaire revealed that the validity for the questionnaire existed for a limited number of items (10 items out of 24). It was inferred that the items in the TALQ were not appropriate to measure the assessment literacy level of Chinese university English language teachers. It was suggested that items for such measurement test be revised deeply to align with Chinese assessment context. Moreover, the findings of the analysed 10 items suggested basic assessment literacy among the participants. It was also noted that a wide range existed in the participants' assessment literacy competencies. In other words, while a minority possessed a good degree of assessment literacy, most of the participants had basic assessment literacy in domains such as choosing appropriate assessment methods, clarity regarding the purpose of assessments, including students to the assessment procedures and valid grading procedures.

Sheehan and Munro (n. d.) did a project on teachers' practices, attitudes, and needs pertinent to assessment. They used individual, focus group, and follow-up interviews, observations, and a workshop as data collection tools. The results of the interviews revealed that many of the participants in the project lacked training in testing and assessment and, therefore, confidence about their assessment practices. Some participants stated that their assessment skills grew through experience, while some others reported that they made inferences from and connections with how they were tested when they were students. It was noted that teachers also depended on

each other to improve their own assessment skills. The findings of the observations implied that the participants, in fact, applied various assessment techniques in class, yet did not perceive them as a constituent of assessment.

**2.5.3 Research on language assessment courses.** A nationwide study by Jin (2010) investigated the place of LTA courses in the training of foreign language teachers at tertiary level. 86 instructors of language testing assessment course throughout China filled in a survey adapted from Brown & Bailey (as cited in Jin, 2010, p. 556). The survey focused on teaching content, teaching methodology, student perception of the courses and teaching materials used in courses. The findings revealed that language testing courses were not compulsory for 60% of the programmes surveyed. The classrooms were too crowded to spend enough time for discussion and practice. The courses mainly focused on the theory and practice of language testing while educational and psychological measurement procedures received less attention. Topics on assessment such as alternative assessment, formative and summative assessment did not have efficient coverage. Classroom practice wasn't given enough priority of the instruction time. Instructors thought that students taking the course had motivation whereas the courses were challenging for them.

Hatipoğlu (2010) conducted a study to find out pre-service teachers' views and evaluations of a testing and assessment course regarding its content and teaching methodology. The study was based on the undergraduate English Language Testing and Evaluation (ELTE) course delivered at Middle East Technical University in Ankara, Turkey. A short questionnaire responded by 81 pre-service teachers who took the course was used as the first source of data. The questionnaire included two main questions. The respondents were asked to name 5 topics they learnt which they thought would help them in their career as an English teacher, and name 3 things to be changed to improve the quality of the course. The respondents were also asked to elaborate on the questions. After the analysis of the questionnaire, 16 volunteer pre-service teachers having taken the course were interviewed to obtain the second source of data. The study showed that the most favourable topics were testing/assessing skills/knowledge, reliability, and validity, respectively. The respondents found these 3 topics very useful while topics such as teaching and testing, kinds of tests, kinds of testing and writing multiple choice item tests were not

found related to their future careers. Moreover, the most distinct suggestion regarding the improvement of the course was to increase the amount of practice. 37% of the respondents thought the balance of theory and practice in the course should be more towards practice than to theory. Furthermore, the respondents did not think they had enough opportunity to enhance their assessment literacy within the limit of the exercises carried out during the course. As a result of the study, necessary steps were put into action for the following academic year.

Jeong (2013) conducted a mixed methods research study on how language assessment course instructors shape and construct their courses depending on their background in language testing and assessment. A survey and a semi-structured interview were used as data collection tools. The survey was made up of questions addressing the amount of time allocated for certain topics and topics which instructors considered important for their students. Two instructor groups were included in the study: language testers and non-language testers. The language testers were those with primary research interest in the area of language testing. The non-language testers were those who had primary interest in other areas of language teaching, but who were involved in language assessment activities. The results of the study indicated that language assessment courses tend to differ regarding the instructor's background in the area of language testing. While structure and organisation of the course for both groups were similar, language testers spent more time on technical topics such as test specifications, test theory and statistics; non-language testers focused more on classroom related areas such as alternative assessment and classroom assessment. Regarding the language assessment textbooks used in courses, non-language testers preferred books which consisted more day-to-day classroom activities. It was pointed out that such differences in courses brings into mind the question "whether instructors of language assessment courses share a common definition of language assessment literacy" (Jeong, 2013, p. 354).

Malone (2013) was involved in a project to contribute to the development of an online tutorial which was established to aid foreign language instructors on language testing and assessment. While developing the tutorial, feedback from language instructors and language testing experts was elicited to improve the content and presentation of the website. Within this frame a research study was carried out to compare how the two groups' reactions differed regarding the content and

presentation of the online tutorial. Group interviews, written feedback, and an online survey were utilized to elicit feedback from 44 language instructors and 30 testing experts in the USA pertinent to their views on the tutorial. The data were analysed analytically and emerging themes were categorised. The findings showed that on their feedback about the tutorial, language testing experts were inclined to emphasize the preciseness of definitions described on the tutorial and features of adequate test use. However, language instructors were more interested in how clearly and briefly the information on the tutorial was presented and delivered. It was concluded that material developers who are devoted to encouraging assessment literacy should keep in mind that the interests and needs of language instructors and language testing experts may be different from each other.

A study by Montee, Bach, Donovan and Thompson (2013) investigated the assessment practices and training needs of Less Commonly Taught Languages (TCTL) teachers and looked into how LTCL teachers' assessment literacy could be improved through a training course. 29 of the LCTL teachers, who attended a training course about assessment were the participants of the study. Regarding the first source of data, the participants were asked to complete a questionnaire so as to find out about their interests, beliefs, needs, and practices on assessment. This questionnaire was also used to shape the course content in accordance with the participants' needs. As part of the course, the participants were required to complete some assessment tasks, which were used as the second source of data. The results of the study indicated that most of the participants had not received any assessment related course during pre-service education. On the other hand, the ones who taken a university course on assessment had a tendency to have more confidence in their assessment practices. Moreover, while many participants were alienated with most of the assessment terms, the analysis of the assessment tasks revealed that their teaching and assessment practices had strong connections.

In his study, Lam (2015) examined not only the general language assessment training landscape in Hong Kong and its impact on the development of pre-service teachers' language assessment literacy, but also instructors' and pre-service teachers' perceptions on course-based language assessment training to promote language assessment literacy. Lam investigated the content of 5 teacher training programmes in Hong Kong and interviewed 9 instructors and 40 pre-service teachers who were



studying in their final year. The results showed that although essential language assessment training was provided for pre-service teachers, there seemed to be a mismatch between theory and practice regarding the improvement of language assessment literacy in the Hong Kong school context and language assessment training was inadequate to promote language assessment literacy.

#### **2.5.4 Research on language teachers' assessment literacy in Turkey.**

Several studies about language assessment and language assessment literacy have been carried out in Turkey in the last few years.

Han and Kaya (2014) examined the assessment perceptions, practices, and preferences of 95 EFL teachers working in primary or secondary schools in a city in Turkey. A questionnaire was utilized to collect the data. It was found out that teachers considered speaking to be the most difficult skill to assess. Moreover, pre- or in-service assessment training did not have an impact on teachers' assessment preferences. Teachers mostly depended on their personal assessment preferences.

In his MA thesis, Yetkin (2015) carried out a study to find out the views of ELT pre-service teachers on assessment literacy and how ELT programmes contributed to their assessment literacy development. Volante and Fazio's (2007) questionnaire was adapted and used to collect data from 30 ELT pre-service teachers in Mersin, Turkey. Besides, 5 pre-service teachers were interviewed with open-ended questions from the questionnaire. The findings revealed that the participants acquired the necessary classroom assessment strategies they would need in their in-service practices. The courses taken in the ELT programme provided the pre-service teachers with an insight understanding of the theories regarding classroom assessment. Moreover, while the most favourite assessment approaches of the participants were found out to be observation techniques, personal communication, performance assessment and portfolio assessment, their least favourite assessment approaches were selected response and constructed response. Finally, the participants' further training needs on classroom assessment were in connection with their favoured assessment approaches. That is, they perceived a need for further training in using observation techniques, performance assessment, and personal communication.

A study carried out by Hatipoğlu (2015) aimed to find out the knowledge of pre-service teachers, who were studying in an ELT department, about language

testing and their thoughts about the content and methodology to be taught in their English Language Testing and Evaluation course. Data was collected data from 124 pre-service teachers at Middle East Technical University using needs analysis survey questionnaires and interviews between 2009 and 2012. The results of the study revealed that after studying four years in an ELT department, a great majority of the pre-service teachers had limited knowledge regarding English language testing and evaluation. Moreover, it was pointed out that pre-service teachers focused more on testing than on evaluation and assessment, which is probably due to their local assessment culture and context, which is highly exam oriented.

Büyükkarcı (2016) looked into foreign language teachers' assessment literacy and if year of experience and post-graduate education created a difference in the assessment literacy of language teachers. The data were collected from 32 in-service English language teachers working in primary schools and universities in Turkey using "Assessment Literacy Inventory" designed by Mertler and Campbell (as cited in Büyükkarcı, 2016, p. 336). It was found out that the average assessment literacy score of the participants was 12.90 out of 35, which revealed that the assessment literacy of the participants was very low. No correlation between year of experience, post-graduate education and high assessment literacy was observed, either.

## **Chapter 3**

### **Methodology**

#### **3.1 Overview**

This chapter starts with philosophical paradigm and research design for the study. Then, it continues with detailed information regarding setting and participants, and procedures adopted for data collection and data analysis.

#### **3.2 Philosophical Paradigm**

Paradigm is defined as “the consensual set of beliefs and practices that guide a field” (Morgan, 2007). The paradigm a researcher adopts is a summary of her or his world view (Doyle, Brady, & Byrne, 2009). The two most common research paradigms are quantitative and qualitative research. Quantitative research, which has been predominantly utilised in education research, is based on positivism. Positivism is a philosophical theory contending that authentic knowledge may be reached through scientific methods such as testing hypothesis and collecting objective data to reach systematic and generalizable conclusions (Ary, Jacobs, & Sorensen, 2010). In quantitative research, the data is gathered in a way that can be transformed into numbers so that it may be used as statistics.

Qualitative research, on the other hand, is different from quantitative research in nature. It aids the researcher to have an insight of individuals’ behaviour, opinions, or motivations, and how individuals see the world around them. Since social reality is considered to be different from individual to individual, events and experiences have different meanings for people. With qualitative research, the researchers aim to portray and interpret the participants’ social reality (Hesse-Biber, 2010).

Apart from qualitative and quantitative research, mixed methods research has gained acceptance as a third methodological movement in the last few decades, which combines both qualitative and quantitative research (Tashakkori & Teddlie, 2003). While mixed methods research as a field is relatively new and there is still debate on its compatibility (Doyle et al., 2009), it is considered as an advantageous research method. Johnson and Onwuegbuzie (2004) note that by adopting mixed methods research in a study, a researcher may combine both qualitative and

quantitative paradigms, taking advantage of each other's strengths and weaknesses. Besides, a wider range of research questions may be investigated as the researcher has the freedom to apply more than one method. Moreover, the numerical data or statistics from quantitative method may be enriched with personal experiences or narratives collected through qualitative method (Johnson & Onwuegbuzie, 2004).

### **3.3 Research Design**

This study aimed to find out EFL teachers' background knowledge base in various areas of LTA, to what extent they received training in these areas during their pre- and in-service education, and their perceived training needs in various fields of LTA. More specifically, the study examined if the reported training levels and perceived training needs of English language instructors in LTA suggested a statistically significant difference regarding their educational background, teaching experience, and LTA responsibilities in testing units.

A concurrent mixed-methods research design was adopted for this study to benefit the advantages of quantitative data in order to have a generalizable overview of the LTA training received and needed by English language instructors working at English Preparatory Programmes in Turkey, and to benefit the advantages of qualitative data which provided detailed analysis of a certain number of informants. That is, both quantitative and qualitative data were collected and analysed separately, and the findings from both set of data were joined in order to make inferences for the research questions. Moreover, the qualitative data was used to serve as complementary in order to clarify and elaborate on the quantitative data. Therefore, it may visually be represented as a QUAN + qual research study according to Morse's (as cited in Ary et al., 2010) notation system.

The questionnaire, designed by Hasselgreen et al. (2004) and adapted by Vogt and Tsagari (2014) served as the quantitative part of the study (see Appendix A). The semi-structured interview, which explored EFL teachers' training biographies individually, the strategies they used to make up for lack of expertise in LTA, and their personal needs pertinent to in-service teacher training in various fields of LTA, represented the qualitative part of the study (see Appendix B). The questions in the semi-structured interview were also adapted from Vogt and Tsagari (2014) study.

### 3.4 Setting and Participants

**3.4.1 Setting.** The study was carried out among English language instructors working at English Preparatory Programmes of universities in Turkey between February 20 and April 18 in 2018.

In Turkey, English language teachers are trained in ELT departments. Currently, 58 universities in Turkey have an ELT department. Pre-service teachers receive a four-year-education in these programmes and successful graduates are entitled as English language teachers. These graduates are eligible for working at any level in Turkish education system from primary to tertiary level including English Preparatory Programmes of universities. All ELT programmes follow a curriculum standardised by Higher Education Council (YÖK). In other words, all ELT programmes must offer certain must courses to their students to achieve the curriculum standards set by YÖK. However, course contents may be designed or re-designed by the lecturers of those courses pertaining to their students' local needs.

According to the standardised curriculum by YÖK, there are three types of courses in ELT programmes: field courses, education courses, and general knowledge courses. LTA courses offered at these programmes is one of the field courses defined by YÖK (Hatipoğlu, 2010). According to a survey by Hatipoğlu (2010), of the 28 ELT programmes in Turkish state universities, none offers more than one LTA course. Moreover, lecturers who teach this course have to deal with what and how to teach the content areas of LTA in a single course (Hatipoğlu, 2015). Furthermore, while previous studies indicate that pre-service teachers perceive LTA courses highly important as tests and exams are inevitable elements in teaching, especially in Turkish educational system (Uzun, 2016), other studies point out that crowded classrooms and short amount of time (one semester) allocated for LTA courses prevent pre-service teachers from productive course practices (Hatipoğlu, 2015).

Within the light of the background information about ELT programmes in Turkey, there are certain criteria to be eligible for being employed at an English Preparatory Programme. These criteria are also identified by YÖK regulations. An English instructor must get at least 70 points from ALES (Academic Personnel and Postgraduate Education Entrance Exam) and at least 80 points from YDS (Foreign

Language Exam) or an equivalent point from an exam accepted by YÖK. Apart from these standards, universities may demand further qualifications from candidate instructors such as CV, diploma, academic report, military service report (for male candidates), or a report indicating previous teaching experience. Most importantly, candidate instructors are required to hold at least a bachelor's degree in ELT, English Language and Literature, American Culture and Literature, Translation and Interpreting, or Linguistics. Therefore, it is possible to come across many English language instructors working at English Preparatory Programmes who have not received any pre-service education in LTA.

English Preparatory Programmes at Turkish universities provide one year of intensive English course to students who want to study at universities or departments where the medium of instruction is English. Students are placed in levels according to their proficiency or placement test scores at the beginning of an academic year. Some programmes are modular based while others may be semester based. Students receive English language education and take numerous quizzes and exams throughout the year. Those who have acquired the necessary language skills can continue their education in their faculties. Students can study at a preparatory programme a maximum of two years. If they still cannot pass the preparatory programme by the end of the second year, they either continue their education at a department where the medium of instruction is Turkish or they are dismissed from university.

**3.4.2 Participants.** The sample regarding the questionnaire in this study comprised of 246 English language instructors who were working at English Preparatory Programmes of Turkish universities during the 2017 - 2018 Academic Year.

The details about the participants are presented in Table 2. Apart from the 3 participants who preferred to keep their institution names confidential, the 243 participants were from 50 different universities. The majority of these universities were located in Istanbul (66%). Moreover, while nearly 15% of the universities were located in Central Anatolia mainly in Ankara, almost 10% were located in western Anatolia, and the rest (9%) were located in other regions of Turkey. The majority of

the participants (66.7%) were working at a foundation university while the rest 33.3(%) were working at a state university. The names of the universities are not mentioned due to ethical reasons and to provide confidentiality.

Table 2

*Overview of the Participants*

Category	Details	N	%
Age	20 – 25	10	4.1
	26 – 30	94	38.2
	31 – 35	78	31.7
	36 – 40	22	8.9
	41 – 45	22	8.9
	46 – 50	9	3.7
	51 or above	11	4.5
Gender	Male	71	28.9
	Female	175	71.1
Major at university	ELT	125	50.8
	Non-ELT	121	49.2
Highest qualification	B.A.	104	42.3
	M.A.	128	52.2
	Ph.D.	14	5.7
Teaching experience	1 – 5	61	24.8
	6 – 10	118	48
	11 – 15	26	10.6
	16 – 20	22	8.9
	21 or more	19	7.7
Where they teach	State University	82	33.3
	Foundation University	164	66.7
Extra job responsibility in Testing Office	Yes	28	11.4
	No	218	88.6
Separate testing course in pre-service education	Yes	117	47.6
	No	129	52.4
Any previous training on testing and assessment	Yes	131	53.3
	No	115	46.7
TOTAL		246	100

*Note:* N: Number of participants, %: Percentage of participants

Of all the participants, 175 were female (71.1%), 71 were male (28.9%). Age range of the participants was between 20 to 25 and 51 or above. The largest two age groups were 26 to 30 (38.2%) and 31 to 35 (31.7%). The participants whose ages ranged from 51 or above, 20 to 25, and 46 to 50 were the smallest groups with 4.5%, 4.1%, and 3.7% respectively. 125 of the participants graduated from an ELT

department (50.8%) whereas 121 had a non-ELT degree (49.2%). While 104 participants held a B.A. degree (42.3%), 128 held an M.A. (52.2%) and 14 held a Ph.D. degree (5.7%). 164 participants worked at a foundation university (66.7%) whereas 82 worked at a state university (33.3%). 28 of the participants (11.4%) worked at the Testing Office of an English Preparatory Programme. Regarding teaching experience, the majority of the participants (75%) had more than 6 years of experience. Nearly half of the participants (118) had 6 to 10 years of experience and less than 30% had more than 10 years of experience in teaching. For instance, only 22 participants (8.9%) had 16 to 20 years of experience and 19 (7.7%) had 21 years or more experience. On the other hand, almost a quarter of the participants (61) had 1 to 5 years of experience.

The sample regarding the semi-structured interviews comprised of 12 English language instructors who were also working at English Preparatory Programmes of Turkish universities during the 2017 - 2018 Academic Year. Of these interviewees 10 were female and 2 were male. The age of the interviewees ranged from 27 to 51. However, 9 of the interviewees were aged between 27 and 33 while the other 3 were 38, 44, and 51 years old. Regarding teaching experience, 9 of the interviewees had 6 to 10 years of experience. The other 3 had a relatively more teaching experience ranging from 14 to 22 years. The average of interviewees' teaching experience was 10 years. Unlike 5 of the interviewees who graduated from an ELT department, 7 had a non-ELT degree. Moreover, half of the interviewees held a BA degree while the other half had an MA degree.

### **3.5 Procedures**

**3.5.1 Sampling.** A sampling population is a reachable group of people sharing the same aspects of a target population (Weathington, Cunningham, & Pittenger, 2010). It is of critical importance that the sample in a study consists of a representative portion of the target population so that judicious generalizations may be drawn from the sample used in the study.

The procedure of sampling is divided into two main subsections as probability sampling and nonprobability sampling. Probability sampling is defined as a method of sample selection that uses a system of random selection. In other words, each one



of the members in a population possesses a logical or recognized probability of being selected for the sample. Nonprobability sampling, on the other hand, refers to choosing the participants of a sample without random procedures. That is, all the individuals in a population do not have equal probability of being chosen. Convenience sampling is one of the nonprobability sampling types. It is associated with making use of available samples of the target population. In convenience sampling, the samples who are easily accessible for the researcher are selected. As it is extremely demanding to gather data from the whole population, utilizing convenience sampling method tends to be a practical and time saving way of sampling for the researchers (Ary et al., 2010). Therefore, this study adopted convenience sampling. In this study, the target population is English language instructors who work at English Preparatory Programmes at Turkish universities. As for the sampling, both the questionnaire and the interview were conducted with those who were available and fit to the purpose of the study without any randomizing.

**3.5.2 Data Collection Instruments.** A questionnaire and a semi-structured interview were used to collect the data in this study. Below are the details of both instruments.

**3.5.2.1 The questionnaire.** The questionnaire (Appendix A), which represented the quantitative part of the data, comprised two main parts:

- Part I. Background Information, and
- Part II. Questions about Training in LTA, which consisted of three subsections;
  - A. Classroom-focused LTA
  - B. Purposes of testing
  - C. Content and concepts of LTA

The first part of the questionnaire consisted of not only demographic questions such as age (Question 1), gender (Question 2), major (Question 3), highest qualification received (Question 4), teaching experience (Question 5), but also questions related to testing and assessment such as whether the respondents had a separate testing/assessment course in pre-service education (Question 8), or they had attended any professional development training on language testing and assessment (Question 9).

The first subsection in the second part of the questionnaire constituted domains related to classroom-focused LTA. These domains were:

- preparing classroom tests,
- using ready-made tests from textbook packages or from other sources,
- giving feedback to students based on information from tests/assessment,
- using self- or peer-assessment, using informal, continuous, non-test type of assessment, and using the European Language Portfolio, an adaptation of it or some other portfolio.

The second subsection in the second part of the questionnaire included domains related to purposes of testing. These domains were:

- giving grades,
- finding out what needs to be taught/learnt,
- placing students onto courses, programmes, etc. and
- awarding final certificates.

The last subsection in the second part of the questionnaire consisted of domains related to content and concepts of LTA. These domains were:

- testing/assessing receptive skills, productive skills, micro linguistic aspects, integrated language skills, and aspects of culture,
- establishing validity and reliability of tests/assessment, and
- using statistics to study the quality of tests/assessment.

For each subsection in Part II, the respondents were first asked to select the "Not at all", "Poor", "Average", "Good", or "Advanced" option for the training they received in the mentioned domain, then select the "Strongly Disagree", "Disagree", "Undecided", "Agree", or "Strongly Agree" option for the training they needed in the same domains respectively.

As stated earlier, the questionnaire used in this study was designed by Hasselgreen et al. (2004) and then used by Vogt and Tzagari (2014) following slight changes. The questionnaire was shortened to be available for paper-and-pencil format. Besides, the part pertinent to aspects of large-scale testing was omitted, as these domains are not among teachers' daily LTA tasks.

The original version used the terms *none*, *basic*, and *advanced* for the quantification of the training received, and *basic*, *a little*, and *advanced* for the

training needed regarding each domain. These quantifications could cause undesired and various interpretations by the respondents (Vogt & Tsigari, 2014, p. 379). Therefore, the quantification in the questionnaire was slightly altered as described above to reach more quantifiable data that all respondents could agree on.

**3.5.2.2 The semi-structured interview.** Interviews provide researchers with large amount of thorough data and perceptions of interviewees' viewpoints. Moreover, clarification on interviewees' responses may be achieved through instant additional questions (Ary et al., 2010). Therefore, semi-structured interviews were conducted to provide additional data for the study. The questions for the semi-structured interviews were also adapted from Vogt and Tsigari (2014) and constituted the qualitative part of the data. The semi-structured interviews were conducted with 12 voluntary English language instructors to examine their training backgrounds and their current practices in LTA, how they made up for the skills they felt incompetent in LTA, and their individual in-service training needs in LTA. The interviewees were also invited to any additional comments if they had any.

### **3.5.3 Data Collection Procedures.**

**3.5.3.1 The questionnaire.** Web-based surveys provide researchers with such advantages as reaching large number of participants, and thus, large amount of data easily and quickly without any cost (Ary et al., 2010). Therefore, an online version of the questionnaire was formed on *Google Forms* so as to reach a large number of respondents anonymously. Then, three steps were followed to collect the data for the questionnaire. First, the questionnaire link was shared in an email group which was formed for the directors of School of Foreign Languages at Turkish universities. The directors were sent an email with the questionnaire link and kindly requested to share it with the English language instructors in their institutions. Second, the websites of all universities in Turkey were analysed to reach the email addresses of the directors or head of English Preparatory Programmes. Of all the websites investigated, 61 email addresses of the administrative staff were identified. An email requesting to share the questionnaire link with the English language instructors in their institutions was sent to all the identified email addresses. Third, the researcher's network of colleagues and friends was utilized to share the questionnaire link.

All the quantitative data were collected between February 20 and March 16, 2018. A total of 251 respondents filled in the questionnaire. However, 5 of the responses were excluded from the data as they had missing information. The 246 responses to the questionnaire were analysed via IBM SPSS statistical programme. All the participants were informed that their responses would be kept confidential, and would only be used for research purposes.

**3.5.3.2 The semi-structured interviews.** The interviews were carried out in March and April, 2018. 12 English language instructors were kindly requested to take part in the interviews. 2 of the interviewees participated in e-mail interviews and the rest were interviewed at a time and location of their choice (See Appendix B for the consent form). Face-to-face interviews were recorded and took around 15 – 20 minutes. All the interviews were held in English.

**3.5.4 Data Analysis Procedures.** The quantitative data analysis was carried out through IBM SPSS. All the item values in the questionnaire were given numbers. The enumerated data were then, entered into SPPSS programme. Mean scores were utilized to analyse the level of training received and needed by English language instructors in areas of LTA (Research Question 1). To analyse if there is a statistically significant difference in English language instructors' received and needed training in LTA regarding their major, highest degree in education and LTA responsibilities in testing units (Research Questions 2, 3, 4 and 5), the participants were divided into subgroups.

As for major (Research Question 2), the participants were divided into two subgroups as those who studied ELT as a major and those who did not (non-ELT). As for highest degree in education (Research Question 3), all the participants were also divided into another two subgroups as those who held a BA degree and those who held an MA or PhD degree. Regarding LTA responsibilities in testing units (Research Question 4), the participants were grouped as those who were a member of Testing and Assessment Unit (language testers) and those who were not (regular instructors). Then, the mean scores of the subgroups were compared in twos to determine if they were different from each other. Independent samples t-tests were implemented to investigate if the differences between the subgroups were statistically different or not.

To analyse if there is a statistically significant difference in English language instructors' received and needed training in LTA regarding their teaching experience (Research Question 5), the participants were divided into three subgroups as those with a teaching experience of 1 to 5 years, 6 to 10 years, and those with a teaching experience of 11 or more years. One-way ANOVA tests were implemented to investigate if the differences between the subgroups were statistically different or not.

An independent samples t-test and a one-way ANOVA test are parametric tests which require the collected data to be parametric as well. One assumption of these tests is the normality requirement. That is, the two population samples possess an approximately normally distributed data. In this study, the normality requirement was analysed through skewness (lack of symmetry) and kurtosis (pointiness) values. The values of skewness and kurtosis were converted to z-scores by dividing the values by their standard error. Field (2009) stated that "a z-score is a score from a distribution which has a mean of 0 and a standard deviation of 1" (p. 138). When the z-score is less than 1.96, the data is normally distributed ( $p < .05$ ). When the z-score is greater than 1.96, the data is not normally distributed ( $p < .05$ ). Therefore, when the requirements of normality were not met, the Mann-Whitney  $U$  test and Kruskal-Wallis Test, which are the nonparametric counterparts of the independent samples t-test and one-way ANOVA respectively, were implemented (Field, 2009).

Another assumption of an independent samples t-test and one-way ANOVA test is the homogeneity of variance. In other words, the variances should be the same throughout the data. In this study, the homogeneity of variance requirement was examined by a Levene's test. Regarding independent samples t-test, if the Levene's test was not significant ( $p > .05$ ), the assumption of homogeneity of variances was assumed and if the Levene's test was significant ( $p < .05$ ), the assumption of homogeneity of variances was not assumed (the assumption is violated) (Field, 2009). Regarding one-way ANOVA test, if the Levene's test was not significant ( $p > .05$ ), the assumption of homogeneity of variances was assumed and post hoc comparisons using the Tukey HSD test was implemented. If the Levene's test was significant ( $p < .05$ ), the assumption of homogeneity of variances was not assumed (the assumption is violated) and post hoc comparisons using the Games-Howell test was implemented (Field, 2009).

Regarding the qualitative part of the study, the semi-structured interviews were conducted to have a deeper understanding of English language instructors' individual training needs in LTA. 12 instructors participated in the interviews. All the interviews were conducted in English. While 10 of instructors were interviewed face-to-face, 2 were interviewed through e-mail interviews. E-mail interviewees were kindly requested to respond the questions as detailed as possible. The audio-recorded face-to-face interviews were partially transcribed and together with the e-mail interviews, all the qualitative data were analysed through content analysis in order to categorise the themes and patterns emerging from the data.

**3.5.5 Reliability.** The reliability of the questionnaire was tested using Cronbach's alpha in IBM SPSS. Cronbach's alpha for the entire questionnaire was measured .82. The reliability measures of each subsection differed between .81 and .93 according to Cronbach's alpha. The details for all the subsections are presented in Table 3.

Table 3

*Cronbach's Alpha results for the Subsections of the questionnaire*

	Training received	Training needed
Classroom Focused LTA	.85	.87
Purposes of Testing	.81	.86
Content and concepts of LTA	.93	.93

**3.5.6 Trustworthiness.** Credibility, transferability, dependability, and confirmability are four criteria to be considered in order to establish a trustworthy qualitative research study (Ary et al., 2010). The researcher took the following precautions for each one of the criteria to provide trustworthiness for the study.

Credibility refers to the truthfulness and accuracy of the findings from the qualitative data. Member checks strategy was utilised to meet this criterion. Member checking strategy is conducted to prove that what is said by the interviewee matches with how it is interpreted by the researcher (Ary et al., 2010). Hence, interviewees were requested to review the researcher's notes and transcripts of dialogues for precision and meaning.

Transferability concerns the generalizability of the findings. It is the researcher's duty to present correct and in-depth portrayals of the context and

interviewees so that the reader may determine transferability. Therefore, thick description was provided through comprehensive background information about the interviewees (Ary et al., 2010).

As for dependability, which refers to reliability, the whole procedure throughout the study was reported in detail, that is, the research design and the data collection procedures were explained. Furthermore, triangulation method was adapted. Both quantitative and qualitative methods were implemented to increase the reliability of the study.

Confirmability refers to the objectivity of the researcher. After the analysis and coding of the emerging trends from the qualitative data, an external evaluator, who is a colleague of the researcher, took part in the analysis of the study to conduct peer review. The external evaluator was provided with the raw data from the semi-structured interviews along with the researcher's interpretation of the findings. Then, the interpretation of the findings were discussed with the external evaluator to decide that the interpretation of the findings were rational within the light of the raw data and to check confirmability.

### **3.6 Limitations**

It must be pointed out that this study carries some limitations. First, the data for the questionnaire were collected from 246 English language instructors who voluntarily participated in the study. 10 out of 246 English language instructors also volunteered to take part in the semi-structured interviews. Since the sampling is voluntary and self-selecting, the participants may be among those instructors who are already interested in professional development pertaining to testing and assessment practices, and may believe that training in LTA is very important. Therefore it must be noted that, the participants of the study may have attended more training, or may be more inclined to perceive a need for training than average instructors. Second, the institutions where the participants were employed were not mentioned in this study to provide confidentiality and for ethical reasons.

## Chapter 4

### Results

#### 4.1 Overview

This chapter explains the findings from the questionnaire (quantitative data) and the semi-structured interviews (qualitative data), which were implemented to answer the research questions of the study. The first research question aimed to find out the level of training English language instructors received in areas of LTA and their perceived training needs for in-service training in these areas. The third, fourth, and fifth research questions, on the other hand, investigated if there is a statistically significant difference in English language instructors' reported LTA training level and perceived training needs regarding their major, highest degree in education, LTA responsibilities in testing units, and teaching experience respectively.

#### 4.2 RQ 1: Reported Training Level and Perceived Training Needs in LTA

The first research question investigated the level of training English language instructors reported in areas of LTA and to what extent they perceived a need for in-service training in these areas. To explore this question, the questionnaire findings were analysed and descriptive statistics (mean scores) were calculated regarding the reported training level and perceived need for training in LTA areas.

Table 4

##### *Previous training in LTA*

Question	Response	N	%
Separate testing course in pre-service education	Yes	117	47.6
	No	129	52.4
Any previous training on testing and assessment	Yes	131	53.3
	No	115	46.7

*Note:* N: Number of participants, %: Percentage of participants

First of all, the findings indicated that slightly less than half of the participants (117) had a separate testing and assessment course in pre-service education while 52.4% (129) did not. Moreover, it was also found out that slightly more than half of the participants (131) had participated in some training in LTA whereas 46.7% (115)



of the participants had not taken part in any assessment related professional development training (Table 4).

Table 5

*Mean Scores of Reported Training Level and Perceived Training Needs in LTA*

	Training Received (n=246)		Training Needed (n=246)	
	M	SD	M	SD
<b>CLASSROOM-FOCUSED LTA</b>				
Preparing classroom tests	3.08	1.19	3.47	1.01
Ready-made tests	3.31	1.22	2.94	1.14
Giving feedback based on assessment	3.66	1.05	3.30	1.14
Self-/Peer- assessment	3.44	1.07	3.40	1.10
Informal assessment	3.28	1.16	3.42	1.09
ELP or Portfolio	3.07	1.35	3.52	1.07
<b>PURPOSES OF TESTING</b>				
Giving grades	3.54	1.10	2.89	1.13
Finding out what needs to be taught/learnt	3.89	.95	2.97	1.16
Placing students	3.20	1.24	3.26	1.13
Awarding certificates	2.67	1.22	3.40	1.07
<b>CONTENT AND CONCEPTS OF LTA</b>				
Assessing receptive skills	3.59	1.10	3.15	1.15
Assessing productive skills	3.62	1.07	3.23	1.17
Assessing micro linguistic aspects	3.67	1.04	3.11	1.14
Assessing integrated language skills	3.46	1.10	3.37	1.10
Assessing aspects of culture	2.79	1.24	3.68	1.02
Establishing reliability	3.16	1.17	3.57	1.12
Establishing validity	3.15	1.17	3.59	1.12
Using statistics	2.81	1.18	3.68	1.04

*Note:* n: Number of participants, M: Mean, SD: Standard deviation

As for the reported training level in LTA, the findings showed that the domains which the participants thought to be trained the most were *finding out what needs to be taught/learnt* ( $M=3.89$ ,  $SD=.95$ ), *assessing micro linguistic aspects* ( $M=3.67$ ,  $SD=1.04$ ), *giving feedback to students based on information from tests and assessment* ( $M=3.66$ ,  $SD=1.05$ ), *assessing productive skills* ( $M=3.62$ ,  $SD=1.04$ ), *assessing receptive skills* ( $M=3.59$ ,  $SD=1.10$ ) and *giving grades* ( $M=3.54$ ,  $SD=1.10$ ). On the other hand, the participants perceived to be trained the least in *awarding certificates* ( $M=2.67$ ,  $SD=1.22$ ), *assessing aspects of culture* ( $M=2.79$ ,  $SD=1.24$ ), *using statistics* ( $M=2.81$ ,  $SD=1.18$ ), *using the European Language Portfolio, an adaptation of it or some other portfolio* ( $M=3.07$ ,  $SD=1.35$ ), *preparing classroom tests* ( $M=3.08$ ,  $SD=1.19$ ), *establishing validity* ( $M=3.15$ ,  $SD=1.17$ ), and *establishing reliability* ( $M=3.16$ ,  $SD=1.17$ ) (Table 5).

Regarding the perceived need for training, the participants perceived the highest training need for *assessing aspects of culture* ( $M=3.68$ ,  $SD=1.02$ ), *using statistics* ( $M=3.68$ ,  $SD=1.04$ ), *establishing reliability* ( $M=3.57$ ,  $SD=1.12$ ), *establishing validity* ( $M=3.59$ ,  $SD=1.12$ ), *using the European Language Portfolio, an adaptation of it or some other portfolio* ( $M=3.52$ ,  $SD=1.07$ ), *preparing classroom tests* ( $M=3.47$ ,  $SD=1.01$ ). However, the results illustrated that the domains that the participants were the least interested in for training were *giving grades* ( $M=2.89$ ,  $SD=1.13$ ), *using ready-made tests from textbook packages or from other sources* ( $M=2.94$ ,  $SD=1.13$ ), *finding out what needs to be taught/learnt* ( $M=2.97$ ,  $SD=1.16$ ), *assessing micro linguistic aspects* ( $M=3.11$ ,  $SD=1.14$ ), and *assessing receptive skills* ( $M=3.15$ ,  $SD=1.15$ ) (Table 5).

Table 6

*Mean Scores of Average Reported Training Level and Perceived Training Needs in LTA*

	Training Received (n=246)		Training Needed (n=246)	
	M	SD	M	SD
CLASSROOM-FOCUSED LTA	3.30	.89	3.34	.86
PURPOSES OF TESTING	3.32	.91	3.13	.94
CONTENT AND CONCEPTS OF LTA	3.28	.94	3.42	.92

*Note:* n: Number of participants, M: Mean, SD: Standard deviation

On average, it was found out that the participants perceived to be trained in purposes of testing the most, and content and concepts of LTA the least. They wished to receive training in content and concepts of LTA the most, and purposes of testing the least (Table 6).

#### **4.3 RQ 2: Is There a Statistically Significant Difference in English language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their Major?**

To examine if there is a significant difference in English language instructors' reported LTA training and perceived training needs regarding their major, the participants were divided into two subgroups as those who studied ELT as a major and those who did not study ELT (non-ELT) as a major. Then, the training received and needed in classroom-focused LTA, purposes of testing, and content and concepts of LTA by these subgroups were compared. Table 7 shows the means and standard

deviations of training received and needed by English language instructors who studied ELT and who did not study ELT as a major at university.

Table 7

*Means and Standard Deviations of Training Received and Needed by English Language Instructors Who Studied ELT (n=125) and did not Study ELT (non-ELT) (n=121)*

Subsections	Subgroups			
	ELT		non-ELT	
	M	SD	M	SD
Training received in classroom focused LTA	3.51	.74	3.08	.99
Training needed in classroom focused LTA	3.27	.86	3.41	.85
Training received Purposes of Testing	3.49	.76	3.14	1.01
Training needed in Purposes of Testing	3.11	.89	3.15	.99
Training received in Content and Concepts of LTA	3.47	.81	3.08	1.02
Training needed in Content and Concepts of LTA	3.38	.92	3.46	.93

*Note:* M: Mean, SD: Standard deviation

As Table 7 indicates, the mean scores of received training variables by those who studied ELT are higher. On the other hand, the mean scores of training needed variables by those who did not study ELT are higher.

Either an independent samples t-test or a Mann-Whitney *U* test was conducted to determine if these differences between two subgroups were statistically significant or not depending on the normality and homogeneity of variance requirements of the collected data. As mentioned in Data Analysis Procedures section, the normality requirement was analysed through skewness and kurtosis values. The values of skewness and kurtosis were converted to z-scores by dividing the values by their standard error. When the z-score was less than 1.96, the data was normally distributed ( $p < .05$ ). When the z-score was higher than 1.96, the data was not normally distributed ( $p < .05$ ) and requirements of the t-test were not met (normality was not assumed). In such cases Mann-Whitney *U* test, the nonparametric counterpart of the t-test, was carried out.

The z-scores for skewness and kurtosis values of the training received in classroom-focused LTA, purposes of testing, content and concepts of LTA, and

training needed in content and concepts of LTA variables were found to be higher than 1.96 ( $p < .05$ ) (Tables 8). Therefore, a Mann-Whitney  $U$  test was conducted for these variables. However, as for the training needed in classroom-focused LTA and purposes of testing variables the z-scores for skewness and kurtosis values were found to be lower than 1.96 ( $p < .05$ ). Therefore, an independent samples t-test was conducted for these variables.

Table 8

*Skewness, Kurtosis, and Z-score Values of Distribution*

Subsections	Skewness	Standard Error	Z-score	Kurtosis	Standard Error	Z-score
Training received in classroom focused LTA	-.69	.15	-4.45	.02	.30	.64
Training needed in classroom focused LTA	-.28	.15	-1.83	-.13	.30	-.44
Training received in purposes of testing	-.54	.15	-3.50	.08	.30	.27
Training needed in purposes of testing	-.14	.15	-.94	-.63	.30	-2.05
Training received in content and concepts of LTA	-.56	.15	-3.67	-.11	.30	-.36
Training needed in content and concepts of LTA	-.55	.15	-3.6	-.11	.30	-.36

Moreover, homogeneity of the variances was checked via Levene's test. When Levene's test was not significant ( $p >.05$ ), equal variances were assumed and when the Levene's test was significant ( $p <.05$ ), equal variances weren't assumed (the assumption was violated).

Table 9

*Independent Samples T-test Results for the Difference in Perceived Training Needs of English Language Instructors who Studied ELT (n=125) and Those Who did not Study ELT (non-ELT) (n=121)*

		Independent Samples Test				
		Levene's Test for Equality of Variances		t-test for Equality of Means		
		<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig. (2-tailed)</i>
Training needed in classroom focused LTA	Equal variances assumed	.152	.697	-1.239	244	.217
	Equal variances not assumed			-1.239	243.820	.217
Training needed in Purposes of Testing	Equal variances assumed	1.919	.167	-.304	244	.761
	Equal variances not assumed			-.304	239.447	.762

The skewness and kurtosis values for the training needed in classroom focused LTA variable were  $-.285$  ( $SE=.155$ ) and  $-.136$  ( $SE=.309$ ) respectively. The z-score for skewness value was found to be  $-.285/.155 = -1.83$  ( $p < 0.05$ ) and kurtosis value to be  $-.136/.309 = -.44$  ( $p < 0.05$ ) suggesting a normal distribution. A Levene's test found that the assumption of homogeneity of variance was met,  $p = .697$ ; therefore a two-tailed independent samples t-test based on equal variances was carried out to compare the ELT subgroup and non-ELT subgroups' training needs in classroom focused LTA variable. The test suggested no statistically significant difference between the scores for ELT subgroup's ( $M=3.27$ ,  $SD=.86$ ) and non-ELT subgroup's ( $M=3.41$ ,  $SD=.85$ ) training needs;  $t(244) = -1.239$ ,  $p = .217$ ,  $d = -0.16$  (Table 9).

The skewness and kurtosis values for the training needed in purposes of testing variable were  $-.146$  ( $SE=.155$ ) and  $-.635$  ( $SE=.309$ ) respectively. The z-score for skewness value was found to be  $-.146/.155 = -.94$  ( $p < 0.05$ ) and kurtosis value to be  $-.635/.309 = -2.05$  ( $p < 0.05$ ) suggesting an almost normal distribution. A Levene's test found that the assumption of homogeneity of variance was met,  $p = .167$ ; therefore a two-tailed independent samples t-test based on equal variances was carried out. The

independent samples t-test conducted to compare the ELT subgroup and non-ELT subgroups' training needs in purposes of testing suggested no statistically significant difference between the scores for ELT subgroup's ( $M=3.11$ ,  $SD=.89$ ) and non-ELT subgroup's ( $M=3.15$ ,  $SD=.99$ ) training needs;  $t(244) = -.304$ ,  $p=.761$ ,  $d = -0.04$  (Table 9).

A Mann-Whitney  $U$  test results for the training needed in content and concepts of LTA variable suggested that there was no statistically significant difference between the scores for ELT subgroup ( $Mdn=3.50$ ) and non-ELT subgroup ( $Mdn=3.75$ ) in terms of the training they needed in content and concepts of LTA ( $U=7098.000$ ,  $z = -836$ ,  $p=.403$ ,  $d = -0.08$ ).

Table 10

*The Results of Mann Whitney-U Test with Respect to Training Received in Classroom-focused LTA*

Variables	n	Rank Average	Rank Total	U	p
ELT	125	136.96	17120.00	5880.000	.003
Non-ELT	121	109.60	13261.00		

A Mann-Whitney  $U$  test result for the training received in classroom-focused LTA variable indicated that there was a statistically significant difference between the scores for ELT subgroup's ( $Mdn=3.66$ ) and non-ELT subgroup's ( $Mdn=3.33$ ) training needs ( $U=5880.000$ ,  $z = -3.023$ ,  $p=.003$ ,  $d = 0.49$ ) (Table 10).

Table 11

*The Results of Mann Whitney-U Test with Respect to Training Received in Purposes of Testing*

Variables	n	Rank Average	Rank Total	U	p
ELT	125	135.20	16900.50	6099.500	.008
Non-ELT	121	111.41	13480.50		

A Mann-Whitney  $U$  test results for the training received in purposes of testing variable suggested that there was a statistically significant difference between the scores for ELT subgroup ( $Mdn=3.50$ ) and non-ELT subgroup ( $Mdn=3.25$ ) in terms of the training they received in purposes of testing ( $U=6099.500$ ,  $z = -2.633$ ,  $p=.008$ ,  $d = 0.39$ ) (Table 11).

Table 12

*The Results of Mann Whitney-U Test with Respect to Training Received in Content and Concepts of LTA*

Variables	n	Rank Average	Rank Total	U	p
ELT	125	135.79	16974.00	6026.000	.006
Non-ELT	121	110.80	13407.00		

A Mann-Whitney  $U$  test results for the training received in content and concepts of LTA variable suggested that there was a statistically significant difference between the scores for ELT subgroup ( $Mdn=3.62$ ) and non-ELT subgroup ( $Mdn=3.25$ ) in terms of the training they received in content and concepts of LTA ( $U=6026.000$ ,  $z = -2.758$ ,  $p=.006$ ,  $d = 0.42$ ) (Table 12).

To sum up, while the mean scores of the reported training level by those who studied ELT as a major in classroom-focused LTA, purposes of testing, and content and concepts of LTA were significantly higher than those who did not study ELT as a major, no significant difference was observed between those who studied ELT as a major and those who did not regarding their perceived need for training in all three subsections.

#### **4.4 RQ 3: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their Highest Educational Degree?**

To examine if there is a significant difference in English language instructors' reported LTA training and perceived training needs regarding their highest degree in education, the participants were divided into two subgroups as those whose highest qualification is BA and those whose highest qualification is MA or PhD. Then, the training received and needed in classroom-focused LTA, purposes of testing, and content and concepts of LTA by these subgroups were compared. Table 13 shows the means and standard deviations of training received and needed by English language instructors who hold a BA degree and those who hold an MA or PhD degree. The mean scores of training received by those who hold an MA or PhD degree are higher. On the other hand, the mean scores of training needed by those who hold an MA or PhD degree are higher.

Table 13

*Means and Standard Deviations of Training Received and Needed by English Language Instructors Who Hold a BA Degree (n=104) and Who Hold an MA or PhD Degree (n=142)*

Subsections	Subgroups			
	BA		MA / PhD	
	M	SD	M	SD
Training received in classroom focused LTA	3.26	.94	3.33	.86
Training needed in classroom focused LTA	3.36	.80	3.32	.90
Training received Purposes of Testing	3.28	.97	3.34	.86
Training needed in Purposes of Testing	3.16	.93	3.10	.96
Training received in Content and Concepts of LTA	3.23	.93	3.31	.95
Training needed in Content and Concepts of LTA	3.49	.83	3.37	.98

*Note:* M: Mean, SD: Standard deviation

As mentioned earlier, either an independent samples t-test or a Mann-Whitney U test was conducted to determine if these differences between two subgroups were statistically significant or not depending on the normality and homogeneity of variance requirements of the collected data. The z-scores for skewness and kurtosis values of the training received in classroom focused LTA, purposes of testing, content and concepts of LTA, and training needed in content and concepts of LTA variables were found to be higher than 1.96 ( $p < .05$ ) (Table 8). Therefore, a Mann-Whitney U test was conducted for these variables. However, as for the training needed in classroom focused LTA and purposes of testing variables, the z-scores for skewness and kurtosis values were found to be lower than 1.96 ( $p < .05$ ). Therefore, an independent samples t-test was conducted for these variables.

The skewness and kurtosis values for the training needed in classroom focused LTA variable were  $-.285$  ( $SE=.155$ ) and  $-.136$  ( $SE=.309$ ) respectively. The z-score for skewness value was found to be  $-.285/.155 = -1.83$  ( $p < 0.05$ ) and kurtosis value to be  $-.136/.309 = -.44$  ( $p < 0.05$ ) suggesting a normal distribution. A Levene's test found that the assumption of homogeneity of variance was met,  $p=.186$ ; therefore a two-tailed independent samples t-test based on equal variances was carried out. The independent samples t-test conducted to compare the BA and MA / PhD subgroup training needs in classroom focused LTA suggested that there was no statistically



significant difference between the scores for BA subgroup's ( $M=3.36$ ,  $SD=.80$ ) and MA / PhD subgroup's ( $M=3.32$ ,  $SD=.90$ ) training needs;  $t(244)=.351$ ,  $p=.726$ ,  $d=0.04$  (Table 14).

Table 14

*Independent Samples T-test Results for the Difference between Training Needed by English Language Instructors Who Hold a BA Degree (n=104) and Who Hold an MA or PhD Degree (n=142)*

		Independent Samples Test				
		Levene's Test for Equality of Variances		t-test for Equality of Means		
		<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig. (2-tailed)</i>
Training needed in classroom focused LTA	Equal variances assumed	1.760	.186	.351	244	.726
	Equal variances not assumed			.357	234.195	.721
Training needed in Purposes of Testing	Equal variances assumed	.129	.720	.477	244	.634
	Equal variances not assumed			.480	225.939	.632

The skewness and kurtosis values for the training needed in purposes of testing variable were  $-.146$  ( $SE=.155$ ) and  $-.635$  ( $SE=.309$ ) respectively. The z-score for skewness value was found to be  $-.146/.155 = -.94$  ( $p<0.05$ ) and kurtosis value to be  $-.635/.309 = -2.05$  ( $p<0.05$ ) suggesting an almost normal distribution. A Levene's test found that the assumption of homogeneity of variance was met,  $p=.720$ ; therefore a two-tailed independent samples t-test based on equal variances was conducted. The independent samples t-test carried out to compare the BA and MA / PhD subgroups' training needs in purposes of testing suggested no statistically significant difference between the scores for BA subgroup's ( $M=3.16$ ,  $SD=.93$ ) and MA / PhD subgroup's ( $M=3.10$ ,  $SD=.96$ ) training needs;  $t(244)=.477$ ,  $p=.634$ ,  $d = 0.06$  (Table 14).

A Mann-Whitney  $U$  test results for the training needed in content and concepts of LTA variable suggested that there was no statistically significant difference between the scores for BA subgroup ( $Mdn=3.62$ ) and MA / PhD subgroup ( $Mdn=3.50$ ) in terms of the training they needed in content and concepts of LTA ( $U=7037.500$ ,  $z = -.631$ ,  $p=.528$ ,  $d = 0.13$ ).

A Mann-Whitney  $U$  test result for the training received in classroom focused LTA variable indicated that there was no statistically significant difference between

the scores for BA subgroup's ( $Mdn=3.41$ ) and MA / PhD subgroup's ( $Mdn=3.50$ ) training needs ( $U=7259.000$ ,  $z = -.227$ ,  $p=.820$ ,  $d = -0.07$ ).

A Mann-Whitney  $U$  test results for the training received in purposes of testing variable suggested that there was no statistically significant difference between the scores for BA subgroup ( $Mdn=3.50$ ) and MA / PhD subgroup ( $Mdn=3.50$ ) in terms of the training they received in purposes of testing ( $U=7192,500$ ,  $z = -.349$ ,  $p=.727$ ,  $d=-0.06$ ).

A Mann-Whitney  $U$  test results for the training received in content and concepts of LTA variable suggested that there was no statistically significant difference between the scores for BA subgroup ( $Mdn=3.50$ ) and MA / PhD subgroup ( $Mdn=3.50$ ) in terms of the training they received in content and concepts of LTA ( $U=7121.500$ ,  $z = -.477$ ,  $p=.633$ ,  $d = -0.08$ ).

In short, none of the tests above suggested any significant difference between those with a BA degree and those with an MA / PhD degree in the training they received and needed in classroom-focused LTA, purposes of testing and content and concepts of LTA.

#### **4.5 RQ 4: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their LTA Responsibilities in Testing Units?**

To examine if there is a significant difference in English language instructors' reported LTA training and perceived training needs regarding their LTA responsibilities in testing units, the participants were divided into two subgroups as those who work in the Testing Office (language testers) and those who do not work in the Testing Office (regular instructors). Then, the training received and needed by these two subgroups was compared. Table 15 shows the means and standard deviations of training received and needed by language testers and regular instructors.

As mentioned earlier, either an independent samples t-test or a Mann-Whitney  $U$  test was conducted to determine if these differences between two subgroups were statistically significant or not depending on the normality and homogeneity of

variance requirements of the collected data. The z-scores for skewness and kurtosis values of the training received in classroom focused LTA, purposes of testing, content and concepts of LTA, and training needed in content and concepts of LTA variables were found to be higher than 1.96 ( $p<.05$ ) (Table 8). Therefore, a Mann-Whitney  $U$  test was conducted for these variables. However, as for the training needed in classroom focused LTA and purposes of testing variables, the z-scores for skewness and kurtosis values were found to be lower than 1.96 ( $p<.05$ ). Therefore, an independent samples t-test was conducted for these variables.

Table 15

*Means and Standard Deviations of Training Received and Needed by Language Testers (n=28) and Regular Instructors (n=218)*

Subsections	Subgroups			
	Language Testers		Regular Instructors	
	M	SD	M	SD
Training received in classroom focused LTA	3.31	.93	3.30	.89
Training needed in classroom focused LTA	3.55	.80	3.31	.86
Training received Purposes of Testing	3.30	.95	3.32	.90
Training needed in Purposes of Testing	3.46	.88	3.08	.94
Training received in Content and Concepts of LTA	3.23	1.14	3.28	.91
Training needed in Content and Concepts of LTA	3.59	1.00	3.39	.92

*Note:* M: Mean, SD: Standard deviation

The skewness and kurtosis values for the training needed in classroom focused LTA variable were  $-.285$  ( $SE=.155$ ) and  $-.136$  ( $SE=.309$ ) respectively. The z-score for skewness value was found to be  $-.285/.155 = -1.83$  ( $p<0.05$ ) and kurtosis value to be  $-.136/.309 = -.44$  ( $p<0.05$ ) suggesting a normal distribution. A Levene's test found that the assumption of homogeneity of variance was met,  $p=.703$ ; therefore a two-tailed independent samples t-test based on equal variances was carried out. The independent samples t-test conducted to compare the language testers and regular instructors subgroup training needs in classroom focused LTA suggested there was no statistically significant difference between the scores for language testers subgroup's ( $M=3.55$ ,  $SD=.80$ ) and regular instructors subgroup's ( $M=3.31$ ,  $SD=.86$ ) training needs;  $t(244) = 1.379$ ,  $p=.169$ ,  $d = -0.28$  (Table 16).

The skewness and kurtosis values for the training needed in purposes of testing variable were  $-.146$  ( $SE=.155$ ) and  $-.635$  ( $SE=.309$ ) respectively. The z-score for skewness value was found to be  $-.146/.155 = -.94$  ( $p < 0.05$ ) and kurtosis value to be  $-.635/.309 = -2.05$  ( $p < 0.05$ ) suggesting an almost normal distribution. A Levene's test found that the assumption of homogeneity of variance was met,  $p = .433$ ; therefore a two-tailed independent samples t-test based on equal variances was conducted. The independent samples t-test carried out to compare language testers and regular instructors subgroups' training needs in purposes of testing suggested there was a statistically significant difference between the scores for language testers subgroup's ( $M=3.46$ ,  $SD=.88$ ) and regular instructors subgroup's ( $M=3.08$ ,  $SD=.94$ ) training needs;  $t(244) = 1.982$ ,  $p = .049$ ,  $d = 0.41$  (Table 16).

Table 16

*Independent Samples T-test Results for the Difference between Training Needed by Language Testers (n=28) and Regular Instructors (n=218)*

		Independent Samples Test				
		Levene's Test for Equality of Variances		t-test for Equality of Means		
		<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig. (2-tailed)</i>
Training needed in classroom focused LTA	Equal variances assumed	.146	.703	1.379	244	.169
	Equal variances not assumed			1.456	35.485	.154
Training needed in Purposes of Testing	Equal variances assumed	.616	.433	1.982	244	.049
	Equal variances not assumed			2.094	35.496	.043

A Mann-Whitney  $U$  test results for the training needed in content and concepts of LTA variable suggested that there was no statistically significant difference between the scores for language testers subgroup ( $Mdn=3.93$ ) and regular instructors subgroup ( $Mdn=3.50$ ) in terms of the training they needed in content and concepts of LTA ( $U=2671.000$ ,  $z = -1.080$ ,  $p = .280$ ,  $d = 0.20$ ).

A Mann-Whitney  $U$  test result for the training received in classroom focused LTA variable indicated that there was no statistically significant difference between the scores for language testers subgroup's ( $Mdn=3.58$ ) and regular instructors subgroup's ( $Mdn=3.50$ ) training needs ( $U=2990.000$ ,  $z = -.175$ ,  $p = .861$ ,  $d = 0.01$ ).

A Mann-Whitney  $U$  test results for the training received in purposes of testing variable suggested that there was no statistically significant difference between the scores for language testers subgroup ( $Mdn=3.50$ ) and regular instructors subgroup ( $Mdn=3.50$ ) in terms of the training they received in purposes of testing ( $U=2949.500$ ,  $z = -.290$ ,  $p=.771$ ,  $d = -0.02$ ).

A Mann-Whitney  $U$  test results for the training received in content and concepts of LTA variable suggested that there was no statistically significant difference between the scores for language testers subgroup ( $Mdn=3.62$ ) and regular instructors subgroup ( $Mdn=3.50$ ) in terms of the training they received in content and concepts of LTA ( $U=2966.000$ ,  $z = -.243$ ,  $p=.808$ ,  $d = -0.04$ ).

To summarize the results of the tests mentioned above, except for perceived training need in purposes of testing, where the mean score of language testers was higher than regular instructors, no significant difference was observed between language testers and regular instructors in any of the subsections.

#### **4.6 RQ 5: Is There a Statistically Significant Difference in English Language Instructors' Reported LTA Training and Perceived Training Needs Regarding Their Teaching Experience?**

To examine if there is a significant difference in English language instructors' reported LTA training and perceived training needs regarding their teaching experience, the participants were divided into 3 subgroups as those with a teaching experience of 1 to 5 years, 6 to 10 years, and those with a teaching experience of 11 or more years. Then, the training received and needed by these 3 subgroups was compared conducting a one-way ANOVA test for each variable. Table 17 shows the means and standard deviations of training received and needed by the four subgroups regarding teaching experience.

Either a one-way ANOVA test or a Kruskal-Wallis test was conducted to determine if these differences between four subgroups were statistically significant or not depending on the normality and homogeneity of variance requirements of the collected data. As mentioned in Data Analysis Procedures section, the normality requirement was analysed through skewness and kurtosis values. The values of skewness and kurtosis were converted to z-scores by dividing the values by their

standard error. When the z-score was less than 1.96, the data was normally distributed ( $p < .05$ ). When the z-score was higher than 1.96, the data was not normally distributed ( $p < .05$ ) and requirements of the t-test were not met (normality was not assumed). In such cases a Kruskal-Wallis test, the nonparametric counterpart of the one-way ANOVA test, was carried out.

Table 17

*Means and Standard Deviations of Training Received and Needed by the Three Subgroups Regarding Teaching Experience*

Subsections	Subgroups					
	1-5 years (n=61)		6-10 years (n=118)		11 or more years (n=67)	
	M	SD	M	SD	M	SD
Training received in classroom focused LTA	3.13	.96	3.40	.82	3.29	.94
Training needed in classroom focused LTA	3.43	.82	3.43	.86	3.08	.84
Training received Purposes of Testing	3.07	.90	3.46	.81	3.29	1.02
Training needed in Purposes of Testing	3.27	.87	3.20	.98	2.88	.90
Training received in Content and Concepts of LTA	2.99	1.04	3.33	.88	3.44	.91
Training needed in Content and Concepts of LTA	3.45	.95	3.53	.91	3.18	.88

*Note:* M: Mean, SD: Standard deviation

The z-scores for skewness and kurtosis values of the training received in classroom focused LTA, purposes of testing, content and concepts of LTA, and training needed in content and concepts of LTA variables were found to be higher than 1.96 ( $p < .05$ ) (Table 8). Therefore, a Kruskal-Wallis test was conducted for these variables. However, as for the training needed in classroom focused LTA and purposes of testing variables the z-scores for skewness and kurtosis values were found to be lower than 1.96 ( $p < .05$ ). Therefore, a one-way ANOVA test was conducted for these variables.

A one-way analysis of variance (Table 18) suggested that the amount of teaching experience on the perceived training needs for classroom focused LTA variable was significant,  $F(2, 243) = 4.087, p = .018, \eta^2 = .03$ . A Levene's test found

that the assumption of homogeneity of variance was met,  $p=.515$ ; therefore post hoc comparisons based on equal variances was conducted. Post hoc comparisons using the Tukey HSD test indicated that the mean score for instructors with a teaching experience of 6-10 years ( $M= 3.43$ ,  $SD= .86$ ) was significantly different than the mean score for instructors with a teaching experience of 11 or more years ( $M= 3.08$ ,  $SD= .84$ ). However, no other comparisons reached significance.

Table 18

*One-Way Analysis of Variance of Perceived Training Need for Classroom-focused LTA regarding Teaching Experience*

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Between groups	2	5.907	2.954	4.087	.018
Within groups	243	175.626	.723		
Total	245	181.533			

A one-way analysis of variance (Table 19) suggested that the amount of teaching experience on the perceived training needs for purposes of testing variable was significant,  $F(2, 243) = 3.429$ ,  $p = .034$ ,  $\eta^2 = .02$ . A Levene's test found that the assumption of homogeneity of variance was met,  $p=.331$ ; therefore post hoc comparisons based on equal variances was conducted. Post hoc comparisons using the Tukey HSD test indicated that the mean score for instructors with a teaching experience of 1-5 years ( $M= 3.27$ ,  $SD= .87$ ) was significantly different than the mean score for instructors with a teaching experience of 11 or more years ( $M= 2.88$ ,  $SD= .90$ ). However, no other comparisons reached significance.

Table 19

*One-Way Analysis of Variance of Perceived Training Need for Purposes of Testing regarding Teaching Experience*

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Between groups	2	6.041	3.021	3.429	.034
Within groups	243	214.040	.881		
Total	245	220.081			

A Kruskal-Wallis test showed that there was a statistically significant difference in training needed in content and concepts of LTA between the range of different years of teaching experience  $\chi^2(2) = 6.985$ ,  $p = 0.030$ , with a mean rank

training needed in purposes of testing score of 128.92 for 1-5 years, 131.72 for 6-10 years, and 104.08 for 11 or more years. Follow-up tests were conducted to evaluate pairwise differences among the four subgroups, controlling for Type I error across tests by using the Bonferroni approach. The results of these tests indicated a significant difference between the 6-10 years subgroup and the 11 or more years subgroup,  $p = .033$ .

A Kruskal-Wallis test showed that there was no statistically significant difference in training received in classroom focused LTA between the range of different years of teaching experience  $\chi^2(2) = 3.162$ ,  $p = 0.206$ , with a mean rank training received in classroom focused LTA score of 110.63 for 1-5 years, 130.51 for 6-10 years, and 122.87 for 11 or more years.

A Kruskal-Wallis test showed that there was a statistically significant difference in training received in purposes of testing between the range of different years of teaching experience  $\chi^2(2) = 7.391$ ,  $p = 0.025$ , with a mean rank training received in purposes of testing score of 103.37 for 1-5 years, 133.74 for 6-10 years, and 123.79 for 11 or more years (Table 20). Follow-up tests were conducted to evaluate pairwise differences among the four subgroups, controlling for Type I error across tests by using the Bonferroni approach. The results of these tests indicated a significant difference between the 1-5 years subgroup and the 6-10 years subgroup,  $p = .020$ .

Table 20

*The Results of Kruskal-Wallis Test with Respect to Training Received in Purposes of Testing*

Variables	n	Rank Average	sd	x	p	
Training received in purposes of testing	1-5 years	61	103.37	2	7.391	0.025
	6-10 years	118	133.74			
	11 or more years	67	123.79			

A Kruskal-Wallis test showed that there was a statistically significant difference in training received in content and concepts of LTA between the range of different years of teaching experience  $\chi^2(2) = 6.753$ ,  $p = 0.034$ , with a mean rank



training received in content and concepts of LTA score of 104.43 for 1-5 years, 126.03 for 6-10 years, and 136.41 for 11 or more years (Table 21). Follow-up tests were conducted to evaluate pairwise differences among the four subgroups, controlling for Type I error across tests by using the Bonferroni approach. The results of these tests indicated a significant difference between the 1-5 years subgroup and the 11 or more years subgroup,  $p = .033$ .

Table 21

*The Results of Kruskal-Wallis Test with Respect to Training Received in Content and Concepts of LTA*

Variables		n	Rank Average	sd	x	p
Training received in content and concepts of LTA	1-5 years	61	104.43	2	6.753	0.034
	6-10 years	118	126.03			
	11 or more years	67	136.41			

In short, the perceived training need for classroom-focused LTA by those with a teaching experience of 6 to 10 years was significantly higher than those with a teaching experience of 11 or more years. The perceived training need for purposes of testing by those with a teaching experience of 1 to 5 years was significantly higher than those with a teaching experience of 11 or more years. The perceived training need in content and concepts of LTA by those with a teaching experience of 6 to 10 years was significantly higher than those with a teaching experience of 11 or more years. On the other hand, the reported training level in purposes of testing by those with a teaching experience of 6 to 10 years was significantly higher than those with a teaching experience of 1 to 5 years. The reported training level in content and concepts of LTA by those with a teaching experience of 11 or more years was significantly higher than those with a teaching experience of 1 to 5 years.

#### **4.7 RQ 6: What are English Language Instructors' Perceptions of LTA Training in Pre-service and In-service Education?**

The findings from the semi-structured interviews, conducted with 12 voluntary English language instructors, represented the qualitative part of the study (See

Methodology chapter for detailed information). The questions in Appendix B were used as a guideline to conduct the interviews. The aim of the interviews were to investigate English language instructors' personalized perceptions on their LTA training and training needs and how they have dealt with any lack of training in their LTA practices.

The findings indicated that of the twelve interviewees, only seven received pre-service training in LTA whereas five had had no training at all when they started teaching.

Regarding LTA training in pre-service education, only two interviewees felt positive about their pre-service training. However, these two had some concerns about not having enough practical experience:

*Well, at that time I thought that I was ready. However now, I think I need some kind of training to remember stuff. It's been a long time and I haven't done testing for 3 - 4 years.* (Instructor 12, face to face interview, April 18, 2018)

*Yes, I did. I felt appropriate. However, when I started teaching, I felt that I had enough theoretical knowledge but I lacked experience.* (Instructor 1, face to face interview, March 22, 2018)

However, a majority of the interviewees pointed out that they did not feel appropriately ready for their LTA activities and some were very strong about it:

*Although we were assigned to do some presentations and encouraged to comment on any in-class tasks about testing and evaluation at (...) University, it was never enough, to be sure.* (Instructor 3, e-mail interview, March 29, 2018)

*Absolutely not. Although I got involved in many TA tasks which were satisfactory in terms of quality and quantity, I could not steep myself in LTA. So, I needed to apply what I know into real practice. However, as I began to teach, my perspective towards LTS has changed and I believe that recent technological developments influenced the methods of LTS, so we need to adjust traditional ways to new trends in education.* (Instructor 4, e-mail interview, March 31, 2018)

Moreover, a senior instructor who studied ELT at university bewailed that they did not have a course on language testing and assessment:

*No, honestly I didn't feel appropriately prepared for my LTA tasks after pre-service training in the 1990s. There weren't even any specific courses to prepare ELT students for their LTA tasks. (...) unfortunately many years passed without recognising the factors affecting learners' performance in language tests.* (Instructor 6, face to face interview, April 5, 2018)

Most of the interviewees learned about LTA on the job, through in-service training provided to them, through self-study or from colleagues:

*On-the-job, through my experiences in teaching. I learned various language tests for different proficiency levels, question types for testing all four language skills as well as vocabulary. Knowing only basic principles in LTA do not ensure you that you can be good at it unless you practice it in real learning setting. So, I still keep learning about LTA.* (Instructor 4, e-mail interview, March 31, 2018)

*I learned LTA while working in the Testing and Assessment Unit at my university. Besides this, I attended several seminars/workshops in this field.* (Instructor 5, face to face interview, April 3, 2018)

*I learned about LTA when I started teaching while assessing my students' performance and with guidance from colleagues.* (Instructor 11, face to face interview, April 17, 2018)

Most interviewees stated to be knowledgeable about more recent LTA methods. Interviewees also pointed out that they used at least some of them in various ways and these methods were mostly used for the assessment of writing skill:

*(...) when I try a method of assessment, I mostly do this in writing after they finish writing their portfolio in class by asking a set of questions to check their writing. I can also do this as peer-check when I make them swap their papers and evaluate them by using the same set of questions.* (Instructor 3, e-mail interview, March 29, 2018)

*(I) also encourage my students to do self- and peer-assessment through a checklist paper I provide them. In addition to that, the assessment of writing skill is based on portfolio assessment in the school I work. So each week I give feedback to a collection of writing papers, thereby have an obvious idea about my students' strengths and weaknesses in writing.* (Instructor 4, e-mail interview, March 31, 2018)

*We have in-class and takeaway portfolio assignments and I try to give my students detailed feedback both about the content and the language. While giving feedback, I consider the criteria on the writing rubric. I also use self- and peer-assessment for the writing skill. In order to guide my students during this process, I give them check-lists.* (Instructor 8, face to face interview, April 12, 2018)

*Peer check is my favourite, but unfortunately not for writing classes that is because students generally are not equipped enough to deal with their errors in writing. However, for simple tasks such as rewriting, fill in the blanks - grammar / vocabulary activities - it is easier to do this. First individual work then pairs then in groups of 3 or 4. Finally a whole class check.* (Instructor 12, face to face interview, April 18, 2018)

Almost all of the interviewees expressed that they felt competent about using the types of LTA conducted in their Preparatory School:

*I feel competent about using the types of LTA applied in my school. They are not varied and not very demanding on my part. (Instructor 11, face to face interview, April 17, 2018)*

*I work at (...) University, and I feel quite competent enough using them since I have been working here for almost (...) years. (Instructor 9, face to face interview, April 13, 2018)*

However, two of the interviewees had some concerns and felt hesitant about their competence to some degree:

*I feel competent about using all kinds of tests. However, sometimes I find it difficult to assess the oral work since there are many other variables like anxiety caused by their speaking partners, teachers assessing them, etc. Therefore, it becomes difficult for assessors to think all of them and decide. (Instructor 1, face to face interview, March 22, 2018)*

*I can't say I am fully equipped with the techniques of LTA so I need to learn more about it to be competent. (Instructor 3, e-mail interview, March 29, 2018)*

Almost all the interviewees expressed that they attended some in-service training in LTA at some point in their career. Most of the trainings mentioned by the interviewees were related to the assessment of productive skills, especially assessing writing:

*I have received in-service trainings on the assessment of writing and speaking at (...) University. They were very satisfying. (Instructor 10, face to face interview, April 16, 2018)*

*I have received some workshops that we evaluated portfolio exams. It was good to see how my colleagues were evaluating a writing paper and it was an opportunity to see which details I missed or I was right doing. (Instructor 2, face to face interview, March 26, 2018)*

*During the orientation weeks (at the beginning of an academic year) and spring breaks, my institution provides in-service training on many areas including LTA. (...) The sessions held by Testing and Assessment Unit usually focus on practical subjects such as the process of speaking exam, quiz marking. Other training sessions include subjects such as giving feedback, types of assessment which meet the students' needs, using rubrics for marking. Most of them are satisfying. (Instructor 8, face to face interview, April 12, 2018)*

*I attended in-service training sessions about LTA. The aim of those sessions was to train the staff to have objective and standardised invigilation and marking procedures. (Instructor 9, face to face interview, April 13, 2018)*

The majority of the interviewees stated that they were satisfied with the in-service training they had been offered until then. However, some had concerns about the content and frequency of these trainings.

*There must be more workshops or training programmes for assessing writing and speaking because standardisation is really important for the sake of learners. (Instructor 2, face to face interview, March 26, 2018)*

*Unfortunately, workshops such as the one I attended are very rare in Turkey. Most workshops and training programmes focus on teaching English, but not testing and assessment. There should be more of them and they should cover every aspect of testing and assessment. There can be separate workshops for each area in assessment (for example, on item analysis, assessing writing, etc.) (Instructor 5, face to face interview, April 3, 2018)*

As for the LTA requested by the interviewees in the short-term and/or in the long-term, although responses from interviewees differed according to their individual needs, there were some emerging trends. Most frequent areas that the interviewees wished to receive in-service training were assessing productive skills and new / alternative assessment methods in classroom assessment:

*I'd like to have training in in-class assessment and the types of productive skills assessment. (Instructor 8, face to face interview, April 12, 2018)*

*I would like to learn more of in-class assessment techniques as well as techniques to prepare questions in particular fields such as vocabulary. (Instructor 7, face to face interview, April 10, 2018)*

*I think it is important for an institution to provide its staff with in-service training about LTA, introduce them with current research about new types of assessment while also adopting alternative assessment methods. (Instructor 11, face to face interview, April 17, 2018)*

Some interviewees emphasized that these trainings should be more practical, functional, and “concentrating on real use” than be theoretical. Moreover, it was also mentioned that these trainings should be “regular, systematic and well-rounded” so that all instructors working at the same institution can have conceptualised and standardised understandings of assessment:

*What I strongly believe is that LTA trainings-like every other training related to ELT- should be functional and goal-oriented. I mean, rather than functioning as a philosophical conundrum, it should be concentrating on the real use. (Instructor 10, face to face interview, April 16, 2018)*

*Even if I know some criteria of assessment, I want them to be exemplified or to be shown in a context where we can identify or define it clearly. For example,*

*if "coherence" is a criterion to assess writing skill, I want somebody to redefine it or to tell simply what it is or how we can evaluate it in a piece of writing so as to make sure that every teacher sees it in the same way. When I assess any portfolio, I want to make sure that the other checker really knows what "coherence" means and is able to recognize it in a text as I am. (Instructor 3, e-mail interview, March 29, 2018)*

*I would desire to learn more about LTA because the present in-service teacher training in LTA is not very comprehensive, probably that is the main reason of facing certain difficulties while properly assessing student progress. On the other hand, as the educational background and teaching experiences of instructors differ, some problems may arise in LTA. So more systematic and well-rounded training in LTA should be provided to all instructors regularly. (Instructor 4, e-mail interview, March 31, 2018)*



## **Chapter 5**

### **Discussion and Conclusions**

#### **5.1 Overview**

This study aimed to explore the reported training level and perceived training needs of English language instructors in LTA. More specifically, the study examined if the reported training level and perceived training needs of English language instructors in LTA showed a statistically significant difference regarding their major, highest educational degree, LTA responsibilities in testing units, and teaching experience.

A mixed methods research design was adopted and both quantitative and qualitative data were gathered and analysed for the purposes of the study. Questionnaires submitted by 246 English language instructors and 12 semi-structured interviews made up the quantitative and qualitative data for the study respectively.

This chapter first discusses the results of the research study in relation to each research question as well as pedagogical implications and recommendations for future research are presented.

#### **5.2 Discussion of Results for the First Research Question**

The first research question intended to find out the level of training English language instructors report in various areas of LTA and to what extent they perceive a need for in-service training in these areas. Both quantitative and qualitative data were analysed to answer these questions.

The results from the questionnaire revealed that nearly 48% of the participants did not have a separate testing and assessment course in pre-service education. LTA courses are offered only at ELT departments at Turkish universities. Those who answered 'No' to this question are most probably those who studied a major other than ELT since the percentage of the participants who said 'No' to this question is quite similar to the proportion of the participants (49%) who did not study ELT as a major at university. Moreover, almost 47% of the participants stated that they did not attend any LTA related professional development training. The interviewees' comments were in line with these findings. Only some of the interviewees reported

that they had received training in LTA in their pre-service education and few of them had felt adequately prepared for their LTA practices in their future career after receiving their BA degree. This is not surprising since LTA courses offered both in Turkey and worldwide was not found to be efficient enough to promote the LTA literacy of pre-service teachers (Jin, 2010; Lam, 2015; Hatipoğlu, 2015). However, these findings are relatively lower compared to Vogt and Tsagari's (2014) results in which they noted that the majority of the teachers interviewed received limited training regarding LTA during both their pre-service and in-service training.

As mentioned in Methodology and Results chapters, the second part of the questionnaire consisted of three subsections, namely classroom-focused LTA, purposes of testing, and content and concepts of LTA. Regarding the findings on the participants' reported training level in LTA, the participants perceived themselves the least trained in content and concepts of LTA whereas purposes of testing was the subsection in which the participants perceived themselves the most trained compared to the other two subsections.

Nonetheless, there were some differences among the subsections regarding the participants' perceived training needs in LTA. It was found out that the mean score for the perceived training need in content and concepts of LTA was higher than the perceived training need in both classroom-focused LTA and content and concepts of LTA. The implication is that the participants in this study perceive the highest need for training in content and concepts of LTA domains and the lowest need for purposes of testing domains.

It should also be noted that the findings from both quantitative and qualitative data indicated that the sample in this study perceived a strong need for training in the areas of LTA. This may be attributed to that the sample in this study were self-selecting and were most probably interested in LTA; and as mentioned in Kvasova and Kavytska's (2014) study, they were also highly motivated for professional development. Therefore, the participants in the study agreed with a need for training in most of the domains mentioned in the questionnaire. These initial findings provide a general perspective on the participants' reported LTA training and perceived training needs in the areas of LTA.



### **5.3 Discussion of Results for the Second Research Question**

The second research question aimed to explore if there is a statistically significant difference in English language instructors' reported training level and perceived training needs in LTA regarding their major. The findings from both the quantitative and qualitative data were analysed to answer this question.

As for their major, the participants were divided into two subgroups as those who studied ELT as a major and those who did not (non-ELT). The findings suggested that there was a significant difference between the two groups in relation to their reported LTA training in all three subsections. The mean score of the reported LTA training by ELT subgroup was higher than the non-ELT subgroup's. This finding aligns with Zhang and Burry-Stock's (2003) study, and is not surprising at all given that the participants in the non-ELT subgroup were comprised of instructors who had a degree on a major other than ELT. The instructors in this subgroup most probably did not receive as much formal training in LTA as the ELT subgroup during their pre-service education and the training they received was most probably during their in-service practices. Most of the interviewees, who had a degree on a major other than ELT, also made comments supporting this view.

On the other hand, no significant difference was observed between the two subgroups regarding their perceived training need in any of the three subsections. In other words, the participants from the ELT subgroup perceived similar amounts of training need in all three subsections to the participants in non-ELT subgroup although they had previous training in LTA. Therefore, it may be inferred that the LTA courses pre-service teachers take during their BA studies are not fulfilling and satisfactory enough to prepare pre-service teachers for their future LTA practices. It is also possible to make some deductions through the findings of the interviews about this result. A majority of the interviewees from both backgrounds expressed that they did not feel sufficiently prepared for their LTA activities after their pre-service education. That is probably why the participants from both subgroups perceived similar amounts of need for training in all three subsections.

### **5.4 Discussion of Results for the Third Research Question**

The third research question aimed to explore if there is a statistically significant difference in English language instructors' reported training level and perceived

training needs in LTA regarding their highest degree in education. The participants were divided into two subgroups as those with a BA degree and those with an MA or PhD degree in order to examine if any significant difference existed between the two subgroups. However, the findings indicated no significant difference between the two subgroups in this respect. Similarly, Büyükkarçı's (2016) study found no correlation between post-graduate studies and assessment literacy of the participants. This proposes two implications. First, the participants were not inquired about their MA or PhD majors. That is, they may have specialized in a non-ELT major in their MA or PhD studies. Therefore, their studies may not have had an impact on their perceptions of the training they received and needed in LTA. Second, even if the participants studied ELT for their MA / PhD degree, the practices during their MA / PhD studies may not have had a positive effect on their reported training level and perceived training needs in LTA. Implication is that MA or PhD programmes in Turkey most probably do not give enough importance on LTA courses. More time and effort should be devoted to the delivery of LTA courses in these programmes so that English language teachers may feel more prepared and confident for their LTA practices in their jobs.

### **5.5 Discussion of Results for the Fourth Research Question**

The fourth research question aimed to explore if there is a statistically significant difference in English language instructors' reported training level and perceived training needs in LTA regarding their LTA responsibilities in testing units. The participants were divided into two subgroups as those who work in the Testing Office (language testers) and those who do not work in the Testing Office (regular instructors) in order to examine if any significant difference existed between the two subgroups.

No significant difference was observed between language testers and regular instructors in relation to their reported training level in LTA. However, the training level reported by language testers would be expected to be higher than regular instructors'. Given that the group described as language testers in this study are the English language instructors who work in the Testing Offices at English Preparatory Schools, and that they are presumably more involved in LTA tasks in their daily practices, they should have received more training in the areas of LTA. Therefore,

implication is that language testers should receive more in-service training to develop their LTA practices.

Moreover, the results suggested no significant difference between language testers and regular instructors regarding their perceived training needs in LTA except for the perceived training need in purposes of testing. The perceived need for training by language testers was higher than the regular instructors for this subsection. This may be linked with the overall findings of the questionnaire suggesting that the participants on average felt a less need for training in purposes of testing compared to the other two subsections. In this case, considering that the language testers are more associated with LTA practices, their perceived need for training in all subsections may have been at similar levels even when the overall participants were less enthusiastic about training in purposes of testing. Therefore, when language testers and regular instructors were compared to each other for their perceived training needs in purposes of testing, language testers' training needs may have been observed higher than regular instructors'.

## **5.6 Discussion of Results for the Fifth Research Question**

The fifth research question aimed to explore if there is a statistically significant difference in English language instructors' reported training level and perceived training needs in LTA regarding their teaching experience. The participants were divided into 3 subgroups as those with a teaching experience of 1 to 5 years, 6 to 10 years, and those with a teaching experience of 11 or more years.

The findings indicated no significant difference among the subgroups in relation to their reported training level in classroom-focused LTA subsection. However, the reported training level in purposes of testing by the participants with a teaching experience of 6 to 10 years significantly differed from the participants with a teaching experience of 1 to 5 years. The participants in the 6-10 years subgroup reported that they received significantly more training in purposes of testing than the participants in the 1-5 years subgroup. Moreover, another significant difference was observed between the participants with a teaching experience of 11 or more years and those with 1 to 5 years regarding the reported training level in content and concepts of LTA. The participants with an experience of 11 or more years reported that they were significantly more trained in content and concepts of LTA than the

participants with an experience of 1-5 years. In accordance with these significant differences among subgroups, the mean scores of the reported training level in all three subsections by the three subgroups suggested that the participants with a teaching experience of 1-5 years had the lowest reported training level among the three subgroups. Hence, it may be deduced from the results of this study that the reported training levels in the areas of LTA tend to increase in accordance with teaching experience. This may be attributed to that the participants in this study most probably give importance to professional development and have attended some training in various areas of ELT as well as in LTA.

Regarding the perceived training needs, some differences were found among the subgroups in all three subsections. The perceived training need in classroom-focused LTA by the participants with a teaching experience of 6 to 10 years was significantly different from the participants with a teaching experience of 11 or more years. That is, the participants in the 6-10 years subgroup perceived more need for training in this subsection than the participants in the 11 or more years subgroup. It should also be noted that the mean scores of the perceived training need in classroom-focused LTA for 1-5 years and 6-10 years subgroups were nearly the same. Therefore, it may be concluded that the most experienced participants in this study perceived the least need for training in classroom-focused LTA and that the perceived training need in classroom-focused LTA seems to decrease in relation to teaching experience.

As for purposes of testing subsection, the perceived training needs of the participants with an experience of 1 to 5 years differed significantly from those with an experience of 11 or more years. In other words, the participants in the 1-5 years subgroup perceived more need for training than the participants in the 11 or more years subgroup. Moreover, a gradual decrease in the mean scores of the perceived training need for purposes of testing was observed in accordance with teaching experience. The implication from this finding is that more experienced English language instructors in this study perceive a less need for training in purposes of testing domains.

Regarding content and concepts of LTA, the perceived training needs of the participants with an experience of 6-10 years differed significantly from those with an experience of 11 or more years. That is, the participants in the 6-10 years

subgroup perceived more need for training in this subsection compared to the participants in the 11 or more years subgroup. Moreover, unlike the other two subsections, the mean scores of perceived training need in content and concepts of LTA for 1-5 years subgroup was slightly lower than 6-10 years subgroup'. This suggests that the participants in the 6-10 years subgroup perceived the highest need for training in content and concepts of LTA followed by the participants in 1-5 years subgroup while the participants in the 11 or more years perceived the least need.

The findings of the perceived training needs overall suggest that among the participants of the three subgroups, the participants with the highest amount of teaching experience (11 or more years) perceived the least need for training in all the areas of LTA mentioned in this study. Studies investigating the correlation between teaching experience and professional development training needs (Ekşi & Çapa Aydın, 2012; Karaarslan, 2003; Sentuna, 2002) also showed that need for professional development training was negatively correlated with teaching experience. In other words, those with more experience in teaching perceived less need for professional development training.

### **5.7 Discussion of Results for the Sixth Research Question**

The sixth research question aimed to find out English language instructors' perceptions of LTA training in pre-service and in-service education. The findings from the qualitative data were analysed to answer this question. The findings suggest that only some of the interviewees were trained in LTA in their pre-service education and few of them felt adequately prepared for their LTA practices in their future career after receiving their BA degree. This finding suggests that interviewees in this study did not receive appropriate training during their pre-service education and therefore, had to cope with their assessment practices without sufficient training. In order to compensate for their lack of training, the interviewees learnt about LTA on the job, through in-service training, self-study or with support from colleagues. It may be deduced from this finding that in-service teacher training is of great importance given that pre-service teachers start their career without proper training in LTA.

Regarding the in-service training in LTA, most of the interviewees stated to have attended several in-service training workshops or courses in LTA. The findings

also suggest that the interviewees are knowledgeable about recent LTA methods and use them to some extent in their in-class assessment practices. This may be attributed to that the interviewees in this study comprised of mostly experienced English language instructors who also care about their professional development. However, they would like to receive more practical and regular training especially on assessing productive skills and new / alternative assessment methods in classroom assessment. The implication from this finding is that although they have received training in various areas of LTA, the interviewees are not fully satisfied with the amount and content of the training they have received. Therefore, it may be inferred that research studies such as this current study should be carried out more often in instructors' local context, in other words, where they currently work in order to identify the training needs of English language instructors in that local context. Thus, workshops or training courses tailored for the specific needs of instructors may be arranged for them to benefit in the best way.

### **5.8 Pedagogical Implications**

Findings from this study propose some pedagogical implications for the development of LTA literacy of English language teachers. First of all, it was found out that the participants in this study did not feel sufficiently ready for their LTA practices after their pre-service education. Therefore, the content of testing and assessment courses in the ELT departments of higher education institutions should be revised, so that pre-service teachers may be provided with more practical experience which will prepare them for their in-service LTA tasks.

Second, the findings showed that the participants in this study mostly perceived a need for training in content and concepts of LTA such as assessing aspects of culture, using statistics to analyse test results, establishing validity and reliability of tests, assessing productive and integrated language skills as well as in using alternative and formative types assessment such as portfolio and informal or non-test type assessment. These findings could give some ideas to the administrative staff at English Preparatory Programmes when they plan or design their further in-service training workshops.

## **5.9 Conclusions**

The current study was conducted to find out the level of training English language instructors reported in various areas of LTA and to what extent they perceived a need for in-service training in these areas. More specifically, the study investigated if the reported training levels and perceived training needs of English language instructors in LTA demonstrated a significant difference in relation to their major, highest degree in education, LTA responsibilities in testing units, and teaching experience.

The findings indicated that the participants reported the highest training in classroom-focused LTA domains and perceived the highest need for training in content and concepts of LTA. Moreover, the findings from the qualitative data revealed that the majority of interviewees did not feel prepared for their LTA tasks after pre-service education and developed their knowledge and skills in LTA through their in-service practices on the job. Moreover, both quantitative and qualitative data suggest that there is a solid need for in-service training in almost all areas of LTA mentioned in this study, especially in formative types of assessment and in content and concepts of LTA.

It was also found out that the participants who studied ELT and those who did not significantly differed in their perceived training needs in all the LTA areas mentioned in this study. However, no significant difference was observed between those with a BA degree and those with an MA / PhD degree in their reported LTA training level or perceived training needs in LTA. Moreover, significant difference was observed between language testers and regular instructors in relation to their training needs in purposes of testing. Finally, although some differences were observed in the reported training level of instructors in relation to their teaching experience, it was found out that instructors with more teaching experience perceived less need for professional development training.

## **5.10 Recommendations for Future Research**

This study has several recommendations for future research. First, the findings of the current study indicated that although nearly half of the participants did not receive any formal training in LTA during their pre-service education, the majority

reported a good level of literacy in LTA. The findings suggested that since the sample in this study constituted mostly experienced instructors, the participants developed their LTA literacy on the job through in-service training, self-study or with support from colleagues. Therefore, a further study could be carried out on the level of training novice teachers report in LTA. In addition, their training needs in this area and how they compensate for these needs could be investigated.

Second, this study could not identify a correlation between MA / PhD studies and training levels and needs of the participants. This may be due to the fact that the participants were not enquired about their specialization in their MA or PhD studies. That is, it was not known whether those who had an MA / PhD degree received training in LTA or not. Therefore, no clear assumptions were drawn between two variables. Further studies could consider this gap and investigate what the participants specialized in their MA / PhD studies to find out any possible correlation between training levels and needs of English language teachers who hold a BA degree and those who hold an MA / PhD degree.



## REFERENCES

- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). The Standards for Teacher Competence in Educational Assessment of Students. Retrieved from <http://buos.org/standards-teacher-competence-educational-assessment-students>
- Ary, D., Jacobs, L. C., & Sorensen, C. K. (2006). *Introduction to research in education* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.
- Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum*. Sydney: National Centre for English Language Teaching and Research, MacQuarie University.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and practice*, 30(1), 3 – 12.
- Brown, H. D. (2004). *Language assessment principles and classroom practice*. White Plains, NY: Pearson Education.
- Büyükkaracı, K. (2016). Identifying the areas for English language teacher development: a study of assessment literacy. *Pegem Eğitim ve Öğretim Dergisi*, 6(3), 333-346. <http://dx.doi.org/10.14527/pegegog.2016.017>
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389. doi:10.1191/0265532204lt288oa
- Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.), *Handbook of classroom assessment* (p. 1 – 32). San Diego, CA: Academic.
- Cizek, G. J. (2009). Reliability and validity of information about student achievement: comparing large-scale and classroom testing contexts. *Theory into Practice*, 48(1), 63 – 71.
- Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners*. Ann Arbor, MI: University of Michigan Press.
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43 – 56.

- Dann, R. (2002). *Promoting assessment as learning: Improving the learner process*. London: Routledge / Falmer.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327 – 347. doi: 10.1177/0265532208090156
- Doyle, L., Brady, A., & Byrne, G. (2009). An overview of mixed methods research. *Journal of Research in Nursing*, 14(2), 175 – 185.
- Earl, L., & M, Katz (2006). *Rethinking classroom assessment with purpose in mind: assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship, and Youth, School Programs Division. Retrieved from [http://www.edu.gov.mb.ca/k12/assess/wncp/full\\_doc.pdf](http://www.edu.gov.mb.ca/k12/assess/wncp/full_doc.pdf)
- Ekşi, G. & Çapa Aydın, Y. (2012). English instructors' professional development need areas and predictors of professional development needs. *Procedia - Social and Behavioral Sciences*, 70, 675 – 685.
- Field, A. P. (2009). *Discovering statistics using SPSS: And sex and drugs and rock 'n' roll* (3rd ed.). London; Los Angeles [i.e. Thousand Oaks, Calif.]; SAGE Publications.
- Fletcher, A. K. (2016). Exceeding expectations: scaffolding agentic engagement through assessment as learning. *Educational Research*, 58(4), 400 – 419.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. doi: 10.1080/15434303.2011.642041
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge (England): Cambridge University Press.
- Gupta, K. (2016). Assessment as learning. *The Science Teacher*, 83(1), 43 – 47.
- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education & Literacy Studies*, 3(4), 42 – 48.
- Han, T., & Kaya, H. İ. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education*, 4(1), 77 – 93.
- Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365 – 379.
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs. Part one: General findings*. Retrieved from <http://www.ealta.eu.org/resources.htm>

- Hatipoğlu, Ç. (2010). Summative evaluation of an English language testing and evaluation course for future English language teachers in Turkey. *Elted Journal*, 13, 40 – 51. Retrieved February 21, 2018, from [http://www.elted.net/uploads/7/3/1/6/7316005/v13\\_5hatipoglu.pdf](http://www.elted.net/uploads/7/3/1/6/7316005/v13_5hatipoglu.pdf)
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, 4(2), 111 – 128.
- Hesse-Biber, S. (2010). Qualitative approaches to mixed methods practice. *Qualitative Inquiry*, 16(6), 455 – 468.
- Jannati, S. (2015). ELT teachers' language assessment literacy: perceptions and practices. *Educational Research Association The International Journal of Research in Teacher Education*, 6(2), 26 – 37.
- Jeong, H. (2013). Defining assessment literacy: is it different for language testers and non-language testers? *Language Testing*, 30(3), 345 – 362.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555 – 584.
- Johnson, R.B., & Onwuegbuzie, A.J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Research*, 33(7), 14 – 26.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385 – 402. doi: 10.1177/0265532208090158
- Karaarslan, D. (2003). *Teachers' perceptions of self-initiated professional development*. (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.
- Kiomrs, R., Abdolmedi, R., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: the case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156 – 161.
- Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy, and Practice*, 16(3), 263 – 268.
- Kvasova, O., & Kavytska, T. (2014). The assessment competence of university foreign language teachers: A ukrainian perspective. *Language Learning in Higher Education*, 4(1), 159 – 177.

- Lam, R. (2015). Language assessment training in Hong Kong: implications for language assessment literacy. *Language Testing*, 32(2), 169 – 197.
- Lomax, R. G. (1996). On becoming assessment literate: an initial look at preservice teachers' beliefs and practices. *The Teacher Educator*, 31(4), 292 – 303.
- Lopez Mendoza, A. A. & Bernal Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile*, 11(2), 55 – 70.
- Malone, M. E. (2011). Assessment literacy for language educators. *CALdigest*. Retrieved from <http://www.cal.org/resource-center/briefs-digests/digests>
- Malone, M. E. (2013). The essentials of assessment literacy: contrast between testers and test users. *Language Testing*, 30(3), 329 – 344.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34 – 43. doi:10.1111/j.1745-3992.2003.tb00142.x
- Mertler, C. A. (2005). Secondary teachers' assessment literacy: does classroom experience make a difference? *American Secondary Education*, 33(2), 76 – 92.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101 – 113.
- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 13, 1-31. Retrieved from <http://www.ncolctl.org/files/LCTL-Teachers-Assessment.pdf>
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained. Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48 – 76.
- Navarete, C., Wilde, J., Nelson, C., Martínez, R., & Hargett, G. (1990, Summer). *Informal assessment in educational evaluation: implications for bilingual education programs*. (NCBE Program Information Guide Series, No. 3). Retrieved from: <http://www.ncele.us/files/rcd/BE017505/PIG3.pdf>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381 – 402.

- Plake, B. S., Impara, J. C., & Fager, J. J. (1993), Assessment Competencies of Teachers: A National Survey. *Educational Measurement: Issues and Practice*, 12, 10 – 12. doi:10.1111/j.1745-3992.1993.tb00548.x
- Popham, W. J. (2006). Needed: a dose of assessment literacy. *Educational Leadership*, 63, 84 – 85.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W. J. (2009). Assessment literacy for teachers: faddish or fundamental? *Theory into Practice*, 48, 4 – 11.
- Rea-Dickens, P. (2007). Classroom-Based Assessment: Possibilities and Pitfalls. In J. Cummins, & C. Davison (Eds), *International Handbook of English Language Teaching* (pp. 505 – 520). Springer, Boston, MA: Springer International Handbooks of Education
- Sentuna, E. (2002). *Interests of EFL instructors in Turkey regarding INSET content*. (Unpublished master's thesis). Bilkent University, Ankara, Turkey.
- Sheehan, S., & Munro, S. (n. d.). Assessment: attitudes, practices and needs. *ELT Research Papers*, 17.08. Retrieved from [https://englishagenda.britishcouncil.org/sites/default/files/attachments/g239\\_elt\\_ra\\_sheehan\\_and\\_munro\\_final\\_web.pdf](https://englishagenda.britishcouncil.org/sites/default/files/attachments/g239_elt_ra_sheehan_and_munro_final_web.pdf)
- Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan*, 72(7), 534 – 539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *The Phi Delta Kappan*, 77(3), 238 – 245.
- Stiggins, R. J. (2002). Assessment crisis: the absence of assessment for learning. *The Phi Delta Kappan*, 83(10), 758 – 765.
- Stiggins, R. J. (2005). From formative assessment to assessment for learning: a path to success in standards-based schools. *The Phi Delta Kappan*, 87(4), 324 – 328.
- Stiggins, R. J. (2007). Assessment through the student's eyes. *Educational Leadership*, 64(8), 22 – 26.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21 – 36. doi:10.1017/S0267190509090035
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. In C. Teddlie & A. Tashakkori (Eds.), *Handbook of mixed methods in social & behavioural research*, (3-50). Thousand Oaks, CA: Sage.

- Tsagari, D. (2016). Assessment orientations of state primary EFL teachers in two mediterranean countries. *Center for Educational Policy Studies Journal*, 6(1), 9 – 30.
- Uzun, L. (2016). Evaluation of the latest English language teacher training programme in Turkey: Teacher trainees' perspective. *Cogent Education*, 3(1), 1147115, 1 – 16. <https://doi.org/10.1080/2331186X.2016.1147115>
- Vogt, K. & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European Study. *Language Assessment Quarterly*, 11(4), 374 – 402.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749 – 770.
- Weathington, B. L., Cunningham, C. J., & Pittenger, D. J. (2010). *Research Methods for the behavioural and social sciences*. Hoboken, NJ: John Wiley & Sons.
- Webb, N. L. (2002, April, 1 – 5). *Assessment literacy in a standards-based education setting* (WCER Working Paper No. 2002-4). Retrieved from Wisconsin Center for Education Research website: [http://wcer.wisc.edu/docs/working-papers/Working\\_Paper\\_No\\_2002\\_4.pdf](http://wcer.wisc.edu/docs/working-papers/Working_Paper_No_2002_4.pdf)
- Wininger, S. R., & Norman, A. D. (2005). Teacher candidates' exposure to formative assessment in educational psychology textbooks: a content analysis. *Educational Assessment*, 10(1), 19 – 37.
- Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 33 – 158. Retrieved February 25, 2018, from [http://arts.unimelb.edu.au/\\_data/assets/pdf\\_file/0011/2349947/6\\_1\\_SI7XuBrown.pdf](http://arts.unimelb.edu.au/_data/assets/pdf_file/0011/2349947/6_1_SI7XuBrown.pdf)
- Yetkin, C. (2015). *An investigation on ELT teacher candidates' assessment literacy* (Unpublished master's thesis). Çağ University, Mersin, Turkey.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323 – 342. 10.1207/S15324818AME1604\_4

## APPENDICES

### A. Survey on Language Testing and Assessment Needs

Dear Colleague,

The survey below was adapted from Vogt & Tsagari (2008) to investigate to what extent EFL teachers working at English Preparatory Programmes at Turkish Universities received training in various areas of Language Testing and Assessment (LTA) during their pre-service and in-service education and also to find out their training needs in these areas.

The survey takes about 5 - 6 minutes to complete and comprises two main parts:

- Part I. Background Information, and
- Part II. Questions about Training in LTA
  - A. Classroom-focused LTA
  - B. Purposes of testing
  - C. Content and concepts of LTA

For each subsection in Part II, you are kindly requested to choose "Not at all", "Poor", "Average", "Good", or "Advanced" for the training you received in the mentioned domain, then choose "Strongly Disagree", "Disagree", "Undecided", "Agree", or "Strongly Agree" for the training you need in the mentioned domain.

Pre-service education refers to the period before you have undertaken any teaching. In-service education refers to the period during which you have been employed.

All your responses will be kept confidential, and will only be used for research purposes.

If you have further questions, please do not hesitate to contact me.

Thank you for your contribution!

Mehmet Türk  
MA Student  
Graduate School of Educational Sciences  
Bahçeşehir University  
mehmetturk@beykent.edu.tr

Supervisor: Assist. Prof. Hatime ÇİFTÇİ  
Graduate School of Educational Sciences  
Bahçeşehir University

**Part I. Demographic Information**

1. Age:       20-25       26-30       31-35       36-40  
                  41-45       46-50       51 or above
2. Gender:     Male       Female
3. What did you study at university?  
                  ELT                       Literature                       Linguistics  
                  Translation and Interpreting                       Other: .....
4. What is your highest qualification?  
                  B.A.       M.A       Ph.D.
5. Years of experience in teaching:  
                  1-5       6-10       11-15       16-20  
                  21 or more
6. Where do you teach?  
                  Preparatory School of a state university  
                  Preparatory School of a private university  
                  Other: .....
7. Please write the name of your institution. ....
8. Extra job responsibilities:  
                  Head of department or school  
                  Mentor  
                  Level Coordinator  
                  Testing and Assessment Office  
                  Material and Curriculum Development Office  
                  Other: .....
9. Did you have a separate testing / assessment course in pre-service education?  
                  Yes  
                  No
10. Have attended any professional development training on language testing and assessment?  
                  Yes  
                  No



## Part II. Questions about Training in LTA

### 1. Classroom-focused LTA

1.1. I **was trained** in the following domains:

	Not at all	Poor	Average	Good	Advanced
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					

1.2 I **need training** in the following domains:

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
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					

### 2. Purposes of testing

2.1. I **was trained** in the following domains:

	Not at all	Poor	Average	Good	Advanced
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 I **need training** in the following domains:

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
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)					

### 3. Content and concepts of LTA

3.1. I **was trained** in the following domains:

	Not at all	Poor	Average	Good	Advanced
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 I **need training** in the following domains:

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
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					

## **B. CONSENT FORM & INTERVIEW QUESTIONS**

### *CONSENT FORM – INTERVIEW WITH AUDIORECORDING*

#### **Consent to Participate in Research**

##### **Introduction and Purpose**

My name is Mehmet Türk. I am a graduate student at Bahçeşehir University in the Department of English Language Teaching. I would like to invite you to take part in my research study regarding testing and assessment practices of English language instructors working at English Preparatory Programmes of Turkish universities.

##### **Procedures**

If you agree to participate in my research, I will conduct an interview with you at a time and location of your choice. The interview will involve questions about your work conditions. It should last about 15 minutes. With your permission, I will record and take notes during the interview. The recording is to accurately record the information you provide, and will be used for transcription purposes only. If you choose not to be recorded, I will take notes instead. If you agree to being recorded but feel uncomfortable at any time during the interview, I can turn off the recorder at your request. Or if you don't wish to continue, you can stop the interview at any time.

##### **Confidentiality**

Your study data will be handled as confidentially as possible. If results of this study are published or presented, individual names and other personally identifiable information will not be used.

##### **Rights**

*Participation in research is completely voluntary.* You are free to decline to take part in the project. You can decline to answer any questions and are free to stop taking part in the project at any time.

##### **Questions**

If you have any questions about this research, please feel free to contact me. I can be reached at [turkmemet@gmail.com].

##### **CONSENT**

You will be given a copy of this consent form to keep for your own records. If you wish to participate in this study, please sign and date below.

\_\_\_\_\_  
Participant's Name

\_\_\_\_\_  
Participant's E-mail Address

\_\_\_\_\_  
Participant's Signature

\_\_\_\_\_  
Date

**Name & Surname:**

**Your age:**

**Gender:**

**What did you study at university?**

**How long have you been teaching?**

**What is your highest qualification?**

- 1.** During your pre-service teacher training did you learn about language testing and assessment (LTA)?
- 2.** Did you feel appropriately prepared for your LTA tasks after pre-service training?
- 3.** If not, how did you learn about LTA?
- 4.** How do you assess your students' learning progress during in-class activities? Do you know about more recent LTA methods such as portfolio assessment, self- or peer-assessment? Have you ever tried them?
- 5.** Do you feel competent about using the types of LTA conducted in your Preparatory School?
- 6.** Have you received in-service training in LTA? If yes, what was the focus of this training? Please give the details about the training.
- 7.** What LTA training would you like in the short-term and/or in the long-term?
- 8.** Do you have any other comments?

## CURRICULUM VITAE

### PERSONAL INFORMATION

Surname, Name: Türk, Mehmet

Nationality: Turkish (T.C.)

Date and Place of Birth: 24 February 1986, Alaşehir

Marital Status: Married

Mobile: (+90) 554 899 7910

E-mail: [turkmemet@gmail.com](mailto:turkmemet@gmail.com)

### EDUCATION

Degree	Institution	Year of Graduation
BA	Marmara University	2008
High School	Alaşehir Şehitleri Anatolian High School	2004

### WORK EXPERIENCE

Year	Place	Enrolment
2011 -	Beykent University School of Foreign Languages	English Instructor

### FOREIGN LANGUAGES

Advanced English

### COURSES

- Oxford Teachers' Academy, Making Digital Sense (April 2013, Beykent University, Istanbul)
- Oxford Teachers' Academy, Teaching English To Adults (April 2012, Beykent University, Istanbul)
- Introduction to Teaching English as a Foreign Language (May 2009, Thanet College, Broadstairs, the United Kingdom)
- The Year 2 Phonics Workshops for Parents (October 2008, Priory Infant School, Ramsgate, the United Kingdom)

- English (July 2006 – September 2006, International Center New York, New York, the USA)

### **CONFERENCES**

- IATEFL 51st international Conference and ELT Exhibition (April 2017 – Glasgow, the United Kingdom)
- 7 th International ELT Conference “CEFR: From a TEACHing to a LEARNing Curriculum” (March 2013 – Beykent University, Istanbul)
- 6th International ELT Conference on The Use of Technology in Student-Centered Learning (March 2012 – Beykent University, Istanbul)
- Together Everyone Achieves More (June 2008 - Doğa Schools, Istanbul)
- Language, Culture and Communication in ELT (February 2008 - Çevre Schools, Istanbul)
- Inspiration, Motivation, Management in ELT (February 2007 - Çevre Schools, Istanbul)

### **PERSONAL INTERESTS**

Reading, Cinema, History, Geography, Jogging, Fitness