# HACETTEPE ÜNİVERSİTESİ
# EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Department of Foreign Language Education

English Language Teaching

# A STUDY INTO LANGUAGE ASSESSMENT LITERACY OF PRESERVICE ENGLISH AS A FOREIGN LANGUAGE TEACHERS IN TURKISH CONTEXT

Gamze SARIYILDIZ

Master's Thesis

Ankara, 2018

With leadership, research, innovation, high quality education and change,

*To the leading edge... Toward being the best...*

# HACETTEPE ÜNİVERSİTESİ
# EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Department of Foreign Language Education

English Language Teaching

A STUDY INTO LANGUAGE ASSESSMENT LITERACY OF PRESERVICE
ENGLISH AS A FOREIGN LANGUAGE TEACHERS IN TURKISH CONTEXT

TÜRKİYE'DEKİ İNGİLİZCE ÖĞRETMEN ADAYLARININ DİL
DEĞERLENDİRME OKURYAZARLIĞI ÜZERİNE BİR ARAŞTIRMA

Gamze SARIYILDIZ

Master's Thesis

Ankara, 2018

# Acceptance and Approval

To the Graduate School of Educational Sciences,

This thesis entitled "A Study into Language Assessment Literacy of Preservice English as a Foreign Language Teachers in Turkish Context" has been approved as a thesis for the Degree of **Master in English Language Teaching, Foreign Language Education** by the members of the Examining Committee.

Chair                      Assoc. Prof. Dr. Kadriye Dilek Akpınar

Member (Supervisor)    Assoc. Prof. Dr. Hüseyin Öz

Member                    Assoc. Prof. Dr. Nuray Alagözlü

This is to certify that this thesis/dissertation has been approved by the aforementioned examining committee members on 04/06/2018 in accordance with the relevant articles of the Rules and Regulations of Hacettepe University Graduate School of Educational Sciences, and was accepted as a **Master's Thesis** in the **Program of English Language Teaching** by the Board of Directors of the Graduate School of Educational Sciences on ...../...../.........

Prof. Dr. Ali Ekber ŞAHİN
Director of Graduate School of Educational Sciences

**Abstract**

This study aimed to investigate language assessment literacy of preservice English-as-a-foreign-language (EFL) teachers at a major state university in Turkey. The study also sought to find out their perceptions as regard to the extent of theoretical and practical training of assessment they received in their teacher education program, and also whether they perceived a need for further training in language assessment. Another aim of this study was to find out how preservice English language teachers evaluated school experience course in relation to English Language Testing and Evaluation (ELTE) course. The study used a mixed-methods research design in which both quantitative (questionnaire) and qualitative (interview) data were collected. A total of 101 4th grade preservice teachers enrolled in the ELT department of Middle East Technical University participated in the study. In addition, a total of 25 participants were involved in the qualitative part of the study. Quantitative data were analyzed using descriptive statistics such as frequencies and percentages, and means. Qualitative data gathered through semi-structured interviews were analyzed using selective coding. The results indicated that the preservice EFL teachers perceived the training that they received in different domains of language testing and assessment not to be sufficient and they needed further basic training in these domains. The findings of the qualitative data revealed that participants did not have the chance to put theoretical knowledge of language testing and assessment into practice during teaching practicum, and the theory and practice of assessment was not covered much in their school experience course. In addition, the results showed that participants were aware of the importance and contributions of ELTE course to their professional development and future practices as language teachers.

**Keywords:** Pre-service teacher education, EFL pre-service teachers, language testing and assessment, language assessment literacy

# Öz

Bu araştırma Türkiye'deki bir devlet üniversitesindeki İngilizce öğretmen adaylarının dil değerlendirme okuryazarlığını incelemeyi amaçlamaktadır. Çalışma ayrıca İngilizce öğretmen adaylarının öğretmen yetiştirme programlarında değerlendirmenin teorisi ve uygulaması üzerine aldıkları eğitimin ne ölçüde olduğuna ilişkin görüşlerini ve dil değerlendirmesinde ek eğitimi ihtiyaç olarak görüp görmediklerini bulmayı amaçlamaktadır. Bu araştırmanın bir diğer amacı da, İngilizce öğretmen adaylarının okul deneyimi dersini yabancı dil öğretiminde ölçme ve değerlendirme dersine bağlı olarak nasıl değerlendirdikleridir. Bu amaçlar doğrultusunda, bu çalışmada hem nicel araştırma (anket) hem de nitel araştırma (mülakat) yöntemlerinin kullanıldığı karma yöntemler araştırması kullanılmıştır. Araştırmanın nicel kısmına Orta Doğu Teknik Üniversitesi'nin İngiliz Dili Eğitimi Anabilim Dalı'nda 4.sınıfta olan toplam 101 öğretmen adayı, nitel kısmına ise 25 kişi katılmıştır. Nicel veriler sıklık, yüzdeler ve ortalama değer gibi betimsel istatistik yöntemi kullanılarak analiz edilmiştir. Yarı yapılandırılmış görüşmelerden elde edilen nitel verilerin seçici kodlama ile analizi yapılmıştır. Bulgular, İngilizce öğretmen adaylarının dil ölçme ve değerlendirmenin farklı alanlarında aldıkları eğitimi yeterli bulmadıklarını ve bu alanlarda ek eğitime ihtiyaç duyduklarını göstermiştir. Mülakat sonuçları, katılımcıların dil ölçme ve değerlendirmedeki teorik bilgilerini öğretmenlik uygulamasında pratiğe dönüştürme fırsatı bulamadığını ve değerlendirmenin teorisinin ve uygulamasının okul deneyimi dersinde çok işlenmediğini göstermiştir. Ayrıca, sonuçlar katılımcıların yabancı dil öğretiminde ölçme ve değerlendirme dersinin öneminin ve bu dersin onların profesyonel gelişimlerine ve dil öğretmenleri olarak gelecekteki uygulamalarına katkılarının farkında olduklarını ortaya çıkarmıştır.

**Anahtar Sözcükler:** Hizmet öncesi öğretmen eğitimi, İngilizce öğretmen adayları, dil ölçme ve değerlendirme, dil değerlendirme okuryazarlığı

## Acknowledgements

To my beloved family
To my precious Seda

# Table of Contents

# List of Tables

# List of Figures

## Symbols and Abbreviations

**LTA:** Language Testing and Assessment

**EFL:** English as a Foreign Language

**ELT:** English Language Teaching

**LAL:** Language Assessment Literacy

**ELTE:** English Language Testing and Evaluation (ELTE)

# Chapter 1

## Introduction

**Background of the Study**

In today's world, where knowledge and competency of foreign languages and especially of English becomes inevitable, there is an increasing professional demand on the part of language teachers to keep pace with changes and reforms in foreign language education policies all over the world. Keeping pace with these reforms is an essential component of qualified language instruction and learning. Language instructors are supposed to adapt to these new reforms, which include many implementations such as delivering an effective language instruction, teaching strategies and assessing students' language development accurately.

One of the cornerstones of these reforms is tests and assessment that has a fundamental place in learning and teaching processes. If good assessment is done, it benefits teachers in many ways such as enabling them to determine the appropriateness of content, pace of the course, and teaching methods and enabling them to monitor student learning during the course. Likewise, it benefits students in several ways like giving students a chance to regard assessment as part of learning process, helping them monitor their own progress, and using information and feedback from assessment to improve their knowledge. It also enables students to gain confidence when they prepare for national tests (Thomas, Allma, & Beech, 2004).

In view of the fact that there is an increasing importance and growth in tests and assessments all over the world and teachers allocate most of their time for assessment, which was found in studies carried out in various educational settings (Stiggins, 2008), the crucial role of assessment in education system cannot be disregarded. In this regard, teachers are supposed to possess knowledge of sound assessment practice and skills of assessment, and these are the major elements of assessment literacy. To put it another way, teachers are supposed to be competent at assessment, which is one of the requirements placed upon teachers.

Assessment literacy, first coined by Stiggins (1995), refers to knowing the reasons and ways of assessing, being aware of the probable problems of assessment, and ways of avoiding these problems as well as being familiar with both positive and negative consequences of inaccurate assessments. Becoming assessment literate requires instructors to have a grasp of assessment in not only theory but also practice as well as developing skills to accurately administer, interpret and make decisions about assessment. Assessment literacy of teachers is considered the cornerstone of the relation between the quality of assessment and learners' improvement (Mertler, 2002), and instructors are expected to be competent at assessment. With regard to the growing importance of tests and assessment all over the world, there is a need for teachers to be assessment literate, which benefits both students and teachers (Coombe, Davidson, O'Sullivan, & Stoynoff, 2012; Newfields, 2006; Popham, 2011).

In parallel with the growth of assessment in the world, language assessment field has witnessed enormous expansion and increasing growth in terms of moving from periphery to center stage during recent years. This growth does not just mean that language tests and scores of these tests are increasingly used for many purposes such as education, employment, professionalism, migration and citizenship. Furthermore, it means that more and more people are engaged in selecting, developing them and using the scores of them for the purpose of making decisions. However, those people involved in fulfilling these are found to have received little training in this field (Taylor, 2009).

Considering the growth and significance of tests and assessments in language field as well as meeting the need for people who have adequate background and training in language assessment, language instructors are expected to be competent at language assessment, which is termed as "language assessment literacy". Fulcher (2012) defines language assessment literacy (LAL) as follows:

> "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)

Language assessment literacy is crucial in language education (Scarino, 2013) on the ground that it enables teachers to comprehend, figure out and implement information about student achievement with the aim of enhancing instruction (Falsgraf, 2005). LAL also allows language teachers to comprehend the pros and cons of different types of assessment as well as enabling them to determine effective tools of language assessment in order to achieve objectives (Siegel & Wissehr, 2011). Developing LAL allows language teachers to delve into and interpret their prejudgments, and comprehend interpretation-based feature of assessment. Besides, it also enables teachers to realize their own structure of understanding, philosophy and applications. Going through these processes will enable language teachers to develop self-awareness as testers, which constitutes fundamental part of their LAL (Scarino, 2013). Having a high level of LAL is crucial in that it enables English language teachers to establish a link between language teaching approaches and practices of assessment, and to select and design appropriate assessments among alternatives as well as examining the effect of some standardized tests such as TOEFL and IELTS (Herrera & Macias, 2015). Similarly, Malone (2011) highlights the interdependent connection between instruction and assessment as they inform each other. In this regard, it is imperative for EFL teachers to have sound foundations in LAL.

Despite the importance of assessment in educational context, especially in language, instructors are not trained adequately in language testing and assessment (LTA); therefore, most experience lack of knowledge, practice, experience and confidence in assessment. In this regard, quite a few instructors are found to be sufficient for meeting requirements and overcoming challenges of classroom assessment (Stiggins, 2002). In spite of the compelling evidence for assessment literacy, only few instructors are engaged in giving decisions about assessment with adequate instruction (DeLuca, 2012; Lam, 2015). Among the most critical issues in education is the level of assessment literacy, which is extremely low, among educators (Popham, 2010) and such "assessment illiteracy" leads to professional suicide as Popham (2004) highlights.

In spite of the history of advocacy for sound foundations in assessment, and the importance of a thorough training in LTA, teacher training programs often ignored assessment component in the past (Guskey, 2003; Popham, 2011). In this regard, a growing body of research has supported the critical need for assessment education in these programs (Popham, 2004; Volante & Fazio, 2007). In addition to the critical demand for assessment for effective teaching, accountability issue has compelled these programs to place more importance on assessment training compared to the past (DeLuca & Bellara, 2013). While there is a widespread need for emphasis on assessment training in teacher education programs, DeLuca and Klinger (2010) indicated that this emphasis has not been at the extent to which it is required yet by arguing that there are still extremely few cases of mandated and formalized assessment education. Accordingly, many teachers assume that they were not proficient in assessment and they need assistance to assess their students' performance and to make decisions about assessment (Mertler, 1998, 1999; Mertler & Campbell, 2005; Stiggins, 1999). Studies conducted on assessment literacy of teachers' reveal that teachers often state that assessment training received in undergraduate education did not prepare them enough for assessing students' performance and being comfortable with assessment-related decisions; therefore, they feel their lack of preparation stems mainly from inadequate undergraduate training (Mertler, 2009; Plake, 1993).

A growing body of research also contends that prospective teachers graduate from teacher training programs without receiving sufficient training in assessment (Popham, 1999), and they are not much exposed to coursework and other experiences related to assessment during preservice training (Stiggins, 2002). Other research that investigates assessment literacy growth during teacher education program reveals that prospective teachers graduate not only with inadequate understanding of key concepts of assessment, but also with low self-competence in employing assessment (Volante & Fazio, 2007). Indeed, mounting research highlights insufficient preparation and training of language instructors in the field of LTA (Berry & Daughterty, 2011; Chelsea & Jordan, 2012; Lam, 2014) as language teachers often get limited training in this field in their undergraduate education (Malone, 2008). Research also reveals that LAL appears to be

underdeveloped to some degree (Vogt & Tsagari, 2014) on the ground that training in LTA in teacher education programs stays marginalized (Fulcher, 2012; Gu, 2014).

With respect to the limited instruction in LTA, more training for prospective teachers in LTA is required, and teacher education programs are responsible for training prospective language teachers adequately in developing, using, scoring, and interpreting assessments (López & Bernal, 2009). Indeed, preservice language teachers are required to be trained not only in conceptual and theoretical bases of assessment, but also in putting such knowledge into their professional practice (Scarino, 2013). Research into teachers' preparation and training in assessment in teacher education programs also contends that lack of preparation also results from the fact teacher training in this field is "theory-laden" and "disconnected from teachers' daily assessment practices" (DeLuca & Bellara, 2013, p. 2). As these instructors start teaching and are expected to comprehend and apply various classroom assessments that inform and promote student learning in current educational context (Goc Karp & Woods, 2008; Roscoe, 2013), it is imperative to provide preservice teachers with increased practice opportunities in assessment (Talanquer, Tomanek, & Novodvorksky, 2013).

The vast majority of research also emphasizes that studies mostly focus on in-service teachers' knowledge and beliefs instead of their real practices (DeLuca & Klinger, 2010), which calls for the necessity for conducting study on prospective teachers' assessment practices in authentic contexts as little is known about it (Talanquer, Tomanek, & Novodvorksky, 2013). Among the few studies on preservice teachers' practices of assessment in authentic contexts, it is revealed that giving preservice teachers opportunities to create, administer, and analyze assessments in authentic contexts is stated to support the development of assessment literacy (Shepard, Hammerness, Darling-Hammond, & Rust, 2005). It is also claimed that preservice teachers who have been in a field experience and given chance to do assessment in real contexts seem to be better in bases of assessment than those who have not been in a field experience (Alkharusi, Kazem, & Al- Musawai, 2011). As practicum is essential for instructors' continuing professional development (Howitt, 2007; Loughran, Muthall, &Berry, 2008), and functions as a platform for establishing a connection between theory and practice

(Darling-Hammond, 2006), applying assessment in practicum contributes to the professional development. In this regard, practice of assessment in practicum develops preservice teachers' understanding of key terms of assessment and their self-efficacy. Accordingly, all of these point to the critical need for and importance of practice of assessment in practicum for preservice language teachers, which contributes to the development of overall assessment literacy of them, and to their professional preparation.

Assessment literacy is widely debated and emphasized in teacher training; however, few studies have been conducted on LAL over the world. Studies conducted on language assessment literacy of language teachers aimed to increase awareness of the prominence of LAL in teaching EFL, and to determine the level and instruction needs of foreign language teachers in LAL together with their perceptions and practices of English language assessment. Indeed, there are both quantitative and qualitative studies that aim to explore LAL of teachers in terms of these different aspects. However, limited research has been conducted on the LAL of preservice language teachers'. Only few studies have investigated the general training in LTA in teacher education programs with regard to exploring the enhancement of the LAL of preservice language teachers', and examining the extent to which assessment courses that they have taken facilitates or prevents this development (Lam, 2015; Vogt & Tsagari, 2014). Though LAL has a prominent role in language instructor training, literature review illustrates a prominent lack of study on LAL in teacher education.

Few studies have been carried out on LAL of preservice language teachers in Turkey. In fact, studies in this context are mostly limited to language teachers, which aims to examine their LAL levels and whether some factors affect their level of LAL such as teaching experience and having post-graduate training. Accordingly, limited research has been conducted on the LAL of preservice teachers, and their education in assessment in teacher training programs. While some studies underlie the necessity of giving sufficient training in LTA (Hatipoglu, 2015; Oz & Atay, 2017), other studies emphasize the viewpoints of prospective English teachers on the use of assessment strategies, methods and contributions of university programs to their development of LAL (Cirit, 2015; Yetkin, 2015). When studies in this field are examined, what remains unclear and needs to be

investigated is the status quo of prospective teachers' language assessment literacy. Thus, this research seeks to delve into LAL of prospective language teachers by examining the extent to which they are trained in assessment in teacher education program.

In conclusion, it is critical to delve into to what extent prospective EFL teachers receive training to possess and develop LAL, considering the role of language assessment training in shaping preservice teachers' grasp and development of language assessment literacy.

**Statement of the Problem**

This study seeks to examine the language assessment literacy of preservice EFL teachers, and finding out to what extent they have received training in both theoretical and practical terms of assessment in teacher education programs, and they perceive a need for training in language assessment.

Different from some previous studies in the field, this study will not only examine the status quo of preservice language teachers' assessment literacy, but also the extent to which assessment knowledge is put into practice in teaching practicum, and theory and practice of assessment is covered in school experience course and practicum. Moreover, it will attempt to find out whether and how they should be covered in school experience course and practicum. The study will also strive to see the usefulness of English Language Testing and Evaluation (ELTE) course on their school experience course and practicum experience.

**Aim and Significance of the Study**

This study aims to explore language assessment literacy of preservice EFL teachers. To this end, the extent to which they have received training in both theoretical and practical terms of assessment in teacher education programs, and they perceive a need for training in language assessment will be examined. Besides, it aims to explore preservice language teachers' opinions on both whether and how the theory and practice of language assessment was covered and whether and how they should be covered in school experience course, and the usefulness of English Language Testing and Evaluation (ELTE) course on their school experience course in depth.

The significance of this study can be defined from three aspects. First of all, considering the critical importance of assessment in language education, LAL in the context of foreign language teacher training is fundamental and should be possessed by language teachers during their undergraduate education, and developed in the following stages of their professional life. In this regard, their undergraduate training in assessment is crucial for the process of development of LAL, and if it is adequate, it prepares prospective teachers well for their future assessment practices, which influences the quality of language education. Accordingly, investigating the training in assessment that preservice teachers undergo in teacher education programs is imperative in terms of gaining deeper insights into the status quo of preservice language teachers' assesment literacy.

Secondly, since language teachers must engage in assessment practices in authentic contexts, it is fundamental for preservice language teachers to be trained in practice of assessment, and to be provided with the opportunities of putting theory of assessment into practice. With regard to this, teacher education programs should offer practice opportunity to prepare them well for their future profession. Considering the school experience and practice teaching courses offered in teacher training programs to consolidate the skills necessary for being effective language teacher, it is vital for them to incorporate assessment, and give preservice language teachers opportunity to practice assessment in a real context with real students. Accordingly, examining to what extent preservice language teachers implement their assessment knowledge in their practicum, and the extent to which assessment in both theory and practice has been covered in practicum as well as their perceptions on these issues is necessary in that it provides a new perspective on establishing connection between theory and practice training in assessment.

Lastly, very few studies have been conducted on LAL of preservice teachers, and their training in both theoretical and practical terms of assessment in teacher education programs in Turkey. In this regard, this study provides insight into the LAL of preservice EFL teachers, and their training in assessment as well as their training needs. The results of the study may offer new insights for decision-makers, teacher trainers in teacher education programs in Turkey, and make them reconsider the crucial role of assessment in the quality of language

education, the need for training to develop language assessment literacy, and encourage them to take some measures to improve the situation. Therefore, it can make contribution to the development of policy of teacher training programs on assessment training. To conclude, this study will be significant in that it will make contributions to the existing literature in the field of LAL in the world, and provide a new deep insight about LAL into the very limited literature of Turkey.

**Research Questions**

Regarding the research gap on language assessment literacy of preservice language teachers, this study seeks to answer the following research questions:

1) What level of training in areas of language testing and assessment do preservice language teachers state? Do participants regard this training as sufficient?

2) What dimensions of language assessment literacy do preservice English language teachers need training in?

3) How do preservice English language teachers evaluate school experience course in relation to English Language Testing and Evaluation course?

**Limitations of the Study**

The major limitation of this study is that it has been conducted only in the ELT Department of Middle East Technical University. The quantitative data for the present study was collected from 101 participants who were 4th graders in the university and have taken English Language Testing and Evaluation (ELTE) course. Although the number of participants is high, it does not include all preservice language teachers who have taken ELTE course in ELT Departments of different universities in Turkey. The data could have been gathered from the other ELT departments of the universities in Turkey and it would have been much easier to generalize the results for all preservice English language teachers in Turkey.

**Definitions of Terms**

Below are the definitions of the key terms used in the present study:

*Assessment:* Refers to an ongoing procedure of getting, interpreting and using information about students' progress towards goals (Dhindsa, Omar, & Waldrip, 2007; Linn & Miller, 2005).

*Assessment Literacy:* Refers to being knowledgeable about the ways of assessing knowledge and skills of learners as well as analyzing the results and enhancing learning and program qualifications (Webb, 2002).

*Language Assessment Literacy:* Refers to skills, knowledge and abilities for constructing and evaluating tests, and grasp of the theory of assessment in addition to the role of assessment on people and community (Fulcher, 2012).

*Preservice education:* Refers to "course or program of study which student teachers complete before they begin teaching" (Richards & Schmidt, 2002, p. 416).

*Preservice Teachers:* Refers to student teachers who study in teacher training programs and working towards their certification.

# Chapter 2
## Literature Review

### Assessment

Assessment, which constitutes an essential component of education system, is an umbrella term covering a variety of strategies including tests (Clapham, 2000). Tests; on the other hand, are prepared administrative procedures during which students are aware that measurement and evaluation take place regarding their responses. As for the other confusing terms, which are measurement and evaluation, evaluation is described as the interpretation of information gained from the results of tests or other assessment procedures, and it is related to decision making (Bachman, as cited in Brown, 2010) whereas measurement refers to the quantifying observed performance of students either quantitatively or qualitatively (Brown, 2004).

In the educational practice, it is referred to as a continuous process consisting of a variety of techniques such as appraising and commenting on students' response and performance (Brown, 2004). It has been also defined as a broad term including activities carried out by teachers and learners with the aim of gathering information, and diagnosing and informing learning process (Boston, 2002; Black & Wiliam, 1998; McMillan, 2007; Tombari & Borich, 1999). Although the definition of assessment varies, it is broadly accepted that it is an indispensable element of teaching and it gives educators many opportunities such as measuring the level of skills, knowledge, and progress, diagnosing strengths and weaknesses of learners, and motivating them (Taras, 2005; Wojtczak, 2002). Besides, information gathered from assessment can be useful for teachers in providing constructive feedback on students' learning in terms of what and the extent to which they are learning, and reflecting upon their own teaching to deliver more effective instruction.

### The Importance of Assessment in Education

Assessment, whether formal or informal, goes parallel with teaching and learning which are complementary to each other. Therefore, assessment is an indispensable part of curriculum and education system. Mounting research has

examined the necessity and role of assessment in education focusing on different dimensions in that some research focus on the role of assessment in learning whereas others emphasize assessment from the perspective of teaching together with the administration and policies.

With respect to learning, assessment is considered as a long term procedure, which takes place during the instruction and promotes lifelong learning (Mussawy, 2009). Enhancing learning is one of the aims of assessment, which calls for improvement in the assessment system since assessment directs learning. Besides, through effective assessment, students are provided with the knowledge of their own progress in terms of their strengths and weaknesses that should be improved, and creative tension with the aim of motivating them to study much (McKay, 2006).

Similarly, as teachers are engaged in many assessment-related activities during their classroom practices, and make decisions regarding progress of students and the effectiveness of their teaching, assessment plays a fundamental role and functions as a feedback on their own teaching. Therefore, this feedback informs the following teaching decision and guides them regarding teaching-related issues (Conde & Barragán, 2014). On the other hand, as one of the important considerations of administrators, and policy makers as well as learners and instructors is the outcomes of learning in terms of what learners can achieve after learning, assessment is necessary for revealing this information to policy makers and administrators (Mikre, 2010). Taking all of these into account, effective assessment does not only benefit teachers and students, but also the policy makers and administrators as well as the curriculum, and it functions as a helpful tool for giving shape to learning and teaching together with the curriculum (Sah, 2012).

With regard to the influence of testing on education, it is regarded as washback or backwash, which becomes an essential and widespread concept in education. The literature also shows that it is sometimes used interchangeably with the term of "test impact" (Cheng, Watanabe, & Curtis, 2004). With the aim of distinguishing these two terms regarding the scope, Wall (1997) stated that the latter means any effect of the tests that can take place on individuals, society, policies, and educational system whereas the former is described as the influence

of tests on teaching and learning either positively or negatively. Provided that influence is positive, it is called positive or beneficial washback. However, if this is negative, it is referred to as negative or harmful washback (Arshad, 2004).

Providing that teaching and learning is regarded while designing assessment, it will motivate students better (William, as cited in Mussawy, 2009, p. 7). Moreover, a variety of factors such as the way the assessment is constructed, designed and used well as well as the procedures of administrating results in positive effects on education. To illustrate, positive washback is stated to occur when assessment is constructed to reflect the content and skills that are taught during the instruction (Bachman, 1990).

On the other hand, assessment may have negative effects on learning and teaching stemming from a variety of reasons. Among these, teaching to the test, which results from much emphasis on test results and fear of poor results, influences instruction and learning of students in many aspects. It makes teachers narrow down the curriculum and focus on points of knowledge and skills that are tested, which leads to ignorance of other required knowledge and skills. Therefore, students may just learn concrete parts of knowledge that are tested, and may not learn real-world knowledge. Another aspect is related to the motivation and psychology of students in that their learning motivation can become tests instead of learning itself, and they can experience test anxiety that may prevent them from realizing their potential. Similar to students, tests can also bring anxiety to teachers and affect their motivation and priorities in terms of passing topics that are not directly addressed in the exam (Yi-Ching Pan, 2009). In the same vein, if the curriculum is driven by the assessment like in many cases of high stakes testing, it can result in negative washback. Another possible negative washback can take place when there is a distinction between the emphasized goals of instruction and the focus of assessment (Wall, 2012), and when tests are not related to the learning basis and course goals (Cheng, 2005). If there is such a difference, students may also experience lack of motivation to study and regard instruction as not serving for their success and preparing them for real life, which also results in negative backwash.

As for the solution for preventing negative washback and promoting positive washback, more learner centered approach can be adopted (Prodromou, 1995),

and teaching and testing can be linked to instructional objectives. If tests reflect the objectives and goals of the instruction together with a variety of activities to teach the content, positive washback can be achieved. Therefore, it is important to plan the course and assessment at the same time (Rogier, 2014). Similarly, Pan (2009) argues that it is crucial to provide a correspondence between what is assessed and what is taught, which can be achieved through direct testing and making students aware of the test. As for another aspect that can promote positive washback, it is feedback in that it gives learners necessary information about their progress and benefits them when it is given on time. For instance, if the teacher administers a quiz and uses it as a tool for revealing progress of students through feedback on time, it creates a positive washback in terms of giving guidance to students and teachers for further study (Rogier, 2014). In conclusion, when the role of assessment in education along with the effect of assessment on it is taken into consideration, aligning assessment and tests with instruction and learning is crucial for contributing to the education system.

**Purposes of Assessment**

The literature on the purposes of assessment shows a variety in terms of the way they are classified. While many researchers refer to these in a general framework, others categorize these purposes separately for teachers, students, and society. Among these classifications, Hyland (2010) suggests five main purposes of assessment, which are placement, diagnostic, achievement, proficiency, and performance. To start with the first purpose, which is placement, it is highlighted as placing students into appropriate classes and levels as a result of the information obtained from tests. The second purpose is to diagnose students' weaknesses and strengths with the aim of modifying course and giving students feedback on their progress. As for the third purpose, which is achievement, it provides information about what has been learned after the instruction, which also demonstrates progress. While the fourth purpose is proficiency that shows students' level of performance and gives general view of ability, the last one is performance that gives insights into learners' skill to carry out a given task.

On the other hand, Berry (2008) mentions that there are four purposes of assessment which are referred to as four roles. Among these roles, the formative

and the summative roles respectively refer to providing feedback on progress of learners to teachers and learners, and giving information about the level of learners' success during and at the end of the school. As for the other two roles, which are the certification and evaluation role, they respectively point to providing opportunities to select through qualification, and making contributions to the information about the effectiveness and quality of both people and the system. Liljedahl's (2010) categorization, which includes four purposes that are communicating, valuing what teachers teach, reporting out, and not ranking, shares some similarities with that of Berry (2008)'s in terms of the purposes of giving information to learners and instructors, and evaluating. However, the categorization of Liljedahl (2010) also incorporates two other dimensions of purposes of assessment, which are reporting this information to stakeholders other than teachers and students, and not sorting students according to assessment results.

As for the categorization of these purposes made separately for teachers, students, and society, it also includes almost all of the purposes forementioned for teachers and students; however, it also specifically mentions the purposes of assessment regarding society. Among these, credibility, accountability, and making the course or module trustworthy to other institutions are underlined (McCulloch, 2007). In spite of the variety of classification of purposes of assessment, it is commonly agreed that assessment aims to report and contribute to the teaching and learning process in many dimensions; in other words, it contributes to educational improvement.

In addition to the purposes of assessment stated above, the literature also reveals three purposes of assessment, which are assessment for learning (*AfL*), assessment of learning (*AoL*), and assessment as learning (*AsL*) (Conde & Barragán, 2014). To start with the first term, which is assessment for learning, it indicates that assessment takes place throughout the process of learning, which provides information about students' progress towards achieving goals. In this regard, it can be employed to investigate what learners have learnt and can do as well as the areas that require more attention; in other words, it enables students to get constructive and descriptive feedback about their progress, and instructors to vary and improve their approaches, methods, techniques and materials with the

aim of moving student learning forward (Wood, 2007).  That is, it provides the basis for informing teachers about the progress of learning and meeting the learning needs of students (Stiggins, Arter, Chappuis, & Chappuis, 2004).

As for the AoL, it means using assessment to prove what students learned, whether goals and objectives are achieved or not, and determine students' future program or placements. In other words, it is prepared with the aim of giving proof of achievement of students to themselves, teachers, and institutions (Wood, 2007). Therefore, AfL and AoL differ in aims in that the former is built to supply information to students, teachers, and parents on supporting learning, not to others outside the classroom whereas the latter is constructed to provide information on achievement of students to administrators and policy makers with the aim of reporting for accountability (Reeves & Conley, as cited in Wood, 2007, p.18).

With regard to AsL, it highlights assessment as a process of improving and supporting metacognition of students. It also regards students as connector between learning and assessment in that as long as learners are actively engaged assessors, they have the opportunity to use information obtained through assessment to make connections with their schemata and to utilize it for further learning. In other words, within this framework, students monitor their own learning and determine the way that they will use assessment for their new learning (Gurski, 2008).

## Functions of Assessment

With regard to the assessment process and how it is used, two main axis regarding its objectives are identified; in other words, two functions are described in the literature. These are formative and summative assessment (Brown, 2004). There is a considerable amount of literature focusing on the importance and differences of these functions; therefore, it is crucial to identify and distinguish these two functions.

To start with the formative assessment, it is identified as taking place at the beginning or during teaching, and gathers data about learning to make changes to teaching process and give instant evidence for student learning (Cheng, Rogers, & Hu, 2004; Conde & Barragán, 2014). It can take place as both formal and informal

16

tools such as quizzes, self-assessment, self-reflection, observation, and think-aloud.

As for the purposes of formative assessment, various purposes have been put forward. In the first place, formative assessment aims to improve student learning without including evaluation and grading in terms of enabling teachers to determine the knowledge and skills that are not learnt well, and guiding them to take actions while the course goes on. Secondly, it provides programs with crucial information on the course in that it enables them to investigate whether learning goals and objectives are being achieved (Conde & Barragán, 2014). If learning goals and objectives are not met by the current instruction according to formative assessment, programs may reflect on possible adjustments to enhance the quality of instruction and to achieve these goals and objectives while the course is in progress. As Wiliam and Thompson (2008) state, the extent to which information gathered from the assessment is reflected in the system and used to improve it determines whether the assessment is formative or not. Therefore, it is imperative to use the information that formative assessment reveals for the benefit of improving the quality of student learning and instruction.

As to the summative assessment, a considerable amount of research has described it as taking place at the end of the unit or term, and gathers data about student learning to inform and judge knowledge and skills of students (Cheng, Rogers, & Hu, 2004; Taras, 2005). Brown (2004) also argues that it is used after the course is over; therefore, it is not used as a component of the course, but as an isolated activity that assign grades to students. To illustrate, end of unit tests, final exams in a course and general proficiency exams can be used as summative assessment.

With regard to the purposes of summative assessment, a number of purposes have been suggested. To start with the first, what students learnt and did not learn after the course or program are aimed to be determined through summative assessment. In other words, it aims to gauge the level of students' learning according to content standards, which also provides accountability (Conde & Barragán, 2014). Secondly, they serve as tools to evaluate the quality and effectiveness of program or course, improvement goals, alignment of the curriculum as well as student placement in certain program ("Garrison &

Ehringhaus", n.d.). More specifically, summative assessment conducted to check whether students have met curriculum goals and objectives (Bardes & Denton, 2001), and following this, administrative decisions are made like making changes in the course/program and planning for the next courses or programs.

All in all, formative and summative assessments, which are the functions of assessment, differ from each other in their design, scope, purpose, result and direction that they give to teaching and learning processes. In this regard, it is crucial for teachers, administrators, program and policy makers to be aware of these two terms, and the way they function in that they can employ the appropriate assessment tools in accordance with their educational context and purposes, which benefits education system all in all.

**Types of Tests and Testing**

There are different types of tests which can be employed for different purposes (Bachman & Palmer, 2000; Brown & Abeywickrama, 2010; Hughes, 2003). In other words, different kinds of tests can be utilized depending on the purpose of assessment. In this regard, teachers should be knowledgeable in these test and testing types including the purpose they serve for, the differences between them, and the situation they should be used so that they can determine the most appropriate type of test and testing for their context.

As for these types of tests and testing, different categorizations are made in the literature (Brown, 2004; Hughes, 2003; Madsen, 1983) and this study is basically built on the categorization suggested by Hughes (2003) while it has implications of the classification suggested by Brown (2004). That is, the study also draws on one element of the Brown's classification, which is language aptitude test. To start with the classification of types of tests, Hughes (2003) pointed to four types, which are achievement tests, proficiency tests, placement tests, diagnostics tests, and Brown (2004) included language aptitude tests as a fifth category. As for the types of testing, Hughes (2003) put forward the necessity of distinguishing between them. Each of them is explained in detail below.

**Types of Tests.**

*Achievement tests.* Achievement tests are utilized to measure how much each student has learned and achieved in a classroom lesson, unit or curriculum. In other words, they aim to identify whether required knowledge and skills are attained, namely course objectives, after the instruction. In this regard, it is crucial to note that these tests are limited to certain material covered in a course or curriculum, and they are administered at the end of the course or instruction (Brown & Abeywickrama, 2010). As they are administered at the end of a unit, lesson or course, they are often summative. However, they have also a formative role in that they reveal the progress of students and give feedback on students' performance in the unit or course. In accordance with this feedback, future learning and teaching may be planned (Brown, 2010; Madsen, 1983).

Hughes (2003) states that there are two types of achievement tests, which are final achievement tests and progress achievement tests. While the former one aims to measure and evaluate students' progress at the end of the course, the latter aims to measure the progress of students. In this sense, well-defined short-term objectives for progress achievement tests may be established, which is helpful in making progress towards the final achievement test. To put it another way, if these objectives fit in the syllabus and teaching, then progress tests also fit in what has been taught.

*Proficiency tests.* Contrary to the achievement tests, proficiency tests aim to measure the overall ability in a language regardless of whether training is received or not. In this sense, proficiency tests are not built on the objectives or content of the language courses, but built on the description of what test-takers can do in the language to be regarded as competent (Hughes, 2003). TOEFL, Cambridge FCE and CPE are among the examples of proficiency tests.

Proficiency tests are stated to be almost always summative in that results of the test are given as a single score and they are not used for giving diagnostic feedback (Brown, 2004). For instance, in most settings such as getting into the university, taking the proficiency test is obligatory in terms of eliciting the students' level of knowledge and skills in the language. The results of this test are used to

determine whether students can pass to the next stage or not; therefore, they don't function as formative or diagnostic tools.

*Placement tests.* The aim of the placement tests is to place test-takers into the levels or sections of the programs or courses that are most relevant and appropriate to their abilities (Brown & Abeywickrama, 2010; Hughes, 2003). A placement test may consist of a sample of material to be addressed in a number of courses within a curriculum. At this point, it is noted as important that performance of test-takers has to reveal the point at which the material is not either too easy or too challenging for test-takers (Brown, 2004). It is also pointed out that if these tests are designed for specific contexts, they can achieve their purposes successfully. In this respect, it is important to consider the requirements of the program or course while constructing them, and they may be designed in many ways like including written and oral performance to measure comprehension and production, multiple choice questions, and open-ended responses.

*Diagnostic tests.* Diagnostic tests aim to identify areas of a language that the student should improve or the course should involve. In other words, they are utilized for determining the strengths and weaknesses; in this respect, diagnostic test is a part of formative assessment. Besides, they can be used to assess language skills in isolation (Hughes, 2003; Smith & Cumming, 2009). To illustrate, a pronunciation test may be administered and identify the difficult sounds for language learners. Accordingly, these sounds may be covered as part of the course. All in all, diagnostic tests serve as a helpful tool for improving learning and teaching.

*Language aptitude tests.* As to the language aptitude tests, they are used with the aim of measuring language learners' capacity and general ability to learn that language. In other words, it is used to predict the language learner's success before being exposed to the language.  However, there is not any research revealing that this kind of test is successful in predicting the communicative ability in a language (Brown, 2004), and these tests are not used widely. Among the very few examples are MLAT and the PLAB that include several English language tasks.

**Types of Testing.**

***Direct versus Indirect Testing.*** When the test-takers are required to perform exactly the skill that is intended to be measured, testing is stated to be direct. To illustrate, if the aim is to assess test-takers' writing skills and if they are required to write a composition, then direct testing takes place. Similarly, direct testing occurs if the students are required to speak with the aim of assessing their pronunciation of the language. In this regard, it has a number of advantages (Hughes, 2003). In the first place, as practicing for the test includes practice of the skill that is aimed to be improved, it can have beneficial washback. In the second place, forming the conditions that reveal the behavior on which the judgments are made is more straightforward provided that the skill or ability that is aimed to be assessed is determined well. Besides, assessing and interpreting the performance of test-takers', especially in writing and speaking skills, is also very straightforward.

On the other hand, indirect testing is measuring the abilities underlying the skills that are intended to be measured. For instance, one part of the TOEFL requires students to determine the word which is not appropriate in formal English or contains error. By doing so, test-takers' writing skill is aimed to be measured indirectly. As for another case for indirect testing, it is testing pronunciation ability of test takers' by requiring them to distinguish rhyming pairs of words in a paper and pencil exam (Lado, 1961, as cited in Hughes, 2003). Hughes (2003) also noted that, it is preferable as it appears to provide the chance of testing representative sample of a number of abilities contrary to the direct testing which is restricted to a very limited sample of tasks. In this sense, indirect testing is superior to direct testing as the results of it are more generalizable.

***Discrete point versus Integrative Testing.*** Tests may be created on the presumption that language can be broken into its constituents and tested separately, and administered by testing one element at a time, namely, item by item, which is referred to as discrete-point testing. These separate parts are the skills of speaking, listening, writing, and reading, and many units of language such as morphology, phonology, lexicon and discourse. On the other hand, tests may be constructed on the assumption that language is a whole unit and tested by combining different skills together. In other words, tests may require test-takers to combine many language elements to complete the test task, which is referred to as integrative testing (Brown, 2004; Heaton, 1988; Hughes, 2003; McNamara,

2000). To illustrate, cloze test is an example of integrative testing in that it requires test-takers to use different abilities and knowledge together such as knowledge of vocabulary, grammar, discourse and reading skills. Similarly, dictation requires students to use many skills and knowledge together like listening, spelling, and writing. In short, these two types of testing differ in their assumptions about whether language can be divided into its parts and tested separately or not.

*Norm-referenced versus Criterion-referenced Testing.* Tests are also different in comparing the scores of test-takers according to a certain criteria or average scores. In norm-referenced testing, test-takers' scores are interpreted in accordance with a mean, median, standard deviation, and percentile rank (Brown, 2004); to put it another way, each test-taker is compared to other test-takers (Bachman & Palmer, 2000; McNamara, 2000). On the other hand, in criterion-referenced testing, there is a predetermined and defined score, which gives feedback to students usually in the form of grades. Therefore, scores of students are not compared to others'. In this respect, criterion-referenced testing provides direct and relevant results for students as it reveals their level of knowledge and skill in language sufficiently (Hughes, 2003).

*Objective versus Subjective testing.* The difference between objective and subjective testing is the methods of scoring. If scorers are not required to make any judgment, then the scoring is objective; on the contrary, if judgment is required, the scoring is stated to be subjective. To illustrate, teachers can evaluate some kinds of tests like multiple choice without making any judgment whereas they may make judgment in the evaluation of some kinds of tests like writing tests and open-ended questions (Hughes, 2003; Madsen, 1983; McNamara, 2000).

**Language Testing and Assessment**

A growing body of literature investigating the role and importance of language testing and assessment has revealed that it plays a prominent role in the process of language education (Elshawa, Heng, Abdullah, & Rashid, 2016; Prapphal, 1990). In the first place, it enables language teachers and administrators to determine strong and weak sides of language learners, and to take actions on the areas that need improvement. Secondly, it is important in that students can be placed into appropriate courses regarding their levels. By doing so, learners are

given opportunity to get instruction and make progress regarding their current levels. Besides, effective LTA can be supportive in the process of planning and administering language courses and programs, which is really significant in that planning determines whether language programs/courses result in achievement or failure (Prapphal, 1990). Supportively, LTA is considered as a helpful tool for testing or assessing learners' language ability, and provide goals with regard to language teaching and learning (Bachman & Palmer, 1996). All in all, language assessment and testing is an indispensable component of language teaching and learning both in theory and in practice taking the critical role and importance of it into consideration.

The literature also reveals that LTA have undergone enormous expansion in the first decade of the 21st century in parallel with the growing importance of testing and assessment all over the world. Besides, it is stated that the field of LTA has also witnessed an increasing professionalism affected by both within and outside of the field, and it has moved to center stage from periphery in recent years for many reasons such as educational, employment, social and sociopolitical factors (Bachman, 1990). Recent developments and professionalization of this field have led to an increased interest in adequate knowledge of language testing and assessment and effective practices (Taylor, 2009). Therefore, it is critically fundamental, especially for language teachers, to have knowledge and practice of LTA.

In addition to these, in parallel with the relationship between LTA and teaching and learning processes, new theories have been put forward about language learning and teaching. These also led to changes in testing applications (Spolsky, 1995), and LTA trends and practices are stated to follow this shift of methodology historically (Brown, 2004). To start with the era of behaviorism in the 1950s, several presuppositions of behaviorism led to consequences for testing. According to behaviorism, tests should be utilized to check mastery prior to moving to next objective, and they are the direct evidence of learning goals and objectives. Moreover, tests were needed at the end of every lesson so that mastery could be achieved through reteaching and learning failures could be prevented, all of which shows the central role of testing in behaviorist theory (Shepard, 2000). As a result of behaviorism, language testing concentrated on

specific language elements like grammar, phonology, and lexicon, and accuracy was emphasized. More specifically, discrete-point testing, which assumes that language can be broken into its small parts like morphology, phonology, syntax, discourse and lexicon and these parts can be tested separately and adequately, reveals the effect of behaviorism along with structuralism (Brown, 2004).

As for the following era that emphasized communication, authenticity and context, new approaches were delved as behaviorism did not correspond to this era. To illustrate, discrete point testing was regarded as incomplete and decontextualized on the ground that language competence cannot be tested separately and contextualization was needed for the sake of learners. The fact that behaviorism did not get into this new era brought on the advent of communicative language teaching in the late 1970s, and with this development, a need arouse for inventing new theories of language testing. With the aim of overcoming weaknesses of discrete point testing, integrative tests were used such as cloze tests and dictations, and supporters of it emphasized unitary trait hypothesis. According to this hypothesis, language proficiency; namely, four skills of language along with other components such as phonology, grammar and lexicon should not be separated from each other. However, others were still against this hypothesis by arguing its inaccuracy. The field of language testing disregarded these arguments and conflicts, and adopted communicative language testing till the mid-1980s. (Brown, 2004). Within this new development, tests are required not only to assess learners' knowledge of the language, but also to test their use of language in communicative and authentic situations (Fulcher & Davidson, 2007), which shows the importance of performance of learners. According to communicative language testing, four principles are fundamental, which are starting from somewhere, concentrating on content, bias for best, working for washback. While starting from somewhere refers to establishing assessment criteria, concentrating on content pays attention to both topics and task types in terms of considering the age, level, needs of learners. As for the last two principles, bias for best requires test takers to be familiar with and well prepared for the test whereas the last principle describes tests as tools for teaching and feedback ("Nguyen & Le", n.d.).

All in all, these changes and developments in language testing in parallel to theories and developments in language teaching and learning is fundamental and should be considered by language teachers, program and policy makers, and test

administrators while constructing, designing, using, adapting, and interpreting the tests. Above all, being aware of these developments and trends in language testing is critically imperative for language teachers in that grasping the underlying theoretical background and developments of testing contributes to their perspectives of language testing and actual practices.

**Five Principles of Language Assessment**

Defining the purposes and functions of tests, and choosing the most appropriate test types should be done prior to conducting a test. However, they are not enough to get an effective test. In this regard, teachers should pay attention to principles of assessment, which are validity, reliability, practicality, authenticity, and washback. Each of them is explained in detail below.

**Validity.** As one of the keystones of any assessment, validity is concerned about the relation between the purpose and the form of assessment. In other words, a test that achieves the purpose of assessment by measuring the intended skill, knowledge, ability or components is valid (Coombe, Davidson, O'Sullivan, & Stoynoff, 2012; Stobard, 2012).

Brown (2004) also states that a test should have some kinds of evidence to obtain valid inferences from tests. To start with the content-related evidence, also known as content validity, it is the adequacy of a test regarding its content or the sample of behavior it intends to measure. To exemplify, if a test aims to assess test-takers' ability to speak in a conversational context, and requires them to speak in authentic setting, it achieves content validity. On the other hand, if the test does not require test-takers to speak, but to answer multiple choice questions, it lacks content validity. In this regard, content validity is crucial for tests to be accurate (Coombe et al., 2012; Hughes, 2003).

In the second place, criterion-related evidence, also referred to as criterion-related validity, relates to the degree to which the criterion of the test has been achieved. To put it another way, results of tests are compared to those of some independent and highly dependable assessment; for instance, the result of a teacher-made test about present modals is compared to the result of a test of the same topic in a textbook. If both of them give the necessary degree of correlation coefficient, then criterion-related validity is achieved. Indeed, there are two types

of criterion-related validity, which are concurrent and predictive validity (Coombe et al., 2012; Fulcher & Davidson, 2007; Hughes, 2003). While the former refers to administering tests at about the same time, the latter refers to administering them at different times. Besides, the former aims to measure concurrent ability; on the other hand, the latter aims to assess and predict test-takers' possibility of future performance. To give an example of concurrent validity, the validity of high scores on the final exam of English course will be verified by actual proficiency in the language. As for the predictive validity, if a proficiency test, which is administered to students and they get high scores, predicts students' ability to handle a course at university, it reveals that it has predictive validity.

A third kind of evidence is construct-related evidence, which is commonly referred to as construct validity. For instance, communicative competence, fluency and proficiency are among the linguistic constructs. It may sound hard and impossible to measure for teachers; and indeed it is not mandatory for classrooms. Even so, through considering the degree to which a test corresponds to its theoretical foundations, construct validity is stated to be achieved (Brown & Abeywickrama, 2010). As for the consequential validity, it comprises all the consequences of a test. The importance of the consequences of using an assessment is highlighted in the literature; in other words, the research shows that *consequential validity* has an important role in education (Brindley, 2001a; McNamara, 2000). To illustrate, consequential validity is essential for teachers in that if teachers get test scores and interpret them with the purpose of improving learning instead of just judging, the test may result in positive consequences. Similarly, it is significant for stakeholders in terms of benefiting them to develop the curriculum. Accordingly, it is essential to take it into consideration during the process of test construction, administration and interpretation.

With regard to the face validity, it relates to the extent to which assessment is regarded as fair, related, and beneficial for the improvement of learning (Gronlund, 1998). In this regard, it is fundamental for students' motivation and performance on the test. That is, if the test does not have face validity, it might not seem to be valid or acceptable to students or teachers, which may result in low motivation or poor performance. To put it another way, they may not reflect their true abilities in test performance if they regard the test as not achieving face

validity. As a result, teachers should consider the likely effects of *face validity* of tests on students' performance, and construct or rearrange their tests accordingly. In conclusion, validity is one of the cardinal components of teachers' grasp of what makes an effective and accurate test, and using direct testing, scoring according to the purpose of the tests, and achieving reliability may enable tests to be more valid as Hughes (2003) put forward.

**Reliability.** As an another keystone of assessment, reliability is concerned with consistency and dependability of tests, and referred to as the extent to which a test yields similar and consistent results when administered in different times. An ideal assessment or test should be independent of who makes scoring, and when and where it is conducted. However, it is crucial to point out that the time interval between the administrations of two tests should not be either too long or too short (Brown & Abeywickrama, 2010; Hughes, 2003). If this interval is too long, it is highly likely that students will gain new knowledge or skills, which will influence the reliability of the test. Similarly, if the interval is too short, it is highly probable for students to remember the test items, which also affects the reliability of the test. In this regard, it is essential for teachers to take this into consideration while administering tests.

As to the factors that affect reliability, they are categorized into four, which are student-related reliability, rater reliability, test administration reliability, and test reliability (Brown, 2004; Heaton, 1975). In the first place, some student-related factors such as temporary illness, anxiety, fatigue, and motivation can prevent students from performing their true performance. In addition to these, students' "test- wiseness" and strategies for effective test taking can make observed score diverge from test-taker's real score as Mousavi (2002) stated. In the second place, some rater related factors such as subjectivity, bias and human error may affect scoring. In this regard, it is important to note that inter-rater and intra-rater reliability may occur. The former takes place when two or more than two scorers are involved in scoring procedure, and give inconsistent scores of the same test stemming from many factors such as not paying attention to scoring criteria, inattention, inexperience, and bias. On the other hand, the latter is a prevalent situation for classroom teachers and may be affected by some factors like ambiguous scoring criteria, tiredness, carelessness, and bias. To illustrate, when a

teacher is required to read 50 tests in a week time, there may be problems in scoring and reliability. Therefore, it may be effective to read through half of the tests before giving any score, and returning back to the whole tests, which may be beneficial to ensure fair judgment and scoring as Brown (2004) proposed. Besides, determining the scoring criteria beforehand, and using rubrics may be helpful for achieving rater reliability. As for the other factor that affects reliability, it is test administration reliability in that some conditions may lead to unreliability such as differences in temperature and the amount of light in different parts of the same room, conditions of desks and chairs, noise, and variations in photocopying. For this reason, it is essential to consider these possible conditions and the possible effects of them on students' performance in administering the test. As to last factor that affects reliability, test reliability is concerned with the nature of the test itself. In other words, some features of test may lead to unreliability such as the length, structure, and items of tests. For example, if the test is too long, students may be tired towards the end of the test, which may hinder them from showing their true performance. Similarly, if the items of the test are not well constructed and ambiguous, it may affect their performance poorly. All in all, these four factors may influence consistency and dependability of tests to much extent; therefore, they should be considered by instructors and test administrators while interpreting the reliability of tests (Cohen, 2001).

Hughes (2003) states that there are three strategies to obtain test reliability, which are test-retest reliability, split-half reliability method, and parallel forms reliability. To achieve test-retest reliability, the same test should be given to the same group of students; on the other hand, testing the same group with the same questions in the second time may not be preferable for students and teachers. For that reason, split half reliability method can be utilized in terms of dividing the test into two separate parts and giving scores for both parts. Through this way, teachers can obtain test-reliability. As for the parallel forms reliability, two parallel forms of the tests that have a number of common points are prepared, and the correlation between the scores of these two parallel tests is analyzed and used as a measure of reliability (Bachman & Palmer, 2000; Fulcher & Davidson, 2007).

**Practicality.** In addition to the keystones of assessment, which are reliability and validity, an effective test should be practical. The test should be

practical in terms of cost, time, and energy (Brown & Abeywickrama, 2010). In other words, the test should be efficient in constructing, using, and evaluating. If the test is extremely expensive, it is impractical. Likewise, if a final exam takes a student six hours to complete, it is also not practical since it takes more time than required to accomplish its objective. Therefore, it is important to scrutinize this principle in designing tests.

**Authenticity.** The fourth principle is authenticity which points to encountering the test task in the real world. If the test task fails to reflect real-world task, authenticity cannot be achieved. To achieve authenticity, Brown (2004) put forwards some strategies like using natural language as much as possible, contextualizing items of the test and selecting meaningful and relevant topics for test-takers.

**Washback.** The final principle is washback which points that tests can affect teaching and learning in a positive or negative way. Provided that the effect is positive, it is called positive or beneficial washback. It may take place when the assessment is designed, administered, and used for improving learning. However, the learning motivation of students may become test itself instead of learning, and they can just focus on passing it with high scores. In this situation, it has negative effects on learning, which is called as negative or harmful washback (Arshad, 2004). According to Heaton (1975), the effect of washback may be categorized into two, which are macro and micro aspects. The former refers to impacts of test on society and education system like the development of curriculum whereas the latter refers to the impacts of test on individual student and teacher like improvement of teaching and learning process.

Considering the importance of washback on education, assessment should be planned well. To this end, direct testing may be used to assess and evaluate learners' language skills instead of indirect testing. For instance, if the aim is to assess students' writing skills, the assessment should include writing task directly. Besides, criterion-referenced assessment may be used to reveal the level of students' knowledge and skills without determining their success according to their peers' success. By doing so, a reliable picture of the outcomes of the process of learning and teaching can be obtained. In addition to these, providing positive reinforcement like praising students' correct answers may motivate students and

result in positive washback. On the other hand, providing feedback on wrong answers of students instead of just giving scores provides positive washback in that students reflect on their weak points and improve these areas (Brown & Abeywickrama, 2010; Hughes, 2003).

**Alternatives in Language Assessment**

Assessment has witnessed a paradigm shift from a testing culture to an assessment culture in the last decades, especially when the traditional tests have been questioned whether all people and skills could be measured through them (Brown & Abeywickrama, 2010; Lynch, 2001). As a result, alternative assessment, which emerged as a new concept, has become prevalent.

Alternatives in assessment contribute to teaching and learning in that language teachers get a more exact view of students' achievement and may adapt their instruction in accordance with the feedback on students' learning and progress. Besides, through a number of assessment opportunities, different learning styles can be addressed; for instance, some students may be required to write about the topic while the others may be interviewed. All in all, alternatives in language assessment benefit language teachers and learners when designed and utilized in a planned way. As for the examples, portfolios, journals, conferences and interviews, observations, and self and peer assessment may be cited among the examples of alternatives in assessment. Each of them is explained in below.

**Portfolios.** Portfolios provide concrete evidence of students' progress, efforts and achievements in language learning over time. As for the importance of portfolios, they offer various benefits. In the first place, as portfolios reveal students' progress and achievements, instruction may be reshaped according to the needs and weaknesses of learners. In the second place, they foster intrinsic motivation of students by giving the feeling of responsibility and ownership. Moreover, teachers and students communicate with each other more during portfolio development, which facilitates and strengthens the interaction between them. In this respect, the use of portfolios with language learners is extremely beneficial, and language teachers should use it to improve the effectiveness and quality of education.

It is also crucial to point to the European Language Portfolio (ELP). It has been developed by the Council of Europe as a tool for supporting the development of plurilingualism and pluriculturalism (Council of Europe and European Union, 2001), and used widely in Europe since 2000 (Council of Europe, 2011; Little, 2005; Mirici, 2000). It has three obligatory components which are a language passport, a language biography, and a dossier.

ELP serves two functions, which are pedagogical and reporting functions (Council of Europe, 2000). To start with the pedagogic function, it enhances the motivation of learners and enables them to reflect on their own learning and achievement, and to improve plurilingual and intercultural experience like through reading and projects. Shortly, it supports learner autonomy, plurilingualism, and intercultural awareness and competence. As for the reporting function, it documents learner's plurilingual language proficiency in a comprehensive and reliable manner. As it is based on the Common European Framework (CEF) Reference Levels, the levels of competence of learners can be informed to others in an internationally comparable way. Shortly, this function of ELP enables users to record their language learning accomplishments together with their experiences of learning and using languages. Considering the importance and functions of the ELP, especially its function for developing learner autonomy and responsibility, intercultural awareness and competence, teacher candidates should be trained in why the ELP is important, what opportunities it provides, and how it is used.

**Journals.** Another alternative in assessment is journal that is referred to as a written record of person's emotions, beliefs, ideas or improvement towards targets without paying much attention to structural components or accuracy (Brown & Abeywickrama, 2010, p.134). There are various categories and purposes of journals like response to readings, language-learning logs, self-assessment reflections, and grammar journals, and they serve as a helpful tool for students in that they may practice writing, use writing as a thinking process, and reflect on their learning process.

When language teachers plan to make use of journals, they should pay attention to some aspects like stating the objective or objectives of the journal, guiding students on possible topics, specifying the criteria for the assessment of

these journals, and providing formative and washback-giving feedback. Otherwise, journals may not serve for their purposes accurately.

**Conferences and Interviews.** As for the other kind of alternatives in assessment, conferences are usually used in language courses. Especially when the process approach to writing is adopted in writing courses, conferencing takes place through conversations on drafts and ways of improving them. However, conferences may also take place for various purposes such as reviewing journals, making suggestions about oral presentation of students, and setting goals for future learning. In this regard, the role of language teachers is prominent since they act as a guide and facilitator, but not as a test administrator.

With regard to the interview, it is described as a situation where a student is interviewed for previously designed assessment purpose (Brown, 2004). Through interviews, many purposes can be achieved like assessing speaking skills of students, finding out students' learning styles and evaluation of the course.

**Observations.** Making observations is an inevitable part of teaching process; to put it another way, most of the time teachers are engaged in observations whether intentionally or not. However, the observation that is stated to be part of alternatives in assessment refers to a systematic and planned procedure that records verbal and nonverbal behaviors of students. To elicit natural language performance of students, it is important not to make them realize that they are being observed or assessed as Brown (2004) put forward. Otherwise, they could experience anxiety and their performance may not reveal their true performance.

Another important point for teachers is to define the aim of their observation, and accordingly determine the way the observation will be recorded and the elements to be focused during the observation. To illustrate, if the aim is to observe the student's errors in the use of articles "the" and "a/an, the teacher may utilize rubrics or rating scales accordingly. By adhering to these important aspects, a language teacher may obtain a rich and various evidence of students' learning including their strengths and weaknesses; in other words, this concrete evidence plays a formative role and informs learning and teaching process (Maxwell, 2001).

**Self- and Peer- Assessments.** To start with the self-assessment, it functions as a helpful tool for learning and measurement, and it is regarded as a process of formative assessment in which learners make decisions about their own work and learning, and find out their strengths and weaknesses in their work, and go over it (Andrade & Du, 2007). Accordingly, the principles underlying self-assessment are autonomy and intrinsic motivation. As for the advantages of using self-assessment, there are various benefits of self-assessment (Spiller, 2012). In the first place, it promotes active student participation and responsibility, which facilitates more effective learning. In the second place, it improves the evaluative skills of students together with critical thinking skills. In this regard, it is essential for teachers to encourage students to acquire and develop necessary skills and judgments for self-assessment.

As for the peer assessment, it benefits learning and teaching in many ways such as giving an active role to students in learning process, improving students' work and promoting collaborative learning. With regard to this, peer-assessment is also a valuable assessment tool like self-assessment, and teachers should incorporate it into language courses like writing courses.

## Assessment of Language Skills and Language Content

Acquiring language skills together with language content is essential for language learners to be proficient and competent at that language. It is also essential for them to develop these skills through some ways like exposure to the language, instruction, and active use of language. In this sense, language teachers should facilitate the development of language skills: listening, speaking, writing, and reading, and content of language: grammar and vocabulary, and assess them on a regular basis in order to elicit information about their progress and achievements. In this sense, assessment is a valuable tool for informing the development of language skills and progress of language learners. Therefore, language teachers should be competent at assessing language skills and content, and design the most appropriate assessment tasks in accordance with their context. The assessment of language skills and language content will be explained separately in below.

**Assessing Language Skills.**

***Assessing Listening.*** Listening is an inseparable part of speaking in that whenever oral production takes place, aural comprehension also occurs. However, in some situations speaking is not required like listening to the radio, lecture or announcements. Similarly, assessment of oral ability may be regarded as impractical in some situations, and assessment of listening may be included for its washback effect on the improvement of oral skills and diagnostic functions (Hughes, 2003). Therefore, it is essential to address assessment of listening individually.

In this regard, identifying the types of listening and assessment tasks and procedures in accordance with these types is crucial. Brown and Abeywickrama (2010) classified listening into four categories. To start with the intensive listening, it is referred to as listening for the perception of a larger part of language, and it includes some types of assessment tasks like recognition of phonological and morphological elements, and paraphrase recognition. To illustrate, if the aim is to measure students' recognition of morphological pairs during conversation such as "–ed", a language teacher may use a recognition type of assessment. The second type of listening is responsive listening, which is described as listening to a short part of language such as a greeting and comprehension check in order to give a short response. Within this type of listening, some task types may be utilized such as open ended response to the question and appropriate response to the question. For example, if the aim is to measure the recognition of "wh-"questions, test-takers may be required to listen to the question and provided with a number of options to choose the most appropriate response. As for the third type of listening, which is selective listening, it refers to listening to stretches of conversation for scanning for specific information. Listening cloze, information transfer and sentence repetition tasks may be used in this type of listening. To illustrate, information transfer task may be used for measuring whether test-takers can transfer what they hear to some written and visual tasks like finding a figure in a picture and completing a chart. As to the last type, extensive listening refers to listening for developing a top-down and general understanding of spoken language. Therefore, extensive listening comprehension tasks may be used such as dictation and authentic listening tasks including note-taking and editing.

Since assessment tasks and procedures have a prominent role in determining the extent to which the purpose of assessment is achieved, language teachers should design and select them in accordance with the purpose of assessment.

***Assessing Speaking.*** As for the speaking skill, it is a productive skill and can be directly observed (Brown, 2004). Some factors can interfere with these observations such as the accuracy of test-taker's listening skill and the design of elicitation techniques, which influences the reliability and validity of observations and tests of oral production. Therefore, it is essential for language teachers to be aware of the types of oral production, and to design and select assessment tasks and procedures accordingly.

With regard to types of oral production, they are classified under five categories (Brown & Abeywickrama, 2010). To start with the imitative speaking, it refers to imitation of a word, phrase or sentence, and a word and sentence repetition task may be cited as an example for this type of oral production. In other words, test-takers may be called for to repeat given words or sentences, through which their pronunciation may be assessed. The second type is intensive speaking, which is stated as the oral production of short stretches of language. To illustrate, if the aim is to elicit students' oral language performance through short stretches of language, they can be called for to describe the given pictures. As for the third type of speaking, it is responsive speaking that consists of interaction and comprehension at a limited level. Question-and answer, and giving directions and instructions may be made use of within this type of oral production. For instance, test-takers may be required to describe the way to hospital from city centre, through which their oral performance is aimed to be assessed. With regard to the interactive speaking, contrary to the responsive speaking, it requires longer and more complex interaction. In this sense, interviews, role plays, games and discussions may be designed in accordance with the assessment context. If the aim is to provide an authentic context for oral performance, discussions through which the oral ability is assessed can be used. On the other hand, extensive speaking aims to involve complex and longer stretches of language like monologues, oral presentations and translation. Therefore, if the language teacher aims to observe students' oral language performance in a more extensive way,

students may be required to give oral presentations. All in all, knowledge of these types of oral ability and task types guides teachers in deciding how to assess oral performance accurately.

***Assessing Writing.*** As to the writing skill, it is an indispensable part of second and foreign language learning with its own features and principles; therefore, teaching and developing writing skills is imperative (Brown, 2004). In parallel with this, assessing writing skill is also crucial for informing the progress of students and improving their ability in writing accordingly. In this regard, the aim and the objectives of the assessment should be specified precisely and assessment tasks should be designed accordingly. In other words, language teachers should consider the types of writing, and design the assessment tasks and procedures in accordance with the aim and objectives of the assessment.

Regarding the types of writing, the first type is imitative writing and it requires test-takers to master the mechanics of writing skill such as spelling. Within this scope, spelling tasks, picture-cued tasks and matching phonetic symbols may be utilized if the aim is to find out mechanics of writing skill of test-takers. As for the second type, it is intensive writing that requires production of vocabulary, idioms and phrases in accordance with the context and correct grammatical features. Therefore, meaning and context are placed emphasis to some extent. For this type of writing, it is appropriate to use dictation, grammatical transformation and picture-cued tasks. For instance, teachers may call for test-takers to write short sentences to describe the given picture. With regard to the last two types of writing, they are responsive and extensive writing. While the former refers to performing at a limited discourse, and writing two or three paragraphs, the latter refers to performing writing up to the length of essay or project report. As for the assessment tasks of these two types of writing, paraphrasing, paragraph construction and guided question and answer tasks may be made use of. To illustrate, test-takers may be called for to write an essay on the topic of global warming, through which their writing ability can be elicited and assessed to much extent. All in all, the awareness of the types of writing and assessment tasks guides language teachers in the process of assessment of writing skill.

***Assessing Reading.*** Like listening skill, the process and product of reading cannot be observed directly; therefore, assessment of reading should be carried out through inference (Brown & Abeywickrama, 2010). On the other hand, assessment of reading does not only include the measurement of comprehension but also the strategies that test-takers use to achieve the comprehension. Therefore, assessment of reading tasks and procedures should be designed in a way that the aim of assessment can be achieved.

With regard to which assessment procedures and tasks will be designed, considering the types of reading performance is essential as in other three skills. Four types of reading are listed: perceptive, selective, interactive, and extensive. To start with the perceptive reading, it requires paying attention to the components of discourse such as letters, words and punctuation, and includes some tasks like reading aloud and picture-cued tasks. A language teacher may employ pictures in terms of asking students to find the sentence describing the picture, which allows the teacher to make inferences about the reading ability of students. As for the selective reading, it aims to elicit test-taker's recognition of lexical, discourse or grammatical features of the language, and includes tasks like matching, true/false, and editing. To illustrate, to assess linguistic competence in reading, an editing task may be utilized in which test-takers are called for to find the incorrect words among the given sentences. As it involves proofreading, which is one of the real life skills, it provides authenticity. On the other hand, interactive listening aims to make students interact with the text psycholinguistically, and includes larger stretches of the language like several paragraphs. Within this aim, some task types such as scanning, editing longer texts and ordering tasks. For instance, students may be provided with a text and some specific questions about the text like date, name and setting. Through this task, students' scanning strategy may be assessed. As for the last type of reading, extensive reading includes longer texts like essays, articles, stories and novels, and aims to elicit students' overall understanding of the passage. To this end, assessment tasks like skimming, summarizing and responding. To exemplify, students may be provided with some general questions about the passage and their skimming strategies may be assessed in this way. In short, these assessment tasks may be designed or

adapted by language teachers depending on the context and aim of an assessment.

**Assessing Language Content.**

*Assessing Grammar.* The knowledge of grammar is at the core of four language skills (Brown & Abeywickrama, 2010); therefore, it has a prominent role in language teaching. Regarding assessment, it is also at the core of language description and performance of students (Rimmer, 2006). To this end, it is essential to be aware of how best to assess students' knowledge of grammar.

With regard to assessment tasks for assessing grammar, they can be classified under three different categories, which are selected response, limited production, and extended production. To start with the selected response, they require test-takers to select the most appropriate response among the given responses, through which the knowledge of grammatical form or meaning may be measured. To exemplify, multiple-choice tasks, discrimination tasks and noticing tasks may serve for this aim. If the objective of assessment is to elicit test-takers' knowledge of grammatical meaning of modal verb "could", a passage including different usages of "could" may be provided to test-takers to discriminate between these meanings. As for the second type, limited production tasks call for limited language production on the part of test-takers. Within this purpose, short-answer tasks, gap-filling tasks, dialogue-completion tasks may be utilized. As to the last type, extended production calls for larger amounts of language production on the part of test-takers and elicits more authentic use of language. To this end, information gap tasks, role play and simulation tasks may be employed. In particular, information gap task serves a valuable tool for eliciting students' grammar and pragmatic knowledge as well as oral language.

*Assessing Vocabulary.* Vocabulary is the cornerstone of knowledge and use of language; in other words, vocabulary knowledge is at the center of second/ foreign language ability. Considering the importance of vocabulary knowledge in second/foreign language learning, how best to assess students' knowledge of vocabulary is also critically important. To this end, during the design of assessment of vocabulary, some steps may be taken into consideration such as determining the purpose, defining the construct and target words, and choosing

the mode of performance such as receptive and productive vocabulary (Brown & Abeywickrama, 2010).

As for the assessment tasks, they can be labeled as receptive vocabulary tasks and productive vocabulary tasks. To start with the first type, it requires recall of vocabulary and encompasses certain tasks such as vocabulary in one sentence context, vocabulary matching exercise and word association. To exemplify, word association may be employed to assess test-takers' vocabulary size in that they are required to find the word which is associated with the target word. As for the productive vocabulary tasks, they involve both recall and use of vocabulary. Within this aim, fill-in-the-blank and selective deletion cloze task types may be employed. For instance, test-takers may be provided with a text about education and some parts of the sentences may be deleted so as to elicit students' knowledge of vocabulary. All in all, language teachers should be aware of these task types, how they function and in which situations they may be employed so that they can achieve the purpose of assessment.

**Assessment Literacy**

In the literature, there are various definitions of assessment literacy. Among the many, it has been referred to as involving the understanding of sound principles of assessment and the proper use of assessment practices together with the knowledge of theoretical and philosophical foundations in measuring learners' learning (Popham, 2004; Stiggins, 2002; Volante & Fazio, 2007). As for the viewpoint of Newfields (2006), it is defined differently for a university student, a high school teachers and an expert test developer. It is the knowledge of knowing how to be successful in exams from the perspective of students while it is associated with grading students accurately and ethically on the part of teachers. As for the test developers, it includes all aspects of their profession. When these different viewpoints on assessment literacy are taken into consideration, the common point of them is that teachers should be aware of different purposes and functions of assessment and employ them accordingly (Volante & Fazio, 2007).

Traditionally, it was considered as choosing, designing and making judgments of test and assessment process along with scoring and giving grades in accordance with theoretical knowledge. However, the recent paradigms have

developed a more comprehensive understanding of this notion by considering the implications of assessment for teaching. According to Boyles (2005) and Stiggins (1991), this concept extends beyond the theoretical knowledge of choosing and designing appropriate assessment tools to encompass the skill to analyze evidence or data gathered from assessment with the aim of improving teaching. Similarly, Siegel and Wissehr (2011) put forward that for teachers to be assessment literate, not only understanding of the theoretical bases of assessment, but also practical skills and methods of assessment in classroom context are called for; therefore, assessment literacy is not restricted to knowing theoretical foundations of assessment and includes overall competence in all areas of assessment.

**The Importance of Assessment Literacy.** There is a considerable amount of research focusing on the reasons why being assessment literate is crucial and the importance of assessment literacy for teachers in the literature (Wilson, Floden, & Ferrini-Mundy, 2001; Coombe et al., 2012). To start with these reasons, Popham (2011) suggested that there are basically two reasons that require teachers to be assessment literate. The former reason for this requirement stems from the fact that today's teachers are evaluated basically according to their students' scores on accountability tests. It is an undeniable fact that teachers' success is evaluated based on students' performance on accountability tests, which becomes one fundamental reason why todays' teachers should be well-informed of the nature of large-scale educational tests and instruments being used to judge them as professionals. Especially in the era of accountability policies, teachers' assessment literacy should be given priority like instructional competencies. The latter reason is that when tests and assessment are constructed and administered well, it can be vital for improving students' learning, which is one of the aims of education system. Empirical and theoretical body of research reveals that combining assessment and instruction can create a huge power for improving learning outcomes besides engaging students more (William, 2011).

As for another perspective on the reasons for teachers to be assessment literate, Newfields (2006) put forwards three arguments, which are assessment's being a common aspect of most of the education systems, its' being required for

understanding much of educational literature, and its' giving teachers an opportunity to share their own classroom results with other people. When these five reasons are taken into consideration together, they form a plausible rationale for the need for todays' teachers and teacher candidates to attain and develop assessment literacy, and they suggest that assessment literacy is needed by teachers for their own long-term well-being and for students' educational well-being (Popham, 2009).

In addition to these, (Coombe et al., 2012) put forward the importance of being assessment literate by stating a number of reasons. In the first place, as teachers spend most of their time in assessment activities (Stiggins, 1995), they are required to be knowledgeable in theory and practice of assessment, which calls for assessment literacy. In the second place, an increased knowledge in assessment will enable teachers to comprehend the results of assessment better and communicate these to stakeholders accurately, which results in increased validity, reliability, and transparency of tests. Thirdly, teachers with sound base of assessment are capable of integrating assessment with their teaching, through which the quality and effectiveness of instruction will be improved. Besides, assessment is stated to be crucial for instructional process and students' education (Mertler, 2003, as cited in Coombe et al. 2012, pp.2-3; Popham, 2006); therefore, once assessment is accurate and effective, achievement of students is also improved.

In addition to these, assessment literate teachers design and administer assessment in a way that it fulfills learning goals and objectives, and reflects students' progress and performance (Mertler & Campbell, 2005; Stiggins, 2002). In other words, being familiar with assessment tools enables teachers to choose the most appropriate and effective assessment instruments for achieving these objectives, as Gottheiner and Siegel (2012) stated. Moreover, teachers with assessment literacy integrate their teaching with assessment in order to employ relevant forms of teaching (McMillan, 2000); therefore, instruction and assessment may inform and develop each other. Furthermore, considering the time devoted to assessment-related activities, the learning potential of assessment can be increased to much extent if teachers are assessment literate (DeLuca & Klinger, 2010).

In parallel with these, in modern knowledge-based communities teacher-administered and classroom-based formative assessment that provides valuable information is called for, and instruction is fundamental (Andrade & Cizek, 2010; Stiggins, 2005). Based on these, teacher assessment literacy has become a concern for teacher educators as Popham (2004) put forward, and they have to assure that teacher candidates are competent at assessment in terms of having the knowledge of purposes and functions of assessment, designing assessment for learning objectives and learners' levels, monitoring the progress of students towards the objectives, and utilizing assessment for improving learning. Moreover, considering that teacher candidates start their teaching after completing their teacher education programs, assessment education in preservice period has a prominent role in providing strong bases in assessment theory and practice and preparing them for real classroom assessment practices as DeLuca and Klinger (2010) put forward.

**Language Assessment Literacy**

In the literature, language assessment literacy has been defined in various ways. Among the many, LAL is referred to as stakeholders', especially language teachers', familiarity with the testing definitions, practices of measurement, and using it in practice (Inbar-Lourie, 2008; Malone, 2013; Stiggins, 2001; Taylor, 2009). It is also described as having a repertoire of competences that allow teachers to grasp, make judgment, and create language tests together with analyzing these tests in certain situations (Pill & Harding, 2013). In this sense, it involves the acquisition of a variety of skills regarding production, interpretation, evaluation of tests besides the critical comprehension of assessment roles and functions in education and community (O'Loughlin, 2013).

As to what constitutes LAL, Inbar-Lourie (2013a) describes two dimensions of the LAL construct, which are generic and specific. While the first dimension encompasses the knowledge base of assessment that is common in other aspects of educational testing, the second dimension entails the body of assessment knowledge which is called for the implementation and utilization in assessment.

Furthermore, Brindley (2001b) put forwards an outline for development in LTA including the knowledge components called for administering LTA, and this

outline offers a very useful framework in terms of including the components of LAL knowledge base. Drawing on Brindley's (2001b) outline, Inbar-Lourie (2008) analyzes and discusses the competencies of LAL, and states that it incorporates understanding the why, what and how of language assessment. To start with *the why* of language assessment, Inbar-Lourie (2008) argues that the importance of language assessment should be underscored and taken into consideration in the professional development of future language assessors. In this sense, understanding the reason and rationale underlying language assessment is crucial for language teachers in that it shapes their perceptions of assessment and assessment process accordingly. In the second place, *the what* of language assessment implies that language assessors should be expert in current theories and developments in language knowledge and use so that they can apply assessment measures in accordance with the current viewpoints. Moreover, knowledge about the current language teaching pedagogy is also stated to be as a component of LAL, and language teachers should be familiar with new theories of education to design appropriate assessment measures. In other words, to ensure the expertise that LAL requires, these theories and approaches should be transformed into competencies in assessment. As for *the how* of language assessment, language assessors are required to deliver appropriate assessment procedures and make interpretations of assessment. A sound in-depth knowledge foundation in various assessment orientations is required to ensure this. In short, Inbar-Lourie (2008) argues that language assessment including reason, purpose and method should be understood in view of the current theories, insights and findings in language-related areas, and language teachers should be familiar with these issues.

These requirements expected from language teachers stem from three major reasons (Fulcher, 2012). While the first reason is related to the language field, the other two are external to the field. The first and internal subject to the field is assessment for learning. Assessment for learning and its role in promoting student's learning has been highly focused in language programs. As for the second subject, which is external to the field, it is the increased use of assessment tools with the aim of accountability. The last subject stems from the quick development of the use of language tests within the scope of immigration policies and citizenship tests (Kunnan & McNamara, as cited in Fulcher, 2012, p. 2). Due

to the increasing effect of washback of tests on teachers, they are recommended to notice how testing policies and practices affect and shape their work. Taking all of these reasons and facts into account, language teachers are recommended to possess and develop their assessment literacy, which will contribute to improvement of the quality of education.

**The Importance of Language Assessment Literacy.** There is an expanding growth and significance of tests and assessment in language field, and accordingly language testing and assessment has taken the center stage recently. In regard to this expansion and importance of language tests and assessment, there is a need for language teachers who have sound assessment knowledge base and practices; in other words, LAL is called for on the part of language teachers. However, the research on LAL is still in its infancy (Fulcher, 2012); in parallel with this, the importance of LAL has been studied recently (Herrera & Macías, 2015).

Among these recent research, LAL has been highlighted as a crucial component of second/foreign language education on the ground that developing LAL facilitates language teachers to discover and make judgments about their own preconceptions, and to be familiar with knowledge, applications and values (Scarino, 2013). Through this procedure, language teachers improve their awareness as assessors, which is a fundamental part of their LAL. Besides, Falsgraf (2005) argues that LAL enables teachers to grasp, analyze and implement this information on student performance to improve instruction.

As language tests impact stakeholders, it is vital for language teachers to discriminate between the sound and unsound assessment (Stiggins, 1995). In the same vein, Siegel and Wissehr (2011) highlight the importance of it in that if language teachers are familiar with the types, advantages and disadvantages of types, tools, functions, principles, purposes and procedure of assessment, they will gather accurate information about students' progress and communicate this information to students. By doing so, they may adapt their instruction and set both immediate and long-term learning goals. To illustrate, if language teachers possess knowledge about the variety of types of assessment, they choose the most relevant and effective assessment tool to achieve learning goals. All in all, it has crucial benefits in learning and improvements in teaching as the growing body

of research suggests (DeLuca & Klinger, 2010; Wilson, Floden, & Ferrini-Mundy, 2001).

With a solid background and foundation in LTA, language teachers can integrate assessment and instruction, which is essential for effective English language teaching (Coombe et al., 2012). Malone (2013) also calls for the integration of these concepts as they inform and improve each other. In this sense, if language teachers do not have solid foundations of LAL, they are not very likely to establish connection between assessment practices and approaches of language teaching (Herrera & Macías, 2015). Unless they have adequate level of LAL, they cannot integrate assessment and instruction, which influences the effectiveness and quality of language teaching.

The fact that language teachers are engaged in a variety of assessment procedures in their professional life has also made language teachers' competence in this field crucial (Cheng, Rogers, & Hu, 2004; Davison & Leung, 2009; Rea-Dickins, 2004). Indeed, they are often required to organise and administer language testing and assessment activities, and make educational decisions that affect learning and teaching processes. In addition, in some situations they are required to construct and mark high-stakes tests on their own while in others they are expected to take a consultative role in terms of advising students on the choice of external proficiency tests (Tsagari & Vogt, 2014). Therefore, being familiar with these tests such as their content and specifications, strengths and weaknesses, and taking them into consideration while advising students on the choice of exams is necessary. Considering these various responsibilities of language teachers, LAL is really an indispensable part of their competencies.

In addition to these competencies, European Union policies on language learning together with new developments in foreign language teaching call for new competencies of language teachers. To demonstrate, the European Language Portfolio (ELP) emphasizes self-assessment and peer-assessment, which are added to the pedagogic agenda of modern language teacher (Morrow, 2004; Schneider & Lenz, 2001). These new developments require language teachers to acquire new skills and competencies (Edelenbos & Kubanek- German, as cited in Tsagari & Vogt, 2014, p.375). In other words, language teachers are also required

to adapt to these new developments in the field, and possess these new skills and competencies.

All in all, possessing and developing LAL is a critical skill for language teachers (Mertler & Campbell, 2005; Otero, 2006; Siegel & Wissehr, 2011) in terms of contributing to their own professional life and learners' development as well as the programs and institutions where they work (Popham, 2009). Therefore, the enhancement of LAL should be a concern for EFL teachers and accordingly teacher education programs (Siegel & Wissehr, 2011).

**Language Assessment Literacy in Teacher Education.** A growing body of research reveals that teacher training programs should incorporate assessment courses to their curriculum to provide language teachers with solid base of language assessment in both theory and practice; in other words, to develop assessment literacy of them (Brindley, 2001b; Herrera & Macías, 2015; Odo, 2016; Sato, Wei, & Darling- Hammond, 2008; Schafer & Lizzitz, 1987). In light of the fact that teacher candidates are to start teaching profession after completing their undergraduate programs, assessment education in preservice teacher education is pivotal in providing them with a solid foundation for future professional assessment practices (DeLuca & Klinger, 2010).

Recently, many researchers have pointed to future language teachers' preparation in LTA field (Brindley, 2001) and regarded it as a fundamental element of teacher training programs for a number of reasons. First of all, language testing includes a number of technologies and improvements which are not the same as language teaching though they interact with each other in many aspects (Johnson & Johnson, 2001), and these new developments should be introduced to language teachers so that they can keep up with the innovations. Besides, since one of the professional demands placed upon language teachers is to assess students accurately, assessment is becoming an increasingly fundamental field of language teachers' mastery (O' Loughlin, 2006) and teacher training programs should support the development of this mastery. The second reason is that instructors were not born to be tester (Jin, 2010); therefore, training in language assessment concepts, knowledge, skills, strategies and practice is required. In addition to the critical demand for assessment for effective teaching, accountability issue has compelled training programs to give more importance to training in assessment

compared to the past (DeLuca & Bellara, 2013). Furthermore, as teacher training has a central role in preparing proficient instructors (Odo, 2016), and teacher competency in the field of LTA is increasingly needed, incorporation of assessment education to the teacher education programs is necessary. When all of these reasons and facts are taken into consideration together, they call for the need for training programs to offer thorough training in assessment in their curriculum, and to develop LAL among EFL teachers.

Furthermore, as the professional demands on language teachers regarding assessment increase, assessment has been addressed as a fundamental domain of instructors' expertise and professional development in assessment is called for (O'Loughlin, 2006; Newfields, 2006; Brindley, 2001). At this point, Herrera and Macías (2015) specifically reiterate that EFL teacher training programs should provide teacher candidates with adequate preparation and professional development in LAL.

Similarly, Pollock (2011) argues that teacher educators have to ensure that prospective teachers possess classroom assessment competence regarding knowledge of assessment types, functions, purposes, students' ability level and how to support students' progress towards desired outcomes. Indeed, Siegel and Wissehr (2011) take it much further in that LTA should be completely addressed within language methodology courses so that prospective teachers can be more competent in the skills and knowledge.

With regard to the content of assessment training in teacher education programs, Popham makes suggestions in that this content can include many components such as functions of assessment, principles of assessment, elimination of bias in assessment, construction and scoring of test items as well as other forms of assessment (e.g., interviews, self-assessment, etc.) and creating formative assessment. Besides, preparing students for high-stakes tests, interpreting their performance and making decisions about them can also be included in assessment courses.

Furthermore, Herrera and Macías (2015) point out the critical need for training in assessment by drawing on the Stiggins' (1999) content requirements or competencies that aim to promote a sound base in assessment. As for these critical components that should be addressed in teacher education programs, they are setting clear purposes for assessment, making achievement expectations

clear, using appropriate assessment methods, setting quality assessment exercises, criteria for scoring and sampling accurately, avoiding bias in assessment, and informing about student success. Receiving training in these is critically helpful not only for EFL teachers in terms of having a clearer representation of learners' knowledge and abilities, but also for students in terms of being provided with more reliable and accurate assessment tools and practices (Herrera & Macías, 2015).

In addition to these perspectives on what assessment courses should address, Fan, Wang, and Wang (2011) and Popham (2009) offer a new insight in that the content of the assessment courses is required to keep pace with research and innovations in the field of assessment so that language teachers can follow and adapt to the recent changes and developments in their profession.

With regard to the LTA in Turkish education, foreign language teaching and learning has been an important issue since the Ottoman Empire (Kucukoglu, 2012) and several changes and reforms have been made in foreign language education system. Among these, 1997 Education Reform brought about fundamental changes in English teaching policy in that new English curriculum for fourth and fifth grades was developed, and following this, curriculum of education faculties was reshaped in terms of teacher training courses. This reform has been rearranged for further changes in foreign language teaching policy, and one of the major changes taking place in English Language Teaching is assessment. Similarly, curriculum reform of 2005 has brought about important changes in assessment regarding monitoring learners' progress through formative assessment and using authentic assessment instead of traditional assessment (Aksit, 2007). When these reforms and changes are taken into consideration, language teachers in Turkey are required to adapt to them and assess learners' language development accurately. In this regard, responsibilities placed upon teacher education programs to train language teachers in assessment have increased in accordance with changes and reforms taking place in assessment.

Training teachers in Turkey has been under the responsibility of the Faculties of Education (FE) since 1982, and English language teachers are trained in English Language Teaching Departments, which are placed in the Faculties of Education. In Turkey, sixty-four state universities have Faculties of Education, twenty-nine of which offer Bachelor's degrees in English Language Education in

the departments of Foreign Language Education (YÖK 2009, as cited in Hatipoglu, 2010). All English Language Teaching Departments use standardized curriculum, and within this curriculum, three different groups of courses are offered. These are field courses (e.g., Approaches to ELT, Teaching Language Skills, English Language Testing and Evaluation Course), education courses (e.g., Classroom Management), and knowledge courses (e.g., Written Communication in Turkish, Non-departmental electives).

With regard to the assessment courses, it has been found out that there is just one English Language Testing and Evaluation Course (ELTE) after the investigation of the curricula of twenty-nine Foreign Language Education Departments'. In this regard, all sound foundations of assessment including theoretical and practical components are expected to be addressed during this course, and preservice language teachers should be prepared for their future assessment practices through offered courses like ELTE, school experience and practice teaching courses. In other words, language teachers are supposed to be competent at ELTE so that they can assess students accurately and deliver effective teaching after graduating from these departments. Therefore, scrutinizing the extent to which they are trained in language testing and assessment during undergraduate education to possess and develop LAL is essential.

**Related Research Conducted in Turkey and Abroad**

This section presents studies related to assessment education, and particularly research studies on the assessment literacy of preservice teachers studying in different teacher education programs. After reviewing the results of those studies, the section also presents studies related to language assessment training and particularly research studies on the LAL of preservice and inservice language teachers.

**Abroad.** The research studies that focus on preservice teachers' assessment literacy have mostly investigated the effectiveness of assessment education and put forward mixed findings across quantitative and qualitative studies. In most of these studies, surveys were employed to identify preservice teachers' perceived confidence, self-efficacy or attitudes regarding assessment. To start with the study of Volante and Fazio (2007), the development of

assessment literacy of teacher candidates in a teacher education program was explored. An assessment literacy questionnaire was administered to 69 teacher candidates. The results of this questionnaire revealed that the levels of self-efficacy were very low and teacher candidates highlighted the necessity of assessment practices, constructing tools, scoring and administrating. Teacher candidates also argued for the need for specific courses related to assessment and evaluation. In this regard, Volante and Fazio indicated that teacher education programs offering such specific courses should not necessarily suppose that their teacher candidates graduate with an adequate level of assessment literacy; therefore, they proposed that it is necessary for these programs to carry out a systematic gap analysis of preservice teachers in order to determine the match and mismatch between applied assessment curriculum and what they really learnt.

In another study, DeLuca and Klinger (2010) examined the viewpoints of teacher candidates on assessment education program in Canada. In line with this purpose, a questionnaire was employed to report participants' perceived confidence levels in theory, philosophy and practice of assessment. Besides, they proposed their ideas on which assessment topics should be included in preservice assessment course. It was reported that teacher candidates reported much higher levels of confidence in theory and practice knowledge domains than in assessment philosophy domain. In other words, the assessment course was stated to have a greater influence on their confidence levels regarding assessment practice and theory, and to a lesser degree, assessment philosophy. All in all, the confidence levels stated by these participants were really high contrary to the results reported by Volante and Fazio (2007), and the assessment course was influential in teacher candidates' development of confidence in assessment. Moreover, the results of this questionnaire revealed the need for direct instruction in certain topics of assessment such as item reliability, validity, adjusting assessment and reporting achievement in that these topics are essential for developing assessment literacy of preservice teachers.

Similarly, Chen (2005) carried out a study to identify preservice teachers' confidence in assessment through a questionnaire. This was administered to 61 preservice teachers at undergraduate and graduate levels who took the measurement course. It was reported that both groups of teacher candidates were

more confident at the end of the measurement course; however, it was stated that measurement course benefited graduate level teacher candidates more.

In 2013, DeLuca and Bellara carried out a study to analyze the harmony between the accreditation policies of teacher training, professional standards for assessment practice of teachers and the curriculum of LTA course. As limited assessment education was regarded as a factor that leads to lack of assessment competency, the need for thorough preservice assessment education which closely relates to assessment literacy was highlighted in their study. It was also asserted that preservice assessment education programs should be familiar with the importance and scope of assessment literacy, which will support them in incorporating and addressing this crucial teacher competency all through their programs. In this regard, the enhancement of teacher preparation in the field of assessment was put forward.

In a different study, Siegel and Wissehr (2011) examined secondary preservice teachers' understanding of assessment tools and reasons for using assessment as well as the way they incorporate assessment into inquiry-based science units. Based on the analysis of their teaching philosophies, reflective journals and inquiry science units that they developed during the course, it was found that teacher candidates reported a need for adjusting assessment in accordance with learning goals and instruction, and they used various assessment tools along with a strong understanding of purposes of assessment. However, assessment in these units were found not to be fully aligned with their views on assessment in that they utilized traditional assessment methods in their units despite their reflections on the use of various assessment methods. All in all, these researchers argue for the need for teacher training programs to emphasize preservice teachers' development of assessment literacy in that with an adequate level of assessment literacy they are more capable of selecting and administering various appropriate assessment tools to facilitate students' learning.

More recently, Odo (2016) has investigated preservice teachers' development of assessment literacy by employing individualized tutoring and peer debriefing. The aim was to identify preservice teachers' perceptions and process of development of assessment literacy over the literacy assessment course including tutoring and peer debriefing components. In the light of this purpose, an exploratory qualitative study was carried out and five preservice teachers taking

the "individualized assessment for reading instruction" course were chosen. Data was collected through two main sources: three reflection essays in which preservice teachers reflect upon theories addressed in the course and the implementation of these in their tutoring, and individual semi-structured interviews in which they reflect on many aspects like what they learnt during the course and how they used the theoretical knowledge during tutoring. These were analyzed through constant comparative and content analysis methods. It was found out that both individualized tutoring and peer debriefing contributed to the development of assessment literacy in that the former enabled them to put theory into practice and the latter allowed them to share their knowledge and experience with each other and reflect upon ways of improving their tutoring. It was suggested that the crucial role of assessment literacy in today's classrooms should be recognized by teacher educators, and accordingly it should be given priority in teacher education programs. It was also suggested that teachers should be familiar with informal classroom assessment tools in that using informal assessment tools like reading inventories or anecdotal records provides more comprehensive data about students' learning compared to traditional testing.

On the other hand, assessment literacy of preservice and inservice instructors was compared in some studies. In 2003, Mertler conducted a study to make a comparison in the assessment literacy of preservice and inservice teachers. For this purpose, classroom assessment literacy inventory was used and the two groups were compared statistically. It was found that inservice teachers scored higher than preservice teachers on the classroom assessment literacy inventory. Mertler stated that inservice teachers usually feel that they did not get adequate training in teacher training programs, and they are not so confident about their decision making in assessment. Besides, he stated that teachers learn more about assessment in practical experience than in teacher training programs. In this regard, the findings pointed to the need for special attention on assessment in preservice programs. In contrast, the study by Alkharusi, Kazem, and Al Musawai (2011) indicated that preservice teachers scored higher than inservice teachers on assessment knowledge on various measures. In this regard, preservice teachers completing a measurement course were stated to be more assessment literate, and this may be connected with completing measurement course and teaching practicum.

Mounting evidence on assessment literacy of preservice teachers demonstrates that assessment education in preservice teacher education programs foster preservice teachers' grasp of assessment and literacy development (Mertler & Campbell, 2005; Popham, 2009; Siegel & Wissehr, 2011); however, some studies reveal that assessment education in preservice teacher education programs is not adequate and assessment courses continue to be few numerically (DeLuca & Klinger, 2010; MacLellan, 2004). Particularly, the study by MacLellan (2004) examined teacher candidates' assessment literacy in terms of analyzing their personal statements on assessment knowledge qualitatively. Based on this, it was reported that participants expressed a variety of purposes and ways of assessment; however, their knowledge on other fundamental knowledge regarding assessment such as reliability, validity and fairness was found to be very limited. She also asserted that having a low level of assessment literacy may have negative effects on students' learning. Besides, she argues that currently preservice programs continue to disregard the role of assessment in teachers' future professional development. Another study by Kahl, Hofman, and Bryant (2013) also revealed that the preparation of teachers has been "incomplete and superficial" (p. 3) from a historical perspective.

Likewise, previous studies have revealed that instructors regard themselves as unprepared for assessing learners appropriately (Mertler, 1998, 1999; Plake, 1993). Teachers were found to believe that they did not receive sufficient training either in undergraduate or graduate programs; therefore, they did not regard themselves as having adequate skills in administering assessments or making assessment-related decisions. The finding is consistent with the findings of the study by Popham (2004). Popham described two obstacles to teachers' development of assessment literacy, which are not receiving training in assessment in preservice teacher training and the misunderstanding that assessment literacy entails a much focus on knowledge on technical measurement and statistic. He also argued that adequate level of assessment literacy is required for teachers to engage proficiently in choosing appropriate assessments that align with instructional objectives and learning goals. All in all, lack of or inadequate training in assessment during teacher education programs was reported as a barrier to teachers' development of assessment literacy in their future professional experience.

On the other hand, there are some studies on LAL of preservice and inservice language teachers, and these studies focus on different aspects like training received in LTA, need for further training, perceptions of language teachers on assessment, language teachers' level of assessment literacy, LTA courses, and practice of assessment in practicum.

Hasselgreen, Carlsen, and Helness (2004) conducted a study to investigate teachers' LTA literacy and particularly their training needs. They employed a questionnaire to identify the training needs of language teachers, language teacher trainers, and experts that design tests. 914 participants took part in this study and among them 361 were language teachers. In the first part of the questionnaire, some background questions such as in which country they work, which languages they deal with and the degree of formal education or training in LTA were asked to elicit information about the participants. In the second part of the questionnaire, questions were classified into three groups which are classroom-based activities, aims of assessment and contents and concepts of assessment, and each question included three aspects: whether they were involved in the type of given LTA, the degree to which they received training in given areas, and the degree to which they perceived a need for training in these areas. Though the teachers' and teacher trainers' questions had similar themes, experts' questions included different themes like item writing and developing tests. The findings revealed that for *classroom-focused assessment,* teachers did not receive any training in using ready tests and informal/continuous assessment as well as providing feedback. As for the *purposes of assessment*, teachers revealed that they did not receive any training in awarding certificates and placing students, and they needed training in all areas of purposes of assessment. Regarding *contents and concepts of LTA*, teachers were reported to use statistics, establish reliability and assess aspects of culture and integrated skills without any training, and needed training in all areas of contents and concepts of LTA. Finally, some activities like reviewing, writing tests and using statistics were performed without any training, and the need for training in these areas was reported. All in all, the language teachers taking part in this study revealed that they carried out most of the activities regarding LTA without any formal training, and they stated that they need training in these given aspects of the LTA. On the other hand, the researchers paid attention to some issues while interpreting the results. First of all,

they stated the difficulty of separating teachers from the informants since the teachers were not just practicing teachers and may be teacher trainers or experts. Secondly, they stated that these informants may have some background in LTA; therefore, they may skew the target group. In this sense, they argued that investigating regular foreign language teachers may give a more accurate picture of the LAL of practicing teachers.

More recently, building on the study of Hasselgreen et al. (2004), Vogt and Tsagari (2014) have carried out a study to identify the current level of language testing and assessment (LTA) literacy of foreign language teachers and their training needs in this field. The Hasselgreen et al.'s (2004) questionnaire was adapted and designed for the purposes of this study. The participants of this questionnaire were 853 foreign language teachers from different countries. This questionnaire was followed by interviews with 63 instructors. The purpose of this interview was to examine individual teachers' training biographies, their individual needs regarding LTA in inservice teacher training as well as identifying the strategies that they use to make compensations for the lack of expertise in this field. The results of the questionnaire revealed that among the three components of LTA, *purposes of testing* was reported to be the least developed area in that 42.4 % on average stated that they did not receive any training. In this regard, it was stated that *purposes of testing* like giving grades, placing students, and awarding certificates were underdeveloped, and they were mostly learned on the teaching profession according to the the results of the interview. As for the *classroom-focused LTA* practices, they were also stated to be underdeveloped given that most received no or little training, particularly in the alternative assessment like ELP and portfolio. As to the *content and concepts of LTA*, foreign language teachers expressed that they received either no (29.4%) or a little (35.1%) training in this component. Besides, almost half of them (41.4 %) stated the need for advanced training in this component. Based on these findings, they stated that foreign language teachers' assessment literacy was underdeveloped. Also, they argued that though foreign language teachers received some training in these components of LTA, most of them expressed their need for further training in these areas. Although the finding was consistent with the findings of Hasselgreen et al. (2004)' study in that the need for training in almost all areas was indicated by foreign language teachers, some aspects of LTA were regarded as more urgent

than other aspects compared to the study of Hasselgreen et al. Besides, further training in some aspects like four skills, grammar and vocabulary and innovations in LTA like ELP and self/peer assessment were reported in this study in contrast to the study of Hasselgreen et al.

As for the findings of the interview, they revealed that teachers were not prepared adequately for LTA in their preservice teacher education programs, and accordingly most of them did not feel adequately prepared for their LTA practices after graduating from preservice teacher training programs. On the other hand, most were reported to seem to be confused about assessment methods and the ways innovative forms of LTA were used. Also, the majority of them stated that they received limited amount of training in their inservice teacher training programs, which resulted in use of compensation strategies during their teaching profession. Vogt and Tsagari also expressed that these strategies may be helpful to some extent; however, they may prevent more innovative and up-to-date forms of LTA from being part of their assessment practice. In this regard, they suggested that inservice teacher training programs should incorporate these innovative forms and practice in LTA into their curriculum; in other words, teachers should be prepared for these new developments in the field and provided with hands-on training opportunities in LTA.

In a different study, Jannati (2015) examined the viewpoints and practices of English instructors on assessment through semi-structured interviews. Analyzing these interviews through content analysis, it was found that teachers were familiar with the essential concepts of assessment, and different teaching experience did not make an important difference in how they perceive assessment. Furthermore, it was reported that though they were assessment literate and aware of basic concepts and assessment principles, this was not revealed in their assessment practices. To demonstrate, though they were aware of the features of good assessment like validity, reliability and authenticity, they did not transfer such knowledge into real practice.

In 2011, Razavipur, Riazi, and Rashidi (2011) conducted a research to separate the effects of LAL of instructors from the washback of summative tests in Iran. 53 EFL teachers participated in this study and were administered a test of LAL and a survey. The assessment literacy test included 35 multiple choice questions and followed by a number of self-assessment questions regarding the

degree to which they regard themselves as proficient in English, prepared for language teaching and competent at LTA. As for the survey, it was a likert-scale survey that addressed the most typical and possible language teaching practices in the given context, and it consisted of two categories. While the first category included items which were regarded as helpful in making students pass the summative tests directly, the second category included items that were not mainly designed to help students to increase their scores on the final exam. The findings revealed that teachers had a poor knowledge base in assessment; in other words, their level of knowledge in assessment is below the adequate level, and regardless of their level of assessment literacy, they adapt English teaching and testing to the requirements of external tests. On the other hand, EFL teachers with more assessment literacy were regarded to be more likely to incorporate non-washback practices into their instruction. Given the low level of LAL, they argued for thorough courses in language assessment both in preservice and inservice trainings. They also stated that overuse of traditional ways of assessment and lack of practice in communicative assessment may lead to further decline in their assessment competence. Besides, they argued that positive washback and their knowledge base in language assessment may be promoted through the use of communicative language assessment. Furthermore, they suggested that teachers' access to local, national and international testing and assessment journals should be made easier to facilitate the improvement of their AL. All in all, based on the findings, they made suggestions related to teachers' training and professional development in language assessment. Following this research, Razavipour (2014) carried out another study to examine what LAL is for three different groups of stakeholders: LAL test developers, LAL instructors and LAL test-takers. With the aim of this study, a questionnaire was designed to examine the perceptions of both language instructors and test-takers while the recent version of LAL test was used to examine the perceptions of language test developers. Based on the results of these, it was revealed that LAL is mostly a matter of knowledge and theory of language assessment with little focus on skills and principles for the test developers. As for the language testing instructors and test-takers, LAL is mainly a matter of skills. Furthermore, test-takers regarded LAL as the most challenging unit of the test in terms of theories, statistics and their lack of practical experience with language tests. All in all, this study revealed that there was a difference

between what LAL means for test developers and what it means for language testing instructors and test-takers.

On the other hand, Lam (2015) explored the LTA training in Hong Kong teacher training programs and to investigate the perspectives of preservice language teachers and instructors on the degree to which two assessment courses promote or hinder preservice teachers' development of LAL in one teacher education program. With the aim of exploring overall training in language assessment, a survey was administered to these five teacher training programs. As for the investigation of the perspectives of preservice teachers and instructors on to what extent two assessment courses promotes LAL, 9 instructors and 40 preservice teachers in their final year took part in the study and focus group interviews were conducted. Based on the findings, it was reported that LTA training in Hong Kong was not sufficient and the two courses could not bridge the gap between theory and practice in language assessment. Particularly, instructors and preservice teachers revealed that social aspects of LAL like ethics and fairness were not incorporated into most of the assessment courses, and the assessment training was too academic, which resulted in a wider gap between theory and practice during practicum experience.

With regard to this gap, Lam suggested that preservice teachers' assessment knowledge and skills may be included and evaluated in teaching practicum so as to foster the development of LAL in the long term. As the findings demonstrated, when LAL is not evaluated during practicum, it is highly possible that teacher candidates will neglect its influence on teaching and learning processes. In this respect, the incorporation of LAL into the practicum may foster their awareness and competence in language assessment. Lam also revealed that these five teacher education programs may not provide teacher candidates with sufficient language assessment training. Within this context, it was noted that these teacher candidates may not be provided with up to date training in language assessment to develop their LAL. From a different perspective, he suggested that teaching of assessment knowledge and skills should be contextualized to facilitate teacher candidates' LAL. To illustrate, certain assessment concepts like *assessment for learning* may not work well in a test-based learning context. All in all, he asserted that incorporating LAL into language assessment courses is compulsory and indispensable in terms of providing teacher candidates with solid

foundations of knowledge, skills and principles of assessment though it is not the only way to promote LAL.

As for the research on LTA courses, limited research has been conducted on learners' evaluation of these courses. In their first study, Bailey and Brown (1996) investigated the characteristics of these courses regarding instructors, course features and students. The questionnaire employed included various aspects like topics covered, textbooks used, practical experience students got during the course, test validity and students' perceived attitudes towards the course. Although the items in the questionnaire were mostly in Likert-scale format, a number of open-ended questions were also included. As for the second study of Brown and Bailey (2008), it also aimed to explore the features of LTA courses and in which ways these courses have changed since 1996. The same questionnaire was employed with some components added, and administered to 97 participants. The results of these two studies revealed that topics like testing various skills, theoretical bases in reliability and strategies for figuring out test reliability, and the general strategies for revealing validity were widely addressed in most of these courses. Besides, most participants expressed that their students regarded LTA courses as useful and balanced in terms of theory and practice. As for the comparison of the results of these two studies conducted in different years, the researchers concluded that although topics in these courses were overlapping in the 1996 and 2008 studies, new topics like consequential validity and computer based TOEFL were also incorporated into language testing curriculum.

In another study, Jin (2010) investigated the training of tertiary level foreign language teachers regarding LTA courses in China. A survey was designed to find out the content and methodology of these courses and the students' thoughts on these courses with regard to the necessity and usefulness. The study demonstrated that various topics covered in these courses in tertiary level in China were parallel to those reported by Brown and Bailey's studies, and Jin concluded that these courses sufficiently addressed the basic dimensions of theory and practice of LTA. On the other hand, educational and psychological measurement as well as student classroom practice did not receive sufficient attention. Besides, students reported that these courses were necessary and they were motivated to take them mostly because of their theoretical and practical usefulness.

Furthermore, Kleinsasser's (2005) study described how LTA course was transformed from a content-centered to learner-centered and teaching content-focused course with the collaboration of the teacher and the students. The course included many topics like hands-on item writing, giving feedback, construct validity and critical perspectives to language testing. The students and the instructor constructed various assessments during the course and made discussions on testing and assessment, which facilitated the formation of a professional practice community. In another study, O'Loughlin (2006) examined an elective course which was offered in the postgraduate program of the University of Melbourne. The aim of the researcher was to demonstrate how the students' grasp of key concepts and the ability to make judgments about assessment tools improve. Two students' written narratives for 12 weeks were analyzed, and both of the students were reported to have achieved the course objectives; however, they were reported to differ in their eagerness and ability to get new ideas in LTA. The researcher stated that these differences may stem from some factors like their cultural background, previous experience in assessment and the features of the input received in the language assessment courses. In this regard, he asserted that a learner-centered approach should be employed during the design and process of assessment courses. These two research studies provided insights into language assessment courses in that how these courses should be planned and conducted.

On the other hand, some research has been undertaken to study preservice teachers' assessment practices in practicum and the effects of assessment practice on their development of assessment literacy. To start with the study of Anderson, Mathys, and Mills (2014), their study focused on examining how preservice teachers' assess 7th grade students' learning of social studies lessons that they planned and delivered. 52 elementary preservice teachers taking part in this study were required to assess students' work and provide their rationale and explanation for their assessments. In this regard, they were required to provide two examples of students' works that were average, two examples that were above average and two examples that were below average. They analyzed each of them and reflected on their justification for why and how they classified each example accordingly. Among the 429 coded assessment explanations, 240 were connected with students' social studies achievement, 141 were connected with

non-achievement factors, and 48 were connected with achievement factors not related to social studies like effort and participation. The researchers stated that their findings were parallel to the findings of the previous study on assessment practices of inservice teachers in that both preservice and inservice teachers used many factors like students' achievement, effort, participation to evaluate students. They concluded that considering the increasing importance and role of assessment of student learning in education, new teachers should be provided with opportunities to develop their assessment practices. In this sense, teacher educators should combine explicit course on assessment with authentic field experiences for teacher candidates so that they could practice assessing and grading students in an authentic context.

Alkharusi, Kazem, and Al-Musawai (2011) conducted a research to uncover the differences in knowledge, perceived skills and attitudes toward educational measurement between pre and inservice teachers as well as investigating whether practicum and teaching experience made a difference in the given aspects of both groups. The results demonstrated the importance of preservice measurement training. Besides, preservice teachers who did practicum and took the measurement course had higher levels of knowledge and skills in educational measurement and more positive attitudes towards educational measurement compared to the preservice teachers not having teaching practicum while taking the course. In this regard, they argued that teaching practicum and experience were useful in preparing teachers in educational measurement in terms of improving preservice teachers' assessment knowledge, skills and attitude. Therefore, they concluded that teacher education programs should connect assessment courses with field-based experience to maximize the development of assessment literacy of preservice teachers.

On the other hand, Campbell and Evans (2010) investigated the preservice teachers' assessment practices who have taken educational measurement course. For this purpose, they reviewed 65 preservice teachers' lesson plans. It was found that lesson plans did not include all of the criteria established as necessary in the evaluation of students' learning. Especially, it was reported that preservice teachers did not incorporate essential concepts like validity and reliability into the assessment of students' achievement despite their recent training and practice in these concepts. They stated that although preservice teachers completed the

course successfully, their retention of measurement knowledge was not demonstrated in their practice. At this point, they suggested that preservice teachers may have regarded these abstract concepts like validity and reliability as dispensable or not applicable. On the other hand, preservice teachers' knowledge of item and test construction was adequately observed during their practices. All in all, the researchers found that preservice teachers did not follow most of the assessment practices addressed in their measurement course.

In addition, Davin and Heineke (2016) focused on how teacher candidates were prepared to assess students' language development in an accurate and meaningful way, and particularly how teacher candidates' learning of classroom-based language assessment was combined with practice in language assessment in their article. They asserted that using practice-based approach to prepare teachers for classroom-based language assessment enables them to learn authentic ways of using language assessment for facilitating student learning. In addition, they suggested that teacher candidates should grasp the dynamic nature of continuous and meaningful assessment, which meets students' needs and facilitates their learning. They concluded that teacher educators should situate assessment courses in practice like the field work in methodology courses.

Similarly, Lam (2015) asserted that some preservice teachers were doubtful about the application of assessment in their teaching practicum based on his study on language assessment training of teachers. He suggested that program directors, instructors and practicum supervisors may include teacher candidates' assessment knowledge and skills as part of pedagogical content knowledge (Gu, 2014; Inbar-Lourie, 2013b) to establish connection between theory and practice of assessment, and adapt the assessment rubrics of practicum in accordance with the development of LAL in the long term (p. 189). He stated that when LAL was not evaluated in formal means in the practicum, preservice teachers may ignore it. In this sense, he proposed that preservice teachers could be required to videotape their assessment practices during the practicum to assess their LAL. Besides, they may be required to have reflective journals that include their beliefs in assessment of students regarding the knowledge, skills, principles and practices covered in language assessment training.

**Turkey.** A detailed study of the literature revealed that there is a scarcity of research on assessment literacy, especially LAL, of preservice teachers and language assessment courses in Turkey.

One of the few studies conducted by Ogan-Bekiroglu and Suzuk (2014) focused on examining 28 preservice physics teachers' assessment literacy and the practice of it. The study was both quantitative and qualitative. In the quantitative part of the study, the researchers aimed to investigate preservice teachers' assessment literacy after they were registered to the course through an instrument developed by the first researcher and a questionnaire. In the qualitative part of the study, the researchers aimed to validate the results of quantitative research and to investigate their assessment literacy in practical terms. In other words, the data was collected through an instrument, a questionnaire and preservice teachers' project assignment including implementation of assessment. As for the results, quantitative research demonstrated that preservice physics teachers' assessment literacy was very high regarding knowledge of assessment, which was consistent with the results of the qualitative analysis. On the other hand, it also showed that there was a gap between theory and practice in assessment literacy. Though preservice teachers demonstrated grasp of key aspects of assessment literacy in theoretical bases, they had difficulties in putting these aspects into practice. To put it another way, they could not internalize what they learnt regarding the concepts of assessment; therefore, they were reported to have lower assessment literacy in terms of practice. Based on these, they concluded that teacher education programs should give preservice teachers opportunities to reflect on, practice and reconsider assessment methods as well as focusing on theories of assessment.

On the other hand, Cirit (2015) aimed to examine the perceptions of ELT preservice teachers toward different assessment methods: alternative, online and traditional. The study in which 40 ELT preservice teachers participated was conducted throughout the fall semester. The findings of the study revealed that preservice teachers were reported to believe in the importance of using alternative and traditional assessment together; however, they had a positive attitude toward the alternative assessment through Web 2.0 tools more than they did toward traditional and online assessment.

In his thesis, Yetkin (2015) examined the perceptions of English language candidate teachers toward the use of assessment strategies in primary/junior settings. Particularly, this study explored (a) the candidate English language teachers' perceived grasp of the purposes of classroom assessment; (b) the contributions of ELT program to the students' assessment literacy; (c) the candidate teachers' most favored assessment approaches; (d) the candidate English language teachers' perceived needs for further training in the classroom assessment approaches; (e) the candidate English language teachers' suggestions to improve assessment literacy (p. 4). This study in which 30 preservice English language teachers participated was conducted through a questionnaire. This questionnaire consisted of 3 close-ended questions that were asked to 30 teacher candidates and 5 open-ended questions that were conducted with 5 teacher candidates in an interview format. The finding regarding their perceptions of purposes of classroom assessment indicated that they mainly regarded classroom assessment as a tool for observing the development or level of students, observing their success as a teacher, understanding how to increase students' consciousness, changing teaching strategies and motivating students. As for the finding related to the contributions of ELT program to the students' assessment literacy, it was found that they believed the courses and the practicum in the ELT program enabled them to develop their understanding of the theories of classroom assessment. In addition, they found presentations, reviewing articles and preparing term projects as well as rubrics used by instructors as helpful in developing their assessment literacy. The results regarding their most favored assessment approaches indicated that observation technique was the most favored approach whereas selected response and constructed response were the least favored assessment approaches. In addition, the results related to their perceived needs for training in these approaches revealed that they perceived a need for training especially in three assessment approaches which were observation techniques, performance assessment and personal communication. Lastly, it was revealed that they found preservice language teacher education program very helpful for their development of theoretical bases of classroom assessment. Besides, they found practicum experience as useful in terms of contributing to the development of assessment practice.

On the other hand, Hatipoglu's study (2010) focused on the way language testing was taught in language teacher education program and the perceptions of students on the way it should be taught. Particularly, the study focused on an undergraduate ELTE course that was offered at Middle East Technical University, and aimed to students' evaluations of the course regarding its content and teaching methodology. In the light of these purposes, a questionnaire was conducted with 81 FLE students. The questionnaire included two parts. In the first part of the questionnaire, students were required to give information about themselves. In the second part of the questionnaire, participants were required to answer two questions: 1) List five topics that were addressed during the course and which you think would help you as an English language teacher in your future career. Then, state the reason and the ways that these topics would help you in your future profession.; 2) List three things that you think should be changed regarding this course to make it better if it is taught again, and state the reason of them (p. 45). Following the questionnaire, an interview was conducted with 16 volunteer students to triangulate the questionnaire data and obtain more detailed information on the research questions addressed in the study. As for the result of the first question, it revealed that three topics were stated more frequently than the others by students, and these topics were testing skills/knowledge, reliability and validity. These three topics were followed by kinds of tests, writing multiple choice item tests, teaching and testing, and kinds of testing. Regarding testing skills/knowledge, all students expressed that they would use these practices when they became English language teachers. As for the reasons why this topic was useful, reasons were classified into four groups. These reasons were the parallelism between the topics covered in the class and the things the instructor did in the classroom, familiarity with various testing techniques, appropriate tests for various groups of students, and procedures of scoring. Regarding reliability, the majority of the students stated that this was useful for them by providing some reasons like the role of reliability in tests and the importance of statistical analysis. As for the validity, most students also stated the usefulness of validity by providing some reasons like the impacts of tests on students' lives and the preparation of better and fair tests.

Regarding the second question, it revealed that eight different suggestions were made about the course, which were not having enough practice, student

presentations, need for more testing courses, topics addressed, crowded classrooms, assessment system used in the course, and course books and slides before lectures. These suggestions were analyzed and categorized into two groups: content and procedure of the course. Regarding the content of the course, students stated that too many topics were covered in the course and some of which were very abstract like kinds of testing, and they could not use them in their classrooms. In this regard, they suggested that abstract topics should not be covered and more practical topics like testing skills should be covered. Some students also suggested that the main texts used during the course should be changed on the ground that the language of the books was too complicated and heavy. In addition, others expressed that they needed presentations for better understanding of the topics. As for the procedure of the course, some students revealed that the theory and practice was not in balance in this course; therefore, they requested a better combination of these two components. Besides, they recommended changes for student presentations in that these were regarded as boring and not efficient, and they should be made individually instead of groups. Regarding these presentations, while some students stated that these presentations should be removed as they wanted to learn the topic from the instructor, others argued that these presentations should be provided before the discussion of topics. Another suggestion made by students was related to the crowded classrooms in that it did not allow productive classroom discussions and they could not follow the instructor well. On the other hand, they suggested that more than one ELTE course should be offered, and argued that these lessons should be offered beginning from the first year of their education considering the importance of it for their professional development. Lastly, informants suggested that they should be given alternative assessments like projects instead of two formal exams. Hatipoglu (2010) concluded that ELTE course may be adapted and realigned in accordance with these suggestions about the course like demonstrating the importance and relevance of topics covered during the course more, and including practical component of assessment.

In the following study, Hatipoglu (2015) aimed to uncover the needs of preservice English language teachers regarding assessment in Turkey. It also aimed to report the findings of the needs analysis survey that attempted to reveal the beliefs and perceived needs of preservice English language teachers at Middle

66

East Technical University (METU) regarding testing and evaluation. More specifically, it investigated (a) LTA training received by preservice teachers graduating from METU; and (b) the preservice English language teachers' expectations from ELTE course related to its content and methodology. In the light of these purposes, needs analysis survey questionnaires and interviews were carried out with 124 preservice English language teachers. The data for the study was collected between 2009 and 2012, and four groups of students were included. The needs analysis questionnaire was given to students each year at the beginning of the course, and it included three parts. The first part included questions regarding their age, gender and previous education to elicit information about them. As for the aim of the second part, it aimed to elicit information on students' previous experience of training in assessment and to obtain their opinions of the course. The third part of the questionnaire required them to list five topics that should be covered in the course as well as indicating how these would be useful in their future profession. On the other hand, semi structured focus group interviews were conducted with some selected preservice teachers with the aim of obtaining detailed information about their ideas on the design, content and methodology of the course.

The findings demonstrated that most students took just one ELTE course. As for their opinions on the usefulness of taking ELTE course in their development as teachers, all students, except for two, stated that it was useful in their development as language teachers. In this regard, the researcher stated that students were aware of the crucial role of testing in their jobs and the expectations in relation to assessment on the part of themselves as teachers. In spite of these, nearly half of the students stated that only one ELTE course was enough at the beginning of the semester. It was reported that the analysis of the justifications of students who regarded one course as enough indicated that some factors like not being familiar with testing and evaluation field and not having experience in constructing tests might have affected their answers. As for the findings of the last part of the questionnaire, it was indicated that most participants did not suggest any topics that they thought should be incorporated into the course, and only 27.5 % of them stated 4 or 5 topics while the rest of them just listed 1-3 topics. After the analysis of these topics, they were classified into nine different groups, which were general like testing students, testees like the features of target students, specific

67

topics like the way of testing idioms, testing skills like reading, speaking and listening, types of testing techniques like constructing multiple choice items, theoretical knowledge like theories of language testing, test administration like time of testing, test score evaluation like statistics, and alternative ways of testing like portfolios. Based on these, the researcher stated that lecturers should work with students in order to increase the usefulness of ELTE courses, and concluded that training in language testing and evaluation should be given more importance in teacher education programs, which contributes to the development of educational system.

On the other hand, Oz and Atay (2017) investigated Turkish EFL teachers' perceptions towards in-class language assessment and reflection of their beliefs on their actual classroom practice as well as exploring the relationship between perception of in-class language assessment and experience. 12 prep school teachers took part in this study, and among them, 7 were novice and 5 were expert teachers. A semi-structured interview was employed to collect data, and the interviews were recorded and transcribed based on common themes. Based on the findings, it was reported that most of the participants were familiar with basic classroom assessment; however, there was an imbalance between assessment literacy and classroom reflection with regard to actual classroom practice. Besides, it was revealed that experience and assessment perception were not much related. With regard to these, they suggested that institutions should provide some professional development opportunities to language teachers in order to develop their assessment practice and increase their awareness. Besides, they suggested that preservice teacher education programs should give more space to assessment with the aim of enhancing the quality of teacher assessment and actual classroom practice.

# Chapter 3

## Methodology

### Research Design

For almost a century, a majority of researchers have focused on either quantitative or qualitative research methods. However, each of these research methods has both strengths and weaknesses within their nature. As Dörnyei (2007) highlights, quantitative research includes accurate calculation and generates credible and replicable data which can be generalized to other situations as well as being systematic and strongly monitored. However, it cannot give the participants' internal perspectives thoroughly. On the contrary, qualitative research provides in-depth and rich descriptions of the participants' perspectives (Johnson & Christensen, 2004) contrary to the fixed numerical data, and may highlight new hypotheses or assumptions (Mackey & Gass, 2005) as well as focusing on an intensive study with a small number of participants. However, as Dörnyei (2007) puts forward, its findings may not be generalized easily because of the small sample size. Consequently, both quantitative and qualitative research methods have some weak points, which resulted in the emergence of a new research paradigm called the mixed methods research. As Tashakkori and Teddlie (2007) state, mixed methods research allows the researcher to gather and evaluate data as well as enabling them to combine the findings and to make inferences through employing quantitative and qualitative approaches.

Mixed methods research has both advantages and limitations. To start with the advantages that it provides, researchers can assure "complementary strengths and non-overlapping weaknesses" (Tashakkori & Teddlie, 2003) by utilizing mixed methods research. As each research method has inherent strengths and weaknesses, the integration of research methods can enable the researchers to use the strong sides of a research method to cope with the weak sides of another research method (Johnson & Onwuegbuzie, 2004, p. 21). That is, the mixed methods research gives the chance to make use of the strengths of each research method and to compensate for inherent weaknesses that each method has. Accordingly, both qualitative (interview) and quantitative (questionnaire) research methods were employed in this study to use the strengths of both research

69

methods and to avoid the weaknesses and limitations of mono-method research studies. Also, by integrating research methods researchers can answer a wider range of research questions as they are not limited to a single research method. In addition, this integration enables researchers to employ a variety of data collection instruments and analysis techniques, which provides a better and broader understanding of research questions, and more valid and reliable findings (Johnson & Onwuegbuzie, 2004, p. 21). In this regard, this study employed mixed methods research to obtain an in-depth understanding of research questions and a more complete picture of the topic investigated and to get more reliable and valid results as well as an in-depth understanding of the research findings. With regard to these advantages, mixed methods research has gained popularity and importance within the educational community in recent years (Gall, Gall, & Borg, 2007). On the other hand, there are some limitations of it such as the necessity for researchers to be knowledgeable in both research methods and to be acquainted with techniques of data analysis of both research methods as well as taking much time.

There are two major dimensions that shape the design of the mixed methods research, which are "time" and "paradigm" (Johnson & Christensen, 2004, p. 418). To start with the "time" dimension, it is classified into two categories: "concurrent" and "sequential". When the quantitative and qualitative phases of the research study are conducted at approximately the same time, the research is concurrent. On the other hand, when the quantitative and qualitative phases of the research study are conducted one after the other in that one phase informs the other, the research is sequential. As for the "paradigm" dimension, it is classified into two groups as "equal status" and "dominant status". When the quantitative and qualitative phases have approximately equal emphasis for the study, the research has equal status. On the contrary, when one phase has higher priority than the other with regard to responding to the research questions, the research has dominant status. The mixed method design is shown in the table below:

Table 1

*Matrix of Mixed Methods Design*

|  |  | Time Order Decision | |
|---|---|---|---|
|  |  | Concurrent | Sequential |
| Paradigm Emphasis Decision | Equal Status | QUAL + QUAN | QUAL →QUAN QUAN→QUAL |
|  | Dominant Status | QUAL + quan | QUAL→quan qual→QUAN |
|  |  | QUAN + qual | QUAN→qual quan→QUAL |

- The letters qual or QUAL stand for qualitative research.
- The letters quan or QUAN stand for quantitative research.
- Capital letters denote priority or increased weight.
- Lowercase letters denote lower priority or weight.
- A plus sign (+) represents a concurrent collection of data.
- An arrow (→) represents a sequential collection of data (Johnson & Christensen, 2004, p.418).

With regard to these, this study is based on an equal status (QUAL + QUAN) sequential design (QUAN→QUAL) in that both quantitative (questionnaire) and qualitative (interview) research methods have been given equal emphasis in the study and quantitative research has been conducted before the qualitative research method.

**Quantitative Research Design.** Quantitative research design which is based on numerical values and statistics has four main types: experimental research, causal-comparative research, correlation research, and survey research. Among the four, the survey research which is a highly generalizable research method was adopted in this study. As Dörnyei and Csizér (2012) define, survey research is one of the quantitative research methods that seeks to gather self-report data from

people. They also state that questionnaires are typically used to achieve this aim. They also suggest that the features, ideas, manners, and conscious actions of a large group of people could be described and analyzed through survey research. With regard to the nature of survey research, Creswell (2014) highlights its numeric and highly generalizable nature in that the results can be generalized to the whole target population with a representative sample of participants.

Survey research is conducted using two different time frames, which are cross-sectional and longitudinal. While cross-sectional research is conducted to gather data from the previously chosen sample at one time, longitudinal research is conducted to gather the data at different times with the aim of identifying the variations as the time progresses. The present study adopted a cross- sectional design since it did not aim to identify the changes in the variables over time.

A questionnaire was adopted in the present study as it gives the opportunity to reach a great number of participants at a time, which increases the reliability and generalizability of the research study. Dörnyei (2002) asserts that questionnaires can be designed easily and a great deal of data can be collected in short order. Another advantage of questionnaire is that using standardized questions saves time in the administration and scoring procedures, and hinders misinterpretations by the researchers to much extent. In addition, not requiring participants to reveal their names prevents respondent bias to some degree since they may feel more comfortable in expressing their views and answering the questions. In other words, the fact that questionnaires allow for anonymity also increases the reliability of the research. Accordingly, a questionnaire was conducted in the present study for the purpose of investigating preservice language teachers' LAL in terms of the training that they have received in language assessment and the need for further training in this field.

**Qualitative Research Design.** Qualitative research design which is based on non-numerical data is adopted when a detailed exploration and in-depth understanding of an issue is required to study a specific group of people or population (Creswell, 2007). It has many data collection methods such as ethnographies, diaries/ journals, case studies, observational techniques and interviews (Mackey & Gass, 2005). Among these data collection methods, interview was adopted in the present study as it gives an in-depth understanding

and descriptions of a specific group of people's experience and perspectives in a particular domain (Johnson & Christensen, 2004; Seidman, 2006). Also, it gives researchers opportunity to obtain additional data if the previously given answers are incomplete, unclear or not specific enough as it is interactive. Another advantage of interview is that it can be utilized to obtain data from participants who are not comfortable in other ways. In other words, some participants prefer speaking rather than writing, and are more likely to give extended answers in a conversational setting. In addition, interviews can be carried out in participants' native language, which enables them to express themselves better during the interview and eliminates concerns about the proficiency of the participant affecting the quality and quantity of the answers (Mackey & Gass, 2005).

With regard to the types of interviews, there are three main types: structured, semi-structured and unstructured. For the purpose of this study, semi-structured interviews in which questions are planned before the interview and the participants are given opportunity to express their further comments on the questions and elaborate on them were employed in this study. Moreover, the participants can be asked to share even some anecdotes with the researcher.

Besides, as Bryman (2008) asserts, using a structured format may hinder the depth and richness of the answers to be elicited. Therefore, using semi-structured interviews allows the researcher to elicit rich and depth responses from the participants. Accordingly, semi-structured interviews were conducted in the present study for the purpose of examining preservice language teachers' opinions on both whether and how the theory and practice of assessment was covered and whether and how they should be covered in school experience course, and the usefulness of ELTE course on their school experience course or practicum

**Setting and Participants**

The present study was conducted in the ELT Department of Middle East Technical University, one of the leading state universities in Turkey. The aim of this department is to make teacher candidates fully qualified teachers of English in educational institutions. In this faculty, only 4th grade students who have taken English Language Testing and Evaluation (ELTE) Course and School Experience Course participated in the study. The ELTE course offered in the 7th semester was

a compulsory course, and it included 42 hours of face-to-face instruction. As for the school experience course, it was offered in the 7th semester and it included 14 hours of face-to-face instruction and 40 hours of school visit (4 hours per week over 10 weeks).

The participants of this study were chosen based on convenience sampling, which is defined as choosing people who are suitable for the research (Mackey & Gass, 2005, p. 122). This sampling method is quite effective in terms of time, money and effort, and it is among the mostly used sampling methods in educational research (Muijs, 2004).

This study was carried out with 4th grade students in the four sections of ELTE course that was offered in the fall term of 2016-2017 academic year. There were 122 4th grade students in the university at the time this study was conducted. Before giving the questionnaires, the researcher mentioned the aim of this study and the requirements. In total, 101 students participated in the study.

With regard to gender distribution of these participants, 80 (79.2 %) of them were female and 21 (20.8 %) were male. Their ages ranged from 22 to 24. Most of them were graduates of Anatolian Teacher Training High School (79.2 %) while 15.8 % were graduates of Anatolian High School and 5% were graduates of Private High School. With respect to the home country of them, almost all of them (96.03 %) were from Turkey. There were 4 students who were from Bulgaria and Kyrgyzstan.

As for the grade point averages of these participants, the researcher classified these into three different groups: participants who got between 3.50 and 4.00; participants who got between 3.00 and 3.49; participants who got between 2.99 and below. In this regard, 37.6 % of the participants got between 3.00 and 3.49 and 35.6 % of them got between 3.50 and 4.00, whereas 26.7 % of them got between 2.99 and below. The table below shows the profile of the participants according to different variables.

Table 2

*Profile of the Participants (N = 101)*

|  | N | % |
|---|---|---|
| Gender |  |  |
| Female | 80 | 79.2 |
| Male | 21 | 20.8 |
| Age |  |  |
| 22 | 38 | 37.6 |
| 23 | 58 | 57.4 |
| 24 | 5 | 5 |
| High School |  |  |
| Anatolian Teacher Training High School | 80 | 79.2 |
| Anatolian High School | 16 | 15.8 |
| Private High School | 5 | 5 |
| Home Country |  |  |
| Turkey | 97 | 96.03 |
| Others | 4 | 3.97 |
| GPA |  |  |
| 3.50- 4.00 | 36 | 35.6 |
| 3.00- 3.49 | 38 | 37.6 |
| 2.99- below | 27 | 26.7 |

The researcher also collected data about whether the participants have taken any Testing and Evaluation or ELTE course before since it was taken into consideration that this may be influential in eliciting information about their LAL. None of them has ever taken Testing and Evaluation or English Language Testing and Evaluation (ELTE) course before. Accordingly, they have taken ELTE course for the first time. Besides, the researcher collected data about the main coursebook that was used during their ELTE course. Accordingly, two major textbooks were used in this course: Brown & Abeywickrama (2010) and Hughes (2003).

As for the qualitative study, a total of 25 participants who also filled the questionnaire were involved in the study as to specify their opinions on both whether and how the theory and practice of assessment was covered and whether they should be covered in school experience course, and the usefulness of ELTE course on their school experience course or practicum in depth. As conducting

interviews, transcribing and analyzing the recorded data are important concerns regarding time and energy, a limited number of participants (25 preservice language teachers) were selected for the interviews. There were no criteria like gender, age, grade point average, high school and so forth to choose the participants for the study; that is, the volunteer students were chosen for these interviews. On the other hand, at the beginning of the interviews, the researcher collected data about the type of school and level of education where they did practicum since it was considered that these aspects might be influential in the interpretation of the information about the theory and practice of assessment in school experience course. Before conducting the interviews, the researcher also mentioned the aim of the study and stated that their responses would be kept confidential.

In this part of the study, 20 female (80 %) and 5 male (20 %) students were interviewed. With regard to the type of school they did practicum, most of them (76 %) did their practicum in state schools while 24 % of them did their practicum in private schools. As for the level of education where they did practicum, most of them did their practicum either in secondary level of education (44 %) or in primary level of education (36 %). 20 % of them did their practicum in high schools. The table below shows the profiles of the interviewees according to different variables.

Table 3

*Profile of the Participants (N = 25)*

|  | N | % |
|---|---|---|
| Gender |  |  |
| Female | 20 | 80 |
| Male | 5 | 20 |
| The type of school they did practicum |  |  |
| State School | 19 | 76 |
| Private School | 6 | 24 |
| The level of education they did practicum |  |  |
| Primary School | 9 | 36 |
| Secondary School | 11 | 44 |
| High School | 5 | 20 |

**Data Collection Instruments**

**Questionnaire.** The questionnaire employed in this study is comprised of two main parts. The first part of the questionnaire is designed to collect demographic information about the background of participants, and includes questions related to participants' age, gender, high schools they graduated from, grade point averages, home country, and the type of school and level of education where they did practicum. In addition, questions such as whether the participants have taken any Testing and Evaluation or ELTE course before, and the main course book that was employed during their ELTE course were incorporated into the questionnaire by the present researcher since it was considered that these aspects might be influential in eliciting information about LAL of participants.

The second part of the questionnaire was adapted from Vogt and Tsagari's (2014) Teachers' Questionnaire which was developed based on Hasselgreen, Carlsen and Helness' (2004) Questionnaire. The "Teachers' Questionnaire" which was administered to foreign language teachers was adopted and adapted by the researcher because the items of the instrument are parallel to the information this study is seeking, which is to gain an understanding of the current level of preservice language teachers in LAL and to identify their training needs in this area (see Appendix A). In this regard, it fits the purposes of the present study. Another reason for adopting this instrument is that its reliability and construct validity has been proven and guaranteed with high Alpha coefficient (ranging from .80 to .93 for individual scales) at the end of factor analysis process. On the other hand, some necessary adaptations and modifications have been made in accordance with the purposes of the present study by deleting some items and adding new items and section to the questionnaire as well as making changes in the Likert scale. To put it another way, as some items would be too general and not be much relevant to the study's concerns, they were omitted from the questionnaire and made more specific. Some new items were also added to the three sections of the instrument as well as adding a new section to the questionnaire in parallel with the purpose of the study. These new items and the new section were added to the instrument in order to gain a deeper and detailed understanding of the current level of preservice language teachers in LAL and their training needs in this field. That is, the instrument had three sections;

however, a new section titled *"Knowledge of Testing and Assessment"* was added based on the book of Brown & Abeywickrama (2010) on the ground that the items of this section would be necessary and indispensable components of preservice language teachers' LAL. Besides, some changes were made in the Likert scale in that the original instrument required teachers to give answers in one of the three ways and the adapted version required participants to give answers in one of the four ways: None, Little (1-2 days), Sufficient, Advanced.

All in all, with these adaptations and alterations, the second part of the questionnaire consisted of 4 different sections and 112 items in total. Similar to the original instrument, each of these sections included both Part A and Part B in that in Part A participants were required to answer the training that they have received in the given item whereas in Part B they were required to answer the training that they think they need in the same given item. In other words, the items given in Part A and Part B of each section of the questionnaire were the same; however, these items were asked and required to be answered separately in accordance with the training received in Part A and the training needed in Part B. In addition, a 4-point Likert-type scale was used in Part A whereas a 3-point Likert-type scale was employed in Part B.

The first section, which examined the training received and the need for further training in *Classroom-focused LTA* domain, consisted of 36 items (18 items in Part A; 18 items in Part B). In Part A, the participants were required to give an answer for the training received in the given items. That is, participants were asked to rate the training they think they have received in the given item by choosing from a 4-point Likert-type scale (None; Little (1-2 days); Sufficient; Advanced) in this part. In Part B, the participants were required to give an answer for the training needed in the same given items. In other words, they were asked to rate the training they think they need in the given item by choosing from a 3-point Likert-type scale (None; Yes, basic training; Yes, more advanced training) in Part B. 6 of these items in both parts were taken from Vogt and Tsagari's (2014) study as stated above, and the other 12 items in both parts were added based on examining the books of the field in detail (Brown & Abeywickrama, 2010; Hughes, 2003) to get a thorough understanding of the extent to which preservice language teachers' received and need training in this domain.

The second section, which aimed to examine the training received and the need for further training in *Knowledge of Testing and Assessment* domain, was produced based on the book of Brown & Abeywickrama (2010) by the researcher in that these items were considered as necessary and essential components of LAL. This part included 18 items (9 items in Part A; 9 items in Part B) and in Part A preservice language teachers were required to specify the training they think they have received in the given item by choosing from the same 4-point Likert-type scale ranging from None to Advanced. As for the Part B, they were asked to specify the training they think they need in the given item by choosing from a 3-point Likert-type scale ranging from None to Yes, more advanced training.

As for the third section, it included 14 items (7 items in Part A; 7 items in Part B) and aimed to elicit information about the training received and the need for further training in *Purposes of Testing* domain. 4 of these items in both parts were also taken from the study of Vogt and Tsagari (2014) and the other 3 items in both parts were included by the researcher on the ground that these were also related to the types of tests as stated by Brown and Abeywickrama (2010), and accordingly related to the purposes of testing. The participants were asked to state the training they think they have received and they need in the given item by choosing from both the 4-point and the 3-point Likert-type scale separately.

In the last section, preservice language teachers were asked to specify the training received and the need for further training in *Content and concepts of LTA* domain by choosing from both the 4-point and the 3-point Likert-type scale separately. It consisted of 44 items (22 items in Part A; 22 items in Part B). 2 of the items in both parts were directly taken from the questionnaire of Vogt and Tsagari (2014); 5 of them were adapted and some alterations were made like dividing some items into two; thereby making them more specific, and rephrasing some items in order to gain a deeper and detailed understanding of the participants' LAL and training needs in this domain. As for the other 15 items of the two parts, they were created by the researcher after the detailed examination of the books by Brown & Abeywickrama (2010) and Hughes (2003) so that this part would comprise the basics of LTA (e.g., *Different test items/task types to test reading in English, Testing pronunciation in English…*).

To compute the internal consistency reliability of the questionnaire, Cronbach's alpha was used.

Table 4

*The Reliability Scale of the Data Collection Instrument*

|  | Number of Items | Cronbach alpha (α) |
|---|---|---|
| Section 1 (In total) | 36 | .919 |
| Section 2 (In total) | 18 | .836 |
| Section 3 (In total) | 14 | .793 |
| Section 4 (In total) | 44 | .918 |
| All Sections | 112 | .963 |

As it is seen in the table, the internal consistency reliability level for section 1 was α = 0.919, for section 2 was α =0.836, for section 3 was α = 0.793, and for section 4 was α = 0.918. As for the internal consistency reliability of the instrument used as a whole, it was α= 0.963, which indicates a high level of reliability as Dörnyei (2010) states. In conclusion, similar to the original instrument in which Cronbach's alpha score was reported to be ranging from 0.80 to 0.93 for individual scales, the instrument used in the present study had a quite satisfactory value (α = 0.963).

With regard to the content validity of the items, the researcher showed it to 2 experts in the field of foreign language education who approved the items in terms of their convenience for the participants. Besides, the questionnaire was also administered to 3 preservice language teachers who did not take part in this study. They did not find any problems in the questionnaire and did not suggest any alterations. After the piloting, the researcher conducted this questionnaire with preservice language teachers (See Appendix B for the questionnaire)

**The Interviews.** A semi-structured interview was preferred for this study as it provides the researcher with a written list of questions to follow but meanwhile freedom of deviation (Mackey & Gass, 2005). In this regard, it enables the researcher to ask for more information when the responses need more clarification and elaboration, and accordingly the interviewees to express their further comments and even to share some anecdotes regarding the questions, thereby providing more thorough insights and details about their opinions.

The questions in the interviews were produced by the researcher in accordance with the purpose of this study and they were reviewed by the research supervisor (see Appendix C for the interview questions in English). After reviewing, some questions were altered as wording might cause misunderstandings among the interviewees. Also, the questions that would be used in the interviews were translated into Turkish by the researcher (see Appendix D for the interview questions in Turkish) and these were also reviewed by the research supervisor. Besides, the interview process was piloted before administering the interviews to enable the researcher to both find out whether there was a problem in understanding or wording of the questions and to reflect on her interviewing skills. The piloted interview was also recorded so that the researcher could listen to the recording, and accordingly improve her interviewing skills.

There were seven questions in total and these were designed as open-ended questions in order to obtain rich and in-depth data from the participants. All the questions were opinion questions as the purpose of the study is to find out prospective teachers' opinions on both whether and how the theory and practice of assessment was covered and whether and how they should be covered in school experience course, and the usefulness of the ELTE course regarding their school experience course and practicum. The interview questions are given below:

**Interview Questions**

1. Has your school experience course covered anything related to the theory of LTA? If so, please list the topics that you learnt during school experience course.
2. Do you think that topics related to the theory of LTA should be covered in school experience course? Why or why not?
3. Has your school experience course/practicum given you the chance to put what you learned related to language testing and assessment into practice? If so, what are these practices?
4. Do you think that the practice related to the language testing and assessment should be realized in school experience course? Why or why not?
5. Did you have a chance to reflect on issues related to LTA in school experience course? If so, what are they?

6. Which topics related to language testing and assessment should be covered in school experience course, and in which ways do these help you?

7. Did taking ELTE course help you during your school experience/ practicum? If so, in which ways it was useful?

As seen above, the 1st, 2nd, 3rd and 4th questions aim to find out prospective teachers' opinions on whether the school experience course has covered the theory and practice of assessment and if this is the case, what has been covered as well as whether they should be covered and realized in this course. The 6th question aims to find out prospective teachers' opinions on the topics that should be covered in the course and how these topics would be helpful for them. As for the 5th question, it aims to uncover the opinions of prospective teachers of whether they have reflected on issues regarding LTA and if this is the case, what has been reflected. The last question aims to reveal the prospective teachers' opinions of the usefulness of the ELTE course in terms of the school experience and their practicum.

**Data Collection Procedure and Analysis**

In this part, a detailed picture of data collection procedure and a general view of phases in data analyses will be presented.

**Data Collection Procedure.** Prior to data collection process, the researcher applied for the permission of Hacettepe University Ethics Commission. Some documents including the original questionnaire, adaptation consent from the developers of the questionnaire, the adapted version of the questionnaire both in Turkish and English, and the interview questions both in Turkish and English were submitted to the commission. Following the investigation process, the commission approved that this study conformed to the ethical principles of Hacettepe University (See Appendix F) and it could be conducted in the way it was planned. After getting permission from the commission, data collection instruments were employed to gather data.

*Questionnaire.* The participants of this study were chosen through convenience sampling. As the researcher was a research assistant in ELT

Department of Middle East Technical University at that time, she gathered data with the 4th grade students of her own institution. She visited all of the four sections of the ELTE course with the permission of the instructors of this course. At the beginning, the researcher gave information about the purpose of the questionnaire in relation to the aim of the study, and stated that the data from the questionnaires would not be shared by any other people or institutions and it would be used only for the purposes of this study. They were also said that their contribution was valuable for this study; therefore, they were asked to be sincere and realistic while answering the questions. Following these explanations, the participants were asked to sign the official consent form (See Appendix E). In total, 110 students accepted taking part in it. However, 9 of the participants did not fill the questionnaires completely, so they were excluded from the study. In this regard, 101 of the questionnaires were analyzed.

The researcher was always in the classroom during the administration of the questionnaires, which enabled the researcher to answer the questions of the participants and to explain the points that need clarification. The participants were given enough time to complete the questionnaire, which took nearly 20 minutes.

*Interview.* The interviews were carried out with 25 preservice language teachers in an office that would provide a relaxed atmosphere for the participants. The interviews had been done in Turkish, which was the native language of the participants on the ground that students may not express themselves well and not elaborate on their ideas in detail, or may hesitate while speaking in English; in other words, interviews were held in Turkish in order to avoid these possible problems, and to get thorough and sincere answers for the interview questions.

At the beginning of the interviews, the interviewer gave information about the purpose of the interview in relation to the aim of the study and assured that the names of the interviewees would be kept anonymous while reporting the results of the analysis. Since the participants were assured about the confidentiality, audio recording that would be used in the interviews did not bother the participants and all of them allowed their responses to be recorded. Following these, the researcher started to ask semi-structured questions; however, the interviews were not just restricted to these questions, instead the interviewees

were also encouraged to share some anecdotes related to the directed questions and explain their further comments.

During the interviews, the researcher did not interrupt the participants while they were speaking even when they diverged from the issue about which they should be talking. That is, the researcher did not intervene in the procedure and waited for some relevant data to occur, which made transcribing difficult as every word uttered by the interviewees was transcribed. After the interviews which lasted for approximately 20 minutes were completed, the researcher transcribed all of the interviews and translated the extracts taken from the interviews into English for submitting it to an English medium institution and for the non-Turkish readers of this study.

**Data Analysis.** As the present study employed a mixed-methods research design, both quantitative and qualitative data analyses were employed to answer the research questions of the study.

In order to analyze the collected quantitative data, they were entered into Statistical Package for Social Sciences (SPSS). However, prior to analyzing the data on SPSS, a test of normality was performed through Kolmogorov-Smirnov and Shapiro-Wilk tests in order to measure whether the data were normally distributed or not. The table below shows the normality level of the questionnaire.

Table 5

*Normality Test of the Questionnaire*

|  | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
|  | Statistic | Df | Sig. | Statistic | df | Sig. |
| Questionnaire | .064 | 101 | .200 | .980 | 101 | .127 |

a. Lilliefors Significance Correction

As it is seen in the table, both Kolmogorov-Smirnov (p=.200 > .05) and Saphiro-Wilk tests (p=.127 > .05) reveal that the data are normally distributed and parametric statistical tests can be employed in order to analyze the research questions.

Besides the numerical data given above, histogram and normal Q-Q plot provide visual cues about the test of normality. Accordingly, histogram and normal

Q-Q plot given below (Figures 1 and 2) reveal the normality test and indicate a normal distribution, supporting the numerical result provided above.
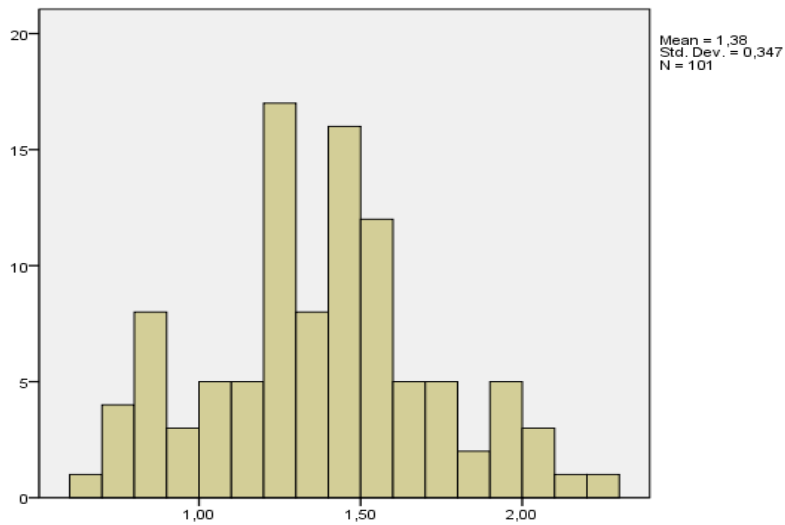


*Figure 1*. Histogram for normality test of the questionnaire.
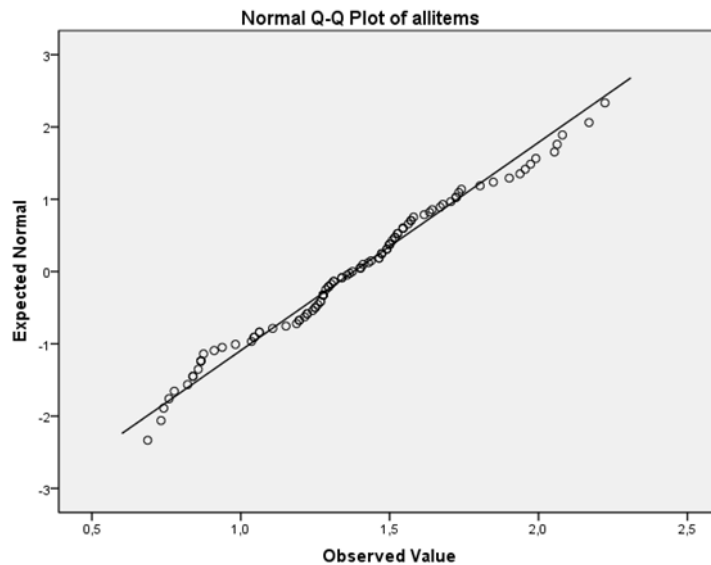


*Figure 2.* Normal q-q plot for normality test of the questionnaire.

Following the test of normality, the quantitative data were analyzed by using descriptive statistics. Initially, the descriptive statistics including percentages and frequencies of participants on various criteria (e.g., gender, age, type of high school, GPA) were calculated. Then, the Likert scale data were analyzed through the means, frequencies and percentages obtained for each item through SPSS program. In this regard, for the 1st and 2nd research questions, descriptive

statistics was applied through calculating the mean values, frequencies and percentages of the training "received" and "needed" in LTA.

As for the qualitative data, it was analyzed by touching on all interview items one by one, and the grounded theory principles were taken into account (Corbin & Strauss, 1990) in that participants' responses were classified in accordance with their similarities and differences. After classifying these responses, selective coding was employed. It is one of the coding types in grounded theory research and focuses on "the main analytic idea presented in the research" (Corbin & Strauss, 1990, p. 14). On the other hand, as the interview was semi-structured and allowed the interviewer not to be strictly restricted to the interview items, the interviewer also encouraged the interviewees to share some anecdotes related to the directed questions, and posed different questions in accordance with the responses given by the interviewees.

**Chapter 4**

**Findings**

**Quantitative Data Analysis**

With the aim of exploring preservice language teachers' LAL in terms of the training that they received in LTA and identifying their training needs in this area, questionnaires were conducted with 101 participants. The questionnaire included 4 different sections in a Likert scale format each including Part A and Part B in that in the first part they were asked to answer the training that they received in the given item whereas in the second part they were required to answer the training that they think they need in the same given item. All the Likert scale data obtained through this questionnaire were analyzed through descriptive statistics using SPSS program. Findings of the quantitative data analysis are presented below under the first and second research questions of the current study.

**Results of Data Analysis for Research Question 1.** The first research question aimed at investigating preservice language teachers' LAL in terms of the training that they have received in LTA. It included four sub-parts. In the first part, it was aimed to find out the level of training preservice language teachers think they have received in *Classroom-focused LTA*. The second part tried to investigate the level of training preservice language teachers think they received in the domain of *Knowledge of Testing and Assessment.* As for the third part, it aimed at finding out the level of training they think they received in the domain of *Purposes of Testing.* The last part aimed at investigating the level of training they think they received in the domain of *Content and concepts of LTA*. Accordingly, the amount of training they have received in different domains of LTA was examined.

Part A of the four sections in the questionnaire including 56 items in total was designed to answer the first research question. To this end, the preservice language teachers' responses to these 56 items were ranked from 0 to 3 (0 referring to None, 1 to Little (1-2 days), 2 to Sufficient, and 3 to Advanced) and presented in four sections.

The data for the first part was gathered through part A of the first section in the questionnaire that include 18 items in total. The table below (Table 6) presents

the results of "*Classroom-focused Language Testing and Assessment*" based on the participants' perceptions of the training they received in this domain.

Table 6

*Participants' Perceptions of Training Received in Classroom-Focused LTA*

|  | N | Mean[a] | SD | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Preparing classroom tests | 101 | 1.73 | .646 | 2 | 32 | 58 | 9 |
| Preparing diagnostic tests | 101 | 1.32 | .706 | 12 | 47 | 40 | 2 |
| Preparing achievement tests | 101 | 1.52 | .672 | 6 | 40 | 51 | 4 |
| Preparing proficiency tests | 101 | 1.28 | .709 | 12 | 52 | 34 | 3 |
| Preparing placement tests | 101 | 1.12 | .791 | 24 | 43 | 32 | 2 |
| Preparing progress tests | 101 | 1.53 | .867 | 12 | 36 | 40 | 13 |
| Preparing language aptitude tests | 101 | 1.04 | .799 | 29 | 40 | 31 | 1 |
| Using ready-made tests from textbook packages or from other sources | 101 | 1.76 | .814 | 7 | 27 | 50 | 17 |
| Adapting ready-made tests for the needs of students | 101 | 1.89 | .786 | 4 | 25 | 50 | 22 |
| Stages of language test construction (e.g. objectives, drawing up test specifications...) | 101 | 1.81 | .796 | 4 | 31 | 46 | 20 |
| Scoring | 101 | 1.63 | .845 | 9 | 34 | 43 | 15 |
| Grading | 101 | 1.56 | .830 | 10 | 36 | 43 | 12 |
| Giving feedback to students based on information from tests/assessment | 101 | 1.57 | .853 | 10 | 37 | 40 | 14 |
| Interpreting test | 101 | 1.47 | .890 | 13 | 42 | 32 | 14 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| scores | | | | | | | |
| Using self/peer assessment | 101 | 1.57 | .898 | 12 | 35 | 38 | 16 |
| Using informal,non-test type of assessment (e.g. essays, presentations, homeworks) | 101 | 1.55 | .900 | 15 | 28 | 45 | 13 |
| Using continuous type of assessment (e.g. quizzes) | 101 | 1.62 | .936 | 13 | 31 | 38 | 19 |
| Using European Language Portfolio | 101 | 1.18 | .942 | 27 | 39 | 25 | 10 |

a.Means are based on a 4-point scale: 0, None; 1, Little (1-2 days); 2, Sufficient; 3, Advanced.

In addition to descriptive statistics regarding the results for training received in Classroom-focused LTA, the overall mean score of 18 items was calculated and found to be 1.50, implying that the preservice language teachers find the training received in Classroom-focused LTA insufficient.

As illustrated in table 6, the preservice language teachers appeared to have the highest mean value in "adapting ready-made tests for the needs of students" ($M$=1.89, SD=.786) indicating that they had received most training in this area of Classroom-Focused LTA. It was followed by "stages of language test construction" ($M$=1.81, $SD$=.796) with a slightly lower mean value. The third highest score was given for the "using ready-made tests from textbook packages or from other sources" component of Classroom-focused LTA with 1.76, while "preparing classroom tests" component had a slightly lower mean score of 1.73. However, the lowest mean score ($M$=1.04) was found for "preparing language aptitude tests", as the majority of the participants (39.6 % and 28.7 %, respectively) reported little or none training in this area. Furthermore, "preparing placement tests" had the second lowest mean score with 1.12, as nearly one fourth (23.8 %) of the participants reported no training in it. The third lowest score was given for "using European Language Portfolio" component with a mean score of 1.18.

Part A of the second section in the questionnaire included 9 items. Table 7 below presents the results of "*Knowledge of Testing and Assessment*" based on the participants' perceptions of the training received in this domain.

Table 7

*Participants' Perceptions of Training Received in Knowledge of Testing and Assessment*

| | N | Mean[a] | SD | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Informal / Formal assessment | 101 | 1.92 | .717 | 0 | 30 | 49 | 22 |
| Formative / Summative assessment | 101 | 1.91 | .750 | 2 | 27 | 50 | 22 |
| Norm /Criterion-referenced assessment | 101 | 1.78 | .795 | 5 | 30 | 48 | 18 |
| Discrete point / Integrative testing | 101 | 1.83 | .813 | 5 | 28 | 47 | 21 |
| Direct / Indirect testing | 101 | 1.96 | .720 | 0 | 28 | 49 | 24 |
| Objective / Subjective testing | 101 | 1.97 | .754 | 2 | 24 | 50 | 25 |
| Approaches to language testing (e.g. integrative,communicative, structuralist) | 101 | 1.61 | .916 | 12 | 33 | 38 | 18 |
| Alternative assessment | 101 | 1.45 | .900 | 18 | 30 | 43 | 10 |
| Computer-based testing | 101 | 1.15 | .910 | 29 | 34 | 32 | 6 |

a.Means are based on a 4-point scale: 0, None;  1, Little (1-2 days);  2, Sufficient; 3, Advanced.

Analysis shows the results for training received in the domain of *Knowledge of Testing and Assessment* with an overall mean score of 1.73, which reveals that the preservice language teachers perceive the training received in Knowledge of Testing and Assessment not to be sufficient.

The results, as indicated in Table 7, demonstrated that the highest mean score was ascribed to the "objective / subjective testing" (M= 1.97, SD=.754) component of Knowledge of Testing and Assessment domain, suggesting that nearly half (49.5 %) of the participants reported sufficient training in it. The second highest mean (*M*=1.96, SD=.72) was found for "direct / indirect testing", indicating that nearly one in two (48.5 %) received sufficient training in this component. It was followed by "informal / formal assessment" (*M*=1.92, SD= .717) and "formative / summative assessment" (*M*=1.91, SD=.75) with slightly lower mean values. In contrast, the lowest mean score was reported in "computer-based testing" (*M*=1.15, SD=.91), indicating that 33.7 % of the participants stated that they had received little training in this component. "Alternative assessment" had the second lowest mean score with 1.45, as nearly one fourth (29.7 %) of the participants reported little training in this component.

Part A of the third section in the questionnaire included 7 items. Table 8 below shows the results of *"Purposes of Testing"* based on preservice language teachers' perceptions of the training that they received in terms of descriptive statistics.

Table 8

*Participants' Perceptions of Training Received in Purposes of Testing*

|  | N | Mean[a] | SD | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Giving grades | 101 | 1.67 | .789 | 7 | 32 | 49 | 13 |
| Finding out what needs to be learned/ taught | 101 | 1.76 | .777 | 5 | 30 | 50 | 16 |
| Placing students onto programs, courses, etc. | 101 | 1.27 | .847 | 20 | 40 | 35 | 6 |
| Testing competence in a language | 101 | 1.58 | .791 | 11 | 28 | 54 | 8 |
| Identifying what has been learned | 101 | 1.83 | .813 | 5 | 28 | 47 | 21 |
| Measuring general ability to learn a foreign language | 101 | 1.39 | .871 | 18 | 34 | 41 | 8 |
| Awarding final certificates (from school, program,; local, regional / national level) | 101 | .89 | .786 | 37 | 38 | 26 | 0 |

a. Means are based on a 4-point scale: 0, None;  1, Little (1-2 days);  2, Sufficient; 3, Advanced.

Descriptive statistics reveals the results of the training received in the domain of Purposes of Testing with an overall mean score of 1.48, indicating that the preservice language teachers perceive the training that they received in this domain not to be sufficient.

As Table 8 indicates, the preservice language teachers appeared to have the highest mean value in "identifying what has been learned" (*M*=1.83, SD=.813). This indicates that nearly 50% of the participants perceive that they have received sufficient training in this component. "Finding out what needs to be learned/ taught" (*M*=1.76, SD=.777) follows "identifying what has been learned" component with a close mean value, suggesting that nearly five in ten (49.5%) of the participants perceive the training in finding out what needs to be learned/ taught as sufficient. In contrast, the lowest mean score (*M*=.89, SD=.786) was observed in "awarding final certificates" as the majority of the participants (36.6% and 37.6%, respectively) reported no or little training in this area. Furthermore, the second lowest mean score (*M*= 1.27) was found for "placing students onto programs,

courses, etc" since nearly four in ten (39.6%) of the participants stated that the training they received in this area was little.

Part A of the last section in the questionnaire consisted of 22 items. Table 9 below presents the results of "*Content and concepts of LTA*" based on the preservice language teachers' perceptions of the amount of training they received in terms of descriptive statistics.

Table 9

*Participants' Perceptions of Training Received in Content and concepts of LTA*

| | N | Mean[a] | SD | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Testing reading in English | 101 | 1.93 | .738 | 0 | 31 | 46 | 24 |
| Different test items/task types to test reading in English | 101 | 1.93 | .752 | 2 | 26 | 50 | 23 |
| Testing listening in English | 101 | 1.91 | .750 | 0 | 33 | 44 | 24 |
| Different test items/task types to test listening in English | 101 | 1.87 | .757 | 2 | 30 | 48 | 21 |
| Testing speaking in English | 101 | 1.77 | .799 | 3 | 37 | 41 | 20 |
| Different test items/task types to test speaking in English | 101 | 1.71 | .804 | 5 | 36 | 43 | 17 |
| Testing writing in English | 101 | 1.80 | .813 | 3 | 36 | 40 | 22 |
| Different test items/task types to test writing in English | 101 | 1.79 | .816 | 4 | 34 | 42 | 21 |
| Testing grammar in English | 101 | 1.85 | .817 | 1 | 39 | 35 | 26 |
| Different test items/task types to test grammar in English | 101 | 1.82 | .792 | 2 | 36 | 41 | 22 |
| Testing vocabulary in English | 101 | 1.82 | .817 | 2 | 38 | 37 | 24 |
| Different test items/task types to test vocabulary in English | 101 | 1.77 | .799 | 3 | 37 | 41 | 20 |
| Testing integrated language skills | 101 | 1.48 | .879 | 15 | 34 | 41 | 11 |
| Testing pronunciation in English | 101 | 1.13 | .913 | 29 | 37 | 28 | 7 |
| Different test items/task types to test pronunciation in English | 101 | 1.20 | .917 | 25 | 40 | 27 | 9 |
| Practicality | 101 | 1.70 | .878 | 5 | 43 | 30 | 23 |
| Reliability (e.g. rater, test administration, test, student-related) | 101 | 2.00 | .849 | 2 | 30 | 35 | 34 |
| Validity (e.g. face, construct, criterion, construct) | 101 | 2.02 | .836 | 2 | 28 | 37 | 34 |
| Authenticity | 101 | 2.03 | .900 | 2 | 33 | 26 | 40 |
| Washback | 101 | 1.88 | .920 | 5 | 34 | 30 | 32 |
| Using statistics to study the quality of tests / assessment | 101 | .80 | .825 | 41 | 44 | 11 | 5 |
| Alternatives in assessment (e.g. portfolios, conferences, interviews, observations, self/peer assessment) | 101 | 1.20 | .883 | 23 | 43 | 27 | 8 |

a.Means are based on a 4-point scale: 0, None;  1, Little (1-2 days);  2, Sufficient; 3, Advanced.

Given the results of descriptive analyses regarding the participants' perceptions of the training received in this domain, it was found that the overall mean score of the 22 items was 1.70, indicating that the preservice language teachers find the training received in Content and concepts of LTA inadequate. As presented in the table, the preservice language teachers appeared to have the highest mean value in "authenticity" ($M$=2.03, SD=.90) since the majority of the participants (39.6% and 25.7%, respectively) reported advanced or sufficient training in this area. "Validity" ($M$=2.02, SD=.836) and "reliability" ($M$=2, SD=.849) follow "authenticity" with very close mean values. Furthermore, in most areas of LTA covered in this domain, nearly half of the preservice language teachers reported that they had received sufficient training: "different test items/task types to test reading in English" (49.5%), "different test items/ task types to test listening in English" (47.5%), "testing reading in English" (45,5%), "testing listening in English" (43.6%), "Different test items/task types to test speaking in English" (42.6%).

On the other hand, the lowest mean score ($M$=.80, SD=.825) was found for "using statistics to study the quality of tests / assessment" since the majority of the participants (43.6% and 40.6%, respectively) stated little or no training in this area.

"Testing pronunciation in English" had the second lowest mean score with 1.13, as more than one fourth (28.7%) of the participants reported no training in it. The third lowest score ($M$=1.20) was given for "different test items/task types to test pronunciation in English" and "alternatives in assessment" components, as nearly four in ten (39.6% and 43.6%, respectively) stated that they had received little training in "different test items/task types to test pronunciation in English" and "alternatives in assessment" components.

All in all, preservice language teachers' responses to these 56 items included in the part A of the four sections in the questionnaire were analyzed to answer the first research question. Table 10 presents the amount of training preservice language teachers state they have received in four areas of LTA in terms of descriptive statistics.

Table 10

*Respondents' Perceptions of Training Received in Four Areas of LTA*

| Areas of Language Testing and Assessment (LTA) Literacy | Received Training | | |
|---|---|---|---|
| | N | Mean[a] | SD |
| Classroom-focused LTA | 101 | 1.5 | .547 |
| Knowledge of Testing and Assessment | 101 | 1.73 | .648 |
| Purposes of Testing | 101 | 1.48 | .633 |
| Content and concepts of LTA | 101 | 1.7 | .637 |
| Total | 101 | 1.61 | .537 |

a.Means are based on a 4-point scale: 0, None;  1, Little (1-2 days);  2, Sufficient; 3, Advanced.

Descriptive statistics shows the results for training received in four domains of LTA literacy and the overall mean score of the 56 items was 1.61, suggesting that the preservice language teachers perceive the training that they received in these domains not to be sufficient.

As presented in the table, preservice language teachers appeared to have the highest mean value in *"Knowledge of Testing and Assessment"* (*M*=1.73, SD=.648) domain. It was followed by *"Content and concepts of LTA"* (*M*=1.7, SD=.637) and *"Classroom-focused LTA"* (*M*=1.5, SD=.547) domains. *"Purposes of Testing"* (*M*=1.48, SD=.633) had the lowest mean value, suggesting that preservice language teachers perceive the amount of training they received in this domain to be less than other three domains.

**Results of Data Analysis for Research Question 2.** The second research question aimed at investigating preservice language teachers' LAL in terms of the further training that they think they need in LTA. It included four sub-parts. The first one aimed to find out the level of further training preservice language teachers think they need in the domain of *Classroom-focused LTA*. In the second part, it was aimed to investigate the level of further training preservice language teachers think they need in *Knowledge of Testing and Assessment* domain. As to the third part, it aimed at investigating the level of further training they think they need in the domain of *Purposes of Testing*. The last part aimed at finding out the level of

further training they think they need in the domain of *Content and concepts of LTA*. Accordingly, the amount of further training preservice language teachers state they need in different domains of LTA was examined.

Part B of the four sections in the questionnaire including 56 items in total was designed to answer the second research question. To this end, the preservice language teachers' responses to these 56 items were ranked from 0 to 2 (0 referring to None, 1 to Yes, basic training, and 2 to Yes, more advanced training) and presented in four sections.

The data for the first part was gathered through part B of the first section in the questionnaire that include 18 items in total. The table below (Table 11) presents the results of "*Classroom-focused LTA*" based on the participants' perceptions of the training needed in this domain.

Table 11

*Participants' Perceptions of Training Needed in Classroom-focused Language Testing and Assessment*

| | N | Mean[a] | SD | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| Preparing classroom tests | 101 | 1.35 | .713 | 14 | 38 | 49 |
| Preparing diagnostic tests | 101 | 1.34 | .621 | 8 | 51 | 42 |
| Preparing achievement tests | 101 | 1.31 | .689 | 13 | 44 | 44 |
| Preparing proficiency tests | 101 | 1.32 | .615 | 8 | 53 | 40 |
| Preparing placement tests | 101 | 1.26 | .688 | 14 | 47 | 40 |
| Preparing progress tests | 101 | 1.18 | .740 | 20 | 43 | 38 |
| Preparing language aptitude tests | 101 | 1.28 | .695 | 14 | 45 | 42 |
| Using ready-made tests from textbook packages or from other sources | 101 | 1.02 | .800 | 31 | 37 | 33 |
| Adapting ready-made tests for the needs of students | 101 | 1.03 | .768 | 28 | 42 | 31 |
| Stages of language test construction (e.g. objectives, drawing up test specifications...) | 101 | 1.05 | .779 | 28 | 40 | 33 |
| Scoring | 101 | 1.10 | .755 | 24 | 43 | 34 |
| Grading | 101 | 1.15 | .740 | 21 | 44 | 36 |
| Giving feedback to students based on information from tests/assessment | 101 | 1.19 | .758 | 21 | 40 | 40 |
| Interpreting test scores | 101 | 1.26 | .702 | 15 | 45 | 41 |
| Using self/peer assessment | 101 | 1.21 | .712 | 17 | 46 | 38 |
| Using informal,non-test type of assessment (e.g. essays, presentations, | 101 | 1.10 | .742 | 23 | 45 | 33 |

| | | | | homeworks) | | | |
|---|---|---|---|---|---|---|---|

| | N | Mean | SD | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| Using continuous type of assessment (e.g. quizzes) | 101 | 1.10 | .728 | 22 | 47 | 32 |
| Using European Language Portfolio | 101 | 1.27 | .691 | 14 | 46 | 41 |

a.Means are based on a 3-point scale: 0, None;  1, Yes, basic training;  2, Yes, more advanced training.

Descriptive statistics reveals the results of the further training needed in the domain of Classroom-focused LTA with an overall mean score of 1.19, indicating that the preservice language teachers perceive a need for further basic training in this field of LTA.

The results, as illustrated in Table 4.6, demonstrated that the highest mean score was ascribed to the "preparing classroom tests" (M=1.35, SD=.713) since the majority of the participants (48.5% and 37.6%, respectively) reported advanced or basic training need in this area. "Preparing diagnostic tests" (*M*=1.34) follows "preparing classroom tests" with a very close mean score. In contrast, the lowest mean score was found for "using ready-made tests from textbook packages or from other sources" (*M*=1.02, SD=.80) as more than one fourth (30.7%) of the participants reported no need for further training in it. "Adapting ready-made tests for the needs of students" had the second lowest mean score with 1.03, as more than one fourth (27.7%) of the participants reported no need for further training in it.

The data for the second part was gathered through part B of the second section in the questionnaire that include 9 items in total. Table 12 below shows the results of *"Knowledge of Testing and Assessment"* based on the participants' perceptions of the amount of training they need further in this domain.

Table 12

*Participants' Perceptions of Training Needed in Knowledge of Testing and Assessment*

| | N | Mean[a] | SD | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| Informal / Formal assessment | 101 | .95 | .779 | 33 | 40 | 28 |
| Formative / Summative assessment | 101 | 1.00 | .800 | 32 | 37 | 32 |
| Norm / Criterion-referenced assessment | 101 | 1.00 | .787 | 31 | 39 | 31 |
| Discrete point / Integrative testing | 101 | 1.00 | .787 | 31 | 39 | 31 |
| Direct / Indirect testing | 101 | 1.01 | .768 | 29 | 42 | 30 |
| Objective / Subjective testing | 101 | .90 | .794 | 37 | 37 | 27 |

| | N | Mean[a] | SD | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| Approaches to language testing (e.g. integrative, communicative, structuralist) | 101 | 1.17 | .762 | 22 | 40 | 39 |
| Alternative assessment | 101 | 1.17 | .694 | 17 | 50 | 34 |
| Computer-based testing | 101 | 1.20 | .735 | 19 | 43 | 39 |

a.Means are based on a 3-point scale: 0, None;  1, Yes, basic traning;  2, Yes, more advanced traning.

Given the results of descriptive analyses regarding the participants' perceptions of the level of training needed in this domain, it was found that the overall mean score of the 9 items was 1.04, which reveals that the preservice language teachers perceive a need for further basic training in this field of LTA.

As Table 12 indicates, the preservice language teachers appeared to have the highest mean value in "computer-based testing" ($M$=1.2, SD=.735). This indicates that nearly four in ten (42.6%) of the participants perceive that they need basic training in this component. The second highest score ($M$=1.17) was found for "approaches to language testing" and "alternative assessment" components. In contrast, "objective / subjective testing" had the lowest mean score with .90, since nearly four in ten (36.6%) of the participants stated that they do not need any further training in it. "Informal / Formal assessment" had the second lowest mean score with .95, indicating that more than one fourth (32.7%) of the participants reported no need for further training in it.

In the questionnaire, part B of the third section included 7 items. Table 13 below presents the results of *"Purposes of Testing"* based on the preservice language teachers' perceptions of the training they need in this domain.

Table 13

*Participants' Perceptions of Training Needed in Purposes of Testing*

| | N | Mean[a] | SD | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| Giving grades | 101 | .96 | .734 | 29 | 47 | 25 |
| Finding out what needs to be learned/ taught | 101 | 1.01 | .755 | 28 | 44 | 29 |
| Placing students onto programs, courses, etc. | 101 | 1.13 | .627 | 14 | 60 | 27 |
| Testing competence in a language | 101 | 1.14 | .749 | 22 | 43 | 36 |
| Identifying what has been learned | 101 | 1.00 | .800 | 32 | 37 | 32 |
| Measuring general ability to learn a foreign language | 101 | 1.11 | .691 | 19 | 52 | 30 |
| Awarding final certificates (from school, | 101 | 1.24 | .723 | 17 | 43 | 41 |

program,; local, regional / national level)

a.Means are based on a 3-point scale: 0, None;  1, Yes, basic training;  2, Yes, more advanced training.

Descriptive statistics reveals the results of the training needed in the domain of Purposes of Testing with an overall mean score of 1.08, which reveals that the amount of further training preservice language teachers need is basic. The results, as illustrated in Table 13, demonstrated that the highest mean score was ascribed to the "awarding final certificates" (*M*=1.24, SD=.723) component of Purposes of Testing domain, suggesting that the majority of the participants (42.6% and 40.6%, respectively) reported basic or advanced training need in it. It was followed by "testing competence in a language" (*M*= 1.14, SD=.749) and "placing students onto programs, courses,etc." (*M*=1.13, SD=.627) with slightly lower mean values. "Giving grades" had the lowest mean score with .96, since more than one fourth (28.7%) of the preservice language teachers reported no need for further training in it.

In the questionnaire, part B of the last section included 22 items. Table 14 below presents the results of *"Content and concepts of LTA"* based on the preservice language teachers' perceptions of the training they need in this domain.

Table 14

*Participants' Perceptions of Training Needed in Content and concepts of LTA*

| | N | Mean[a] | SD | 0 | 1 | 2 |
|---|---|---|---|---|---|---|
| Testing reading in English | 101 | 1.05 | .792 | 29 | 38 | 34 |
| Different test items/task types to test reading in English | 101 | 1.11 | .799 | 27 | 36 | 38 |
| Testing listening in English | 101 | 1.03 | .806 | 31 | 36 | 34 |
| Different test items/task types to test listening in English | 101 | 1.14 | .788 | 25 | 37 | 39 |
| Testing speaking in English | 101 | 1.21 | .779 | 22 | 36 | 43 |
| Different test items/task types to test speaking in English | 101 | 1.21 | .804 | 24 | 32 | 45 |
| Testing writing in English | 101 | 1.14 | .749 | 22 | 43 | 36 |
| Different test items/task types to test writing in English | 101 | 1.20 | .762 | 21 | 39 | 41 |
| Testing grammar in English | 101 | 1.12 | .804 | 27 | 35 | 39 |
| Different test items/task types to test grammar in English | 101 | 1.16 | .784 | 24 | 37 | 40 |
| Testing vocabulary in English | 101 | 1.12 | .765 | 24 | 41 | 36 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Different test items/task types to test vocabulary in English | 101 | 1.16 | .771 | 23 | 39 | 39 |
| Testing integrated language skills | 101 | 1.27 | .747 | 18 | 38 | 45 |
| Testing pronunciation in English | 101 | 1.34 | .697 | 13 | 41 | 47 |
| Different test items/task types to test pronunciation in English | 101 | 1.24 | .737 | 18 | 41 | 42 |
| Practicality | 101 | 1.05 | .792 | 29 | 38 | 34 |
| Reliability (e.g. rater, test administration, test, student-related) | 101 | 1.04 | .824 | 32 | 33 | 36 |
| Validity (e.g. face, construct, criterion, construct) | 101 | 1.10 | .819 | 29 | 33 | 39 |
| Authenticity | 101 | .99 | .806 | 33 | 36 | 32 |
| Washback | 101 | 1.01 | .794 | 31 | 38 | 32 |
| Using statistics to study the quality of tests / assessment | 101 | 1.47 | .657 | 9 | 36 | 56 |
| Alternatives in assessment (e.g. portfolios, conferences, interviews, observations, self/peer assessment) | 101 | 1.30 | .715 | 15 | 41 | 45 |

a.Means are based on a 3-point scale: 0, None; 1, Yes, basic training; 2, Yes, more advanced training.

Descriptive statistics reveals the results of the further training needed in the domain of Content and concepts of LTA with an overall mean score of 1.15, indicating that the preservice language teachers perceive a need for further basic training in this field of LTA.

As shown in Table 14, the results revealed that preservice language teachers appeared to have the highest mean value in "using statistics to study the quality of tests / assessment" (*M*=1.47, SD=.657), suggesting that 55.4% of the participants perceive that they need advanced training in this component. It was followed by "testing pronunciation in English" (*M*=1.34, SD=.697) and "alternatives in assessment" (*M*=1.3, SD=.715) with close mean values. In contrast, the lowest mean score was found for "authenticity" (*M*=.99, SD=.806), since more than one fourth (32.7%) of the participants reported no need for further training.

All in all, preservice language teachers' responses to these 56 items included in the part B of the four sections in the questionnaire were analyzed to answer the second research question. Table 15 presents the amount of training preservice language teachers state they need in four areas of LTA in terms of descriptive statistics.

Table 15

*Respondents' Perceptions of Training Needed in Four Areas of LTA*

| Areas of Language Testing and Assessment (LTA) Literacy | Needed Training | | |
|---|---|---|---|
| | N | Mean[a] | SD |
| Classroom-focused LTA | 101 | 1.19 | .496 |
| Knowledge of Testing and Assessment | 101 | 1.04 | .64 |
| Purposes of Testing | 101 | 1.08 | .562 |
| Content and concepts of LTA | 101 | 1.15 | .592 |
| Total | 101 | 1.14 | .484 |

a.Means are based on a 3-point scale: 0, None;  1, Yes, basic training;  2, Yes, more advanced training.

Analysis shows the results for training needed in four domains of LTA literacy, and the overall mean score of the 56 items was 1.14, which indicates that the preservice language teachers perceive the amount of training that they need further in these domains to be basic.

As shown in the table, preservice language teachers appeared to have the highest mean value in "*Classroom-focused LTA*" (*M*=1.19, SD=.496) domain. It was followed by *"Content and concepts of LTA"* (*M*=1.15, SD=.592) and *"Purposes of Testing"* (*M*=1.08, SD=.562) domains. *"Knowledge of Testing and Assessment"* (*M*=1.04, SD=.64) had the lowest mean value, suggesting that preservice language teachers perceive the amount of training they need further in this domain to be less than other three domains.

**Analysis of the Qualitative Data**

In order to explore preservice language teachers' opinions on both whether and how the theory and practice of language assessment was covered and whether and how they should be covered in school experience course, and the usefulness of English Language Testing and Evaluation (ELTE) course on their school experience course in depth, semi-structured interviews were held with 25 randomly chosen participants who also filled the questionnaire. All the qualitative data gathered through these interviews were analyzed through selective coding.

Findings of the qualitative data analysis are presented below the last research question of the present study.

**Results of Data Analysis for Research Question 3.** Open ended questions were posed to the participants during the interviews and responses obtained through these interviews were categorized under each interview question. Furthermore, the interviewees' names were kept confidential; therefore, they were represented with PT (preservice teacher) and a number so as to achieve anonymity.

**Has your school experience course covered anything related to the theory of LTA? If so, please list the topics that you learnt during school experience course.**

All of the participants stated that their school experience course did not cover anything with regard to the theory of LTA; therefore, the researcher did not ask the following question that requires the list of topics covered.

**Do you think that topics related to the theory of LTA should be covered in school experience course? Why or why not?**

All of the participants indicated that the theory of LTA should be covered in school experience course. When their responses to the reason why they think so were analyzed, it was realized that there were some common points that the participants had touched on. These common points were shown below:

- The relationship between language testing and assessment (LTA) and language teaching and learning processes
- Its importance for their future profession
- The chance to reflect on real language testing and assessment contexts with regard to the theoretical bases in school experience course

*Figure 3.* Comments of the preservice language teachers.

In this regard, the following comments were made:

*"We should cover these because we will need them in our future profession. As teachers, we will not only be engaged in lecturing, but also engaged in assessment both formally and informally. As the school experience course aims to prepare us well for our future profession, it should not ignore such*

*an important component. Without the theoretical knowledge of language testing and assessment, we cannot be effective teachers." (PT12)*

*"Of course, it should be because school experience course should be related to every aspect of language teaching and learning processes and assessment is one of these aspects. Without assessment, language teaching and learning processes will be incomplete and inauthentic because it is an inevitable part of this process." (PT7)*

*"... Because we could have the opportunity to reflect on these again in relation to real teaching and assessment contexts. Like we discuss real teaching contexts and our experience in school experience course, discussing real assessment contexts based on the theoretical framework will be perfect. Talking about real assessment contexts will also contribute to our theoretical understanding." (PT8)*

Moreover, another participant put forward the necessity of covering the theory of LTA in school experience course and explained its importance and rationale by making reference to the necessity of methodological knowledge that each teacher should have. Furthermore, she suggested a way to integrate the theory of LTA into the school experience course.

*"I think that it is as important as other methodological knowledge that we are expected to have. As we know, without language teaching methodology, we are not competent enough. Similarly, without adequate language testing and assessment knowledge, we cannot be competent as language teachers because it is one of the main components of language teaching. As for how it can be, in our observation tasks we can evaluate or analyze one of the tests applied during our practicum. In our class, we did an observation task based on our mentor teacher's assessment, but we could have taken an exam from the school and evaluate and discuss it in terms of the theoretical components we learnt in ELTE course." (PT15)*

**Has your school experience course/practicum given you the chance to put what you learned related to language testing and assessment into practice? If so, what are these practices?**

Most of the preservice English language teachers stated that they did not put what they have learned in terms of the theory of LTA into practice in this

course. Two of them also added that they did nothing related to testing and assessment, but observed their mentor teachers in their practicum. Especially, one of them stated:

*"I cannot say that we did something related to testing and assessment in our practice teaching because the only thing we did was to sit and observe how our mentor teacher lectured and tested her students, so we did not any practice on testing and assessment in real classrooms." (PT6)*

On the other hand, there are some preservice English language teachers who stated that they had the opportunity to put what they learnt with regard to the theory of LTA into practice in their practicum. What was common in the utterances of these preservice teachers was that they had a limited opportunity to apply their knowledge into the real classroom context. Some utterances of the preservice English language teachers who stated that they applied the theory of testing and assessment during their practicum were given below:

*"I think that I applied what I know about language testing and assessment during my practicum; however, this is too limited because I just prepared a short oral quiz for 5th grade students once and a written quiz including multiple choice questions once. I think that this is too limited. I wish we had been engaged in assessment activities more." (PT7)*

*"Although it is restricted, I applied what I learnt about testing and assessment in the real classroom. Once, my mentor teacher asked me to evaluate and give grades to the pop-quiz of students. Although I know the correct answers of the questions in the quiz, it was really hard for me to evaluate them accurately because I did not have such an experience before. Also, I tried to be fair in giving grades, which is another challenge for me. At that moment, I realized that I need more practice on evaluation and giving grades. What I also realized was that testing and assessment is not just limited to preparing and applying exams, but includes the following procedures like scoring, grading and even giving feedback." (PT21)*

**Do you think that the practice related to the language testing and assessment should be realized in school experience course? Why or why not?**

All of the participants mentioned that the practice of LTA should be realized in school experience course/ practicum. When their responses to the reason why they think so were analyzed, it was noticed that there were some common points that the participants had touched on. These common points were given below:

- The importance of gaining experience in the practice of language testing and assessment
- The relationship between the theory and the practice of language testing and assessment
- Its importance for their future profession and practices

*Figure 4.* Comments of participants.

In this regard, the following comments were made:

*"As assessment is one of the most crucial dimensions of language teaching and learning and we will deal with it most of the time, we should be given chance to put what we learnt as theory into practice. Otherwise, it would not be logical and complete. It is true that we need to know theories of assessment; however, having just this knowledge does not make sense and ensure accurate assessment practices. Shortly, we should be aware of the fact that in our real teaching contexts we will deal with testing and assessment most of the time."* (PT23)

*"Without the theory of testing and assessment, we cannot talk about effective and reliable assessment practices. Likewise, without the practice of testing and assessment, the theory of assessment would be meaningless. In this regard, I think that both the theory and practice of assessment are inseparable and they complement each other. Therefore, both of them should be integrated into school experience course at the same time so that it can contribute to us as teacher candidates."* (PT6)

*"... As knowing the theory is not enough as in every theoretical base like methodology or approaches, we should gain experience in practice of*

*testing and assessment. Also, we need to realize our weaknesses and strengths in language testing and assessment as language teachers. Accordingly, we can develop ourselves in our profession." (PT22)*

In addition to these, one student further suggested a way to integrate the practical components of LTA into the school experience course:

*"We could be graded in accordance with the tests or exams that we will prepare as well as lesson plans we prepared. All the time we were evaluated based on our lesson plans and our classroom management, and testing and assessment is ignored. Although we aimed to measure whether students have understood the topic that we lectured them during our teaching practices, we were not asked to prepare a test or any other assessment. Therefore, preparing a test or other forms of assessment for real students can be part of our evaluation in this course and we can get feedbacks on the assessment tools we prepared or adapted. If we practice this, I think that it will contribute to us a lot." (PT17)*

**Did you have a chance to reflect on issues related to LTA in school experience course? If so, what are they?**

Most of the participants stated that their school experience course did not enable them to reflect on issues of language testing and assessment. Especially, PT 19 made the following comment:

*"I don't remember any time that we reflected on something related to assessment and testing. What's more, I was not even aware of its importance until you asked this question to us because we have a tendency to ignore testing and assessment despite its role in our education system."*

On the other hand, there are few preservice English language teachers who mentioned that they found the chance to reflect on issues regarding LTA during their school experience course. What was common in the utterances of these preservice teachers was that one of their observation tasks was about observing their mentor teacher's way of assessing. To illustrate, PT 20 made the following comment:

*"One of our tasks was about observing our mentor's assessment of students. To complete this task, I observed the informal assessment that she used while students were completing a task as a group. After this observation task, our instructor of the school experience course asked us to discuss it and I had a chance to state and reflect on what I observed and thought. Especially, I remember that we discussed whether the teacher should have used a rubric for such informal assessment or not. It was a really fruitful session for me."*

In addition to this, some of them stated that in their school experience course they had a chance to reflect on the LTA system that was applied in the classrooms that they visited during their practicum. Especially, PT 16 stated:

*"In our discussions in school experience course, we discussed language testing and assessment system that we observed during our practicum. Especially, we stated that most of the language assessment were formal and there was a focus on testing grammar and assessment of other language skills was ignored. Considering all of these, we concluded that assessment was generally seen as applying multiple choice tests and giving grades."*

**Which topics related to language testing and assessment should be covered in school experience course, and in which ways do these help you?**

All of the participants stated some topics regarding the language testing and assessment, and the topics listed by all participants are given below:

| *Topics* |
| --- |
| • Assessment types |
| • Testing young learners |
| • Validity |
| • Reliability |

*Figure 5.* Topics listed by all participants.

When they were asked to state how these topics would help them as language teachers and their responses were analyzed, it was realized that there

was a common point that the participants had stated. The common point was these topics' being helpful for their future practices and professions especially in terms of preparing, adapting, analyzing, administering and evaluating tests and other types of assessment in accordance with their contexts. In this regard, the following comments were made:

*"I think we should be taught how to assess young learners particularly because most of the things we learn about assessment may not be valid for assessing young learners. Therefore, we should be taught some special ways of assessing young students so that we adjust our assessment in accordance with their levels, interests and ages." (PT22)*

*"Validity is one of the most problematic and ignored areas, and it is not just enough to cover it in testing and assessment course. Therefore, it should be covered in school experience course, too. One of the most important reasons for this is that it will enable us to prepare valid tests that can assess what we aim to assess in the future. Another reason is that it will enable us to analyze the existing tests in terms of their validity and to decide whether we can use these tests in our contexts or not. Especially, we should discuss and reflect on face, construct, content validity of tests in this course." (PT3)*

*"If we want to become effective language teachers, we should be effective assessors, as well. Reliability is one of the most crucial aspects of effective testing and to become effective assessors, we should know whether a test is reliable or not. Therefore, it should be covered in this course. As for how it will be helpful, there are many ways. For instance, it will help us to analyze existing tests in terms of reliability and determine whether we will use it or not. Also, after using these tests, we can look at its reliability by using different methods and reflect on students' scores in this exam accordingly. Maybe the low scores result from the test itself or the test administration environment. Besides, it will help us to get more precise and clear results from the tests, and provide beneficial washback for our*

*students. Shortly, we should know it to great extent and school experience course can provide this opportunity to us." (PT12)*

In addition to these topics, one of the participants also mentioned that testing grammar and vocabulary should be covered in school experience course. As for how it will be helpful, he made the following comment:

*"I think that especially testing grammar and vocabulary should be covered in this course. No matter what we claim the contrary, we will test our students' grammar and vocabulary knowledge most of the time. In this regard, knowing how to test vocabulary and grammar and various ways of doing so will help us to assess better and more accurately. Just knowing and using multiple choice exam or fill in the blank types to assess their language knowledge will not serve for our aim because it will be too limited. Therefore, we should cover this topic in detail for better and accurate assessment." (PT17)*

Moreover, two of the preservice English language teachers specified testing pronunciation to be covered in school experience course. When they were asked to indicate how this topic would help them, they stated that it would help them to design and prepare pronunciation tests which are appropriate for students' levels, interests and age groups. Especially, PT 25 made the following comment:

*"Pronunciation is an essential aspect of foreign language learning. Similarly, testing our students' pronunciation is essential for us as language teachers. In this regard, knowing various ways of testing pronunciation depending on the classroom context, namely students' levels, interests and age groups, enables me to test their pronunciation more comprehensively and accurately. For example, if my students are adults and their levels are A2, I should design and prepare pronunciation tests accordingly, which would be different from the test of pronunciation which was prepared for young learners."*

**Did taking ELTE course help you during your school experience/ practicum? If so, in which ways it was useful?**

Most of the participants stated that taking ELTE course did not help them during their school experience course or practicum. Two of them also added that they did not see any relation or benefit of ELTE course in their school experience course. Especially, PT21 stated:

*"I do not think that ELTE course helped me in my practicum or school experience course because I did not see any relation, implementation or advantage of this course in my school experience course. If I had seen any connection between the two, I could have said that it was useful."*

On the other hand, there are some preservice English language teachers who stated that ELTE course helped them during their school experience course or practicum. When their responses regarding how it was useful for them are analyzed, what is common in the utterances of these participants is that ELTE course both helped them in the completion of one of their observation tasks and made them aware of and reflect on important concepts related to testing and assessment during their practicum experience. Some utterances of the preservice English language teachers who stated that taking ELTE course was useful for their practicum and school experience course were given below:

"One of our observation tasks was about observing our mentor teacher's way  of assessment and I could complete it thanks to the knowledge I obtained in ELTE course. Without this course, it would be hard to  evaluate and reflect on the way assessment was implemented by my mentor teacher. Shortly, without the contributions of this course, I could not be aware of and critically evaluate some practices of my mentor teacher." (PT20)

"Yes, I gained a critical perspective on evaluating the written exams and pop-  quizzes of the students through the things I learnt in ELTE course. Even  though I was not allowed to grade them in my practicum, I was able to evaluate the reliability, practicality and many other elements of these exams. In this regard, it is possible for me to say that this course contributed to my practicum experience and professional development as a teacher candidate." (PT14)

"Taking this course made me realize that we have to consider lots of things while we are assessing students like our objective, the appropriateness

and reliability of the test. Also, I was not aware of many task types to assess language skills and after being aware of them, I observed them in my practicum. Also, once we talked about an issue in terms of assessment and testing system used in Turkey during school experience course, I remembered our discussions in ELTE course, which was very fruitful. Building on these discussions and expressing my own ideas on such issues really contributed to me and broadened my perspective." (PT3)

Besides, one preservice English language teacher narrated one of her stories regarding this issue:

*"In one of my observations, my mentor teacher decided to give a test for students to check their progress. When I paid attention to the items of the test, I found some problematic items. For example, some multiple choice items were problematic in that they included ungrammatical options. As I learned that it should be avoided as it would make students exposed to inaccurate language input, I immediately realized it. After the lesson, I wanted to share this with my mentor teacher and when I said this to him, he said that he did not know such a rule before and from that moment he would pay attention to this. Without this course, I would not know and reflect on such issues, too." (PT16)*

As seen in the comments above, the preservice English language teachers are aware of the importance and contributions of ELTE course to their practicum experience and school experience course. Furthermore, their responses reveal that they are aware of the critical contributions of this course to their professional development and future practices as language teachers.

**Chapter 5**

**Discussion, Conclusion and Suggestions**

**Discussion of the Results**

The first research question aimed at investigating the preservice English language teachers' LAL in terms of the training they received in four components of LTA.

Firstly, it aimed at finding out the level of training preservice language teachers think they received in *Classroom-focused LTA*. In order to answer this question, 18 items were included in part A of the first section in the questionnaire. When the overall mean score of these 18 items was calculated, it was found to be 1.50. It can be concluded from the results for these 18 items that in general the preservice English language teachers find the training received in this domain of LTA inadequate. Especially in "preparing language aptitude tests" component of this domain, the majority of the preservice English language teachers (39.6% and 28.7%, respectively) reported that they received little or none training. Furthermore, in some areas of LTA covered in this domain the majority of the preservice English language teachers find the training they received quite little: "preparing placement tests", "using European Language Portfolio", "preparing proficiency tests", and "preparing diagnostic tests". However, it is crucial to note that in some components of this domain nearly half of the preservice English language teachers reported that the training they received was adequate: "adapting ready-made tests for the needs of students", "stages of language test construction", "using ready-made tests from textbook packages or from other sources" and "preparing classroom tests".

Considering that LTA activities in the questionnaire constitute a vital part of preservice language teachers' assessment practices, it can be asserted that preservice language teachers' classroom-focused LTA practices, such as preparing language aptitude and placement tests, using European Language Portfolio and preparing proficiency and diagnostic tests, are underdeveloped. This finding is actually consistent with the results of Vogt and Tsagari's study (2014), in which they investigated the current level of LTA literacy of foreign language teachers and their training needs in three domains of this field and one of these

domains was classroom-focused LTA. Their study revealed that in the domain of classroom-focused LTA, most of the teachers received no or little training, especially in the alternative forms of assessment like portfolio and European Language Portfolio. Based on this finding, they pointed out that foreign language teachers' LAL was underdeveloped in this domain. As reported by Vogt and Tsagari (2014), the finding of this study also points to how little developed preservice English language teachers' LAL was in this domain given that training in language assessment plays a crucial role in shaping preservice language teachers' grasp and development of LAL.

It also tried to investigate the level of training preservice English language teachers think they received in the domain of *Knowledge of Testing and Assessment.* To this end, 9 items were included in part A of the second section in the questionnaire. The responses to these items were analyzed based on descriptive statistics and the overall mean score of these 9 items was found to be 1.73. The present finding suggests that in general preservice English language teachers find the training that they received in this domain of LTA insufficient. Especially "computer-based testing" and "alternative assessment" components of this domain seemed to be much neglected LTA aspects because more than one fourth of the preservice English language teachers find the training they received in these components inadequate". As for the other components of this domain, which are "norm/criterion-referenced assessment", "discrete point/integrative testing" and "approaches to language testing like integrative, communicative and structuralist", more than one fourth of the preservice teachers reported that the training they received in these components was little. On the other hand, in some areas of LTA covered in this domain nearly half of the preservice English language teachers find the training they received sufficient: "objective/subjective testing", "direct/indirect testing", "informal/formal assessment" and "formative/summative assessment".

Based on these findings, it can be concluded that in the domain of knowledge of testing and assessment certain elements of preservice language teachers' LTA expertise are not developed on the ground that LTA components of this domain constitute an essential part of preservice language teachers' expertise. The results of the present study also suggest that preservice English language teachers' LAL in this domain is underdeveloped considering the vital role

of adequate level of training in LTA in promoting the development of LAL. This finding is in line with the previous finding of the present study in that preservice English language teachers' LAL in these two domains were found to be underdeveloped.

Also, it tried to find out the amount of training preservice English language teachers think they received in the domain of *Purposes of Testing*. For this purpose, 7 components of this domain were included in Part A of the third section in the questionnaire. When the responses to these items were analyzed, the overall mean score was found to be 1.48. Drawing upon this, it can be concluded that preservice English language teachers perceive the training that they received in this domain not to be sufficient. Especially in "awarding final certificates" component of this domain, the majority of the participants reported that they received no (36.6%) and little (37.6%) training. In the other components of this domain, such as "placing students onto programs, courses, etc" and "measuring general ability to learn a foreign language", more than one fourth of the preservice teachers find the training they received inadequate. However, in some areas of LTA covered in this domain nearly half of the preservice English language teachers find the training they received sufficient: "identifying what has been learned", "finding out what needs to be learned/taught", and "giving grades".

Considering the vital role of LTA components of this domain in their expertise it can be concluded that certain elements of preservice language teachers' LTA expertise in the domain *of purposes of testing* are not well-developed. This finding is in line with that of Vogt and Tsagari (2014) that report that in most areas of the domain of purposes of testing foreign language teachers received no training, especially in "giving grades", "placing students onto courses, programs" and "awarding final certificates". Accordingly, they concluded that fundamental elements of teachers' LTA literacy in this domain are not developed enough and suggested that these are most probably learned in the profession.

The results of the present study also offer vital evidence for how little developed preservice English language teachers' LAL is in this domain on the ground that adequate level of training in components of LTA is required for the development of LAL.

Lastly, it aimed at investigating the level of training preservice English language teachers think received in the domain of *Content and concepts of LTA*.

In order to answer this question, 22 items were included in part A of the last section in the questionnaire. When the overall mean score of these 22 items was calculated, it was found to be 1.70. It can be concluded from the results for these 22 items that in general the preservice English language teachers find the training received in this domain of LTA insufficient. Particularly "using statistics to study the quality of tests/assessment" component seemed to be a much neglected LTA aspect since the majority of the prospective teachers (43.6% and 40.6%, respectively) stated they received little or no training. In the other components of this domain, which are "testing pronunciation in English", "different test items/task types to test pronunciation in English" and "alternatives in assessment", nearly one fourth of the preservice teachers reported that they did not receive any training.

On the other hand, in some areas of LTA covered in this domain the majority of the preservice English language teachers find the training they received quite sufficient: "authenticity", "validity", and "reliability". Especially, the majority of the participants (39.6% and 25.7%, respectively) reported advanced or sufficient training in "authenticity". It is also crucial to note that the amount of training in some components of this domain was found to be slightly sufficient by nearly half of the prospective teachers, such as "different test items/task types to test reading in English", "different test items/ task types to test listening in English", "testing reading in English", "testing listening in English", "different test items/task types to test speaking in English".

Given that only certain components of preservice language teachers' LTA expertise in the domain of *content and concepts of LTA* are found to be developed, it is critical to highlight that in general preservice language teachers' LTA expertise in this domain of LTA is not well-developed. Accordingly, it can be concluded that LAL of these prospective teachers is underdeveloped as sufficient level of training and expertise in the components of LTA is fundamental for the development of LAL. This finding is actually parallel with that of Vogt and Tsagari (2014) in that foreign language teachers' LTA literacy in *content and concepts of LTA* area was reported to be underdeveloped. Besides, it was reported that in almost all components of this domain foreign language teachers received on average either no or little training with the exception of advanced training received in some components like receptive, productive skills and microlinguistic aspects reported by some teachers.

Overall, to answer the first research question of the present study preservice English language teachers' responses to these 56 items were analyzed. When the overall mean score of these 56 items was calculated, it was found to be 1.61. Drawing upon the results of analyses regarding the amount of training preservice English language teachers think they received in four areas of *LTA*, it can be concluded that they find the training that they received in these four domains inadequate.

It is critical to highlight that the level of training and expertise in the components of LTA in these four domains was found to be insufficient and underdeveloped. As mounting evidence on assessment literacy of preservice teachers reveals, adequate level of training in assessment in preservice teacher education programs fosters preservice teachers' understanding of testing and assessment and assessment literacy development (Mertler & Campbell, 2005; Popham, 2009; Siegel & Wissehr, 2011). In this sense, as sufficient level of training and expertise in the LTA components of these four domains is fundamental for preservice language teachers' LAL development, the present finding provides evidence that LAL of the preservice English language teachers taking part in this study was not well-developed.

Especially, it is important to note that of all four LTA components included in the questionnaire, the area that appears to be the least developed is *"purposes of testing"*, which was consistent with the finding of Vogt and Tsagari (2014) in that among the three areas of LTA *purposes of testing* was found to be least developed. All in all, it was really interesting to find how little developed preservice English language teachers' LAL was, which was consistent with the overall finding of Vogt and Tsagari (2014).

This main result corroborates findings from other studies that also investigate the amount of training and LAL of language teachers like Cheng, Rogers, and Hu (2004), DeLuca and Klinger, (2010), Inbar-Lourie and Levi (2015), Lam (2015), Mertler (1998, 1999, 2003), Plake (1993), Popham (2004), Razavipur, Riazi, and Rashidi (2011). For instance, in the study of Popham (2004), it was seen that teachers were found to regard themselves as not having received adequate training in their undergraduate programs and not having adequate skills in making assessment-related decisions and administering assessment. Drawing upon these findings, the lack of adequate training in assessment in teacher

education programs was reported as a barrier to teachers' development of assessment literacy, which is in good agreement with the results of the present study.

On the other hand, in the study of Hasselgreen, Carlsen, and Helness (2004) that also investigated the training that teachers received and needed in three areas of LTA, it was seen that teachers did not receive any formal training in these three areas of LTA and they carried out most of assessment-related activities without any prior formal training. In this regard, the result of Hasselgreen, Carlsen, and Helness (2004) differs considerably from that of the present study since preservice language teachers taking part in the present study were found to have received training in areas of LTA though it is not sufficient enough.

The second research question tried to identify preservice English language teachers' LAL in terms of the further training that they think they need in four components of LTA.

It tried to investigate the level of further training preservice language teachers think they need in the domain of *Classroom-focused LTA.* For this purpose, 18 items which were also in part A of the first section were included in part B of the first section. When the responses to these items were analyzed, the overall mean score was found to be 1.19. The present finding suggests that in general preservice English language teachers perceive a need for further basic training in this domain of LTA. In all components of this domain, the majority of the preservice English language teachers find the amount of further training they need basic.

Especially, the majority of the preservice English language teachers (44.6%, 45.5%, and 46.5%, respectively) reported a need for basic training for "preparing language aptitude tests" and "using European Language Portfolio" as well as for the "preparing placement tests" since equally as many (39.6%, 38.6%, and 42.6%, respectively) received little training in these areas.

It is crucial to note that in some components of this domain nearly half of the participants reported that the training they received was adequate: "adapting ready-made tests for the needs of students" (49.5%), "stages of language test construction" (45.5%), "using ready-made tests from textbook packages or from other sources" (49.5%). As a consequence, more than one fourth of the

participants (27.7%, 27.7% and 30.7%, respectively) did not wish to receive further training in these LTA aspects. However, the picture is somewhat different with "preparing classroom tests", where even though more than half (57.4%) of the participants reported sufficient training, equally as many participants (48.5%) expressed a need for more advanced training in this LTA aspect.

Overall, the training needs of participants in the domain of Classroom-focused LTA established in the present study corroborate the findings of the study by Vogt and Tsagari (2014) to a certain extent. For instance, in both studies a need for training is expressed for all aspects of this domain of LTA. More specifically, the participants in general expressed a need for basic or more advanced training in almost all aspects of this domain of LTA covered in the questionnaire, a finding similar to that of Vogt and Tsagari. However, there are some aspects that are perceived as "less urgent" or "more urgent" than others. Unlike the finding in Vogt and Tsagari's study, for "using ready-made tests" more training need is expressed by the preservice language teachers in the present study. In other areas like in "ELP or Portfolio" and "preparing classroom tests", preservice language teachers expressed less training need compared to teachers in Vogt and Tsagari's study. All in all, given that in general the preservice English language teachers find the training they received in this domain of LTA inadequate, it was surprising to find that the need for further training expressed by them is basic. A possible explanation for this result may be the lack of awareness of the importance of LTA on their future profession and practices as language teachers.

It also aimed at investigating the level of further training participants think they need in the domain of *Knowledge of Testing and Assessment.* To this end, 9 items which were also in part A of the second section were included in part B of the second section in the questionnaire. The responses to these items were analyzed based on descriptive statistics and the overall mean score of these 9 items was found to be 1.04. Drawing upon this, it can be concluded that in general preservice English language teachers perceive a need for further basic training in this domain of LTA.

Especially, the majority of the preservice English language teachers (42.6%) reported a need for basic training for "computer-based testing" since the majority of them (33.7% and 28.7%, respectively) received little or no training in

this area. As for "alternative assessment", the majority of them (29.7% and 17.8%, respectively) received little or no training in this area. As a result, most of them wished for more training, both on basic (49.5%) and advanced (33.7%) training levels. On the other hand, it is important to highlight that in some components of this domain nearly half of the preservice English language teachers reported that the training they received was adequate: "objective / subjective testing" (49.5%), "formative / summative assessment" (49.5%), "direct / indirect testing" (48.5%), and "informal / formal assessment" (48.5%). As a consequence, about one third of the participants (36.6%, 31.7%, 28.7%, and 32.7%, respectively) did not wish to receive further training in these LTA aspects.

In parallel with the previous finding of the present study, in which a basic training need is expressed for *Classroom-focused LTA*, a need for basic training is clearly expressed for *Knowledge of Testing and Assessment* domain of LTA. The majority of the participants did not express a pronounced need for advanced training in this domain of LTA even though equally as many of them reported lack of adequate training in this domain.

Also, it tried to find out the amount of further training preservice English language teachers think they need in the domain of *Purposes of Testing*. For this purpose, 7 items which were also in part A of the third section were included in part B of the third section in the questionnaire. When the responses to these items were analyzed, the overall mean score was found to be 1.08. The present finding suggests that in general preservice English language teachers perceive a need for further basic training in this domain of LTA.

Especially, "placing students onto programs, courses, etc.", "testing competence in a language" and "awarding final certificates (from school, program; local, regional / national level)"were three areas where 59.4%, 42.6% and 42.6%, respectively, of the preservice teachers expressed a pronounced need for basic training.

There are certain LTA components that even though participants received substantial training, they still voiced a need for more. For instance, for "identifying what has been learned", while most participants (46.5% and 20.8%, respectively) reported having received sufficient or advanced training, equally many of them wished for basic (36.6%) and advanced (31.7%) training in this aspect. Also, in "finding out what needs to be learned/ taught",even though most participants

reported having received adequate (49.5%) or advanced (15.8%) training, the majority of them wished for basic (43.6%) and advanced (28.7%) training in this area. As for "giving grades", nearly half (48.5%) of them reported sufficient training; however, they wished for further training both in basic (46.5%) and advanced (24.8%) levels.

Overall, the training needs of participants in the domain of *Purposes of Testing* established in the present study corroborate the findings of the study by Vogt and Tsagari (2014) to a certain extent. In both studies, a need for training is clearly expressed for all components of this domain. However, in Vogt and Tsagari's study, the participants tended to wish for more advanced training in these components. More specifically, most participants in the present study expressed a moderate need for training in the given areas of this domain, asking for basic training.

Lastly, it aimed at investigating the amount of further training preservice English language teachers think they need in the domain of *Content and concepts of LTA*. In order to answer this question, 22 items which were also in part A of the last section were included in part B of the last section in the questionnaire. When the overall mean score of these 22 items was calculated, it was found to be 1.15. Drawing upon this, it can be concluded that preservice English language teachers perceive a need for further basic training in this field of LTA.

Especially, the majority of the participants (35.6%, 40.6%, and 40.6%, respectively) reported a need for basic training for "using statistics to study the quality of tests/assessment" and "testing pronunciation in English" as well as for the "alternatives in assessment" since equally as many (43.6%, 36.6%, and 42.6%, respectively) received little training in these areas.

On the other hand, there are certain LTA components that even though preservice English language teachers received substantial training, they still voiced a need for more. For instance, for "reliability', while most participants (34.7% and 33.7%, respectively) reported having received sufficient or advanced training, equally many of them wished for basic (32.7%) and advanced (35.6%) training in this aspect. Also, in "validity", even though most participants reported having received adequate (36.6%) or advanced (33.7%) training, the majority of them wished for basic (32.7%) and advanced (38.6%) training in this area.

Besides, it is crucial to note that in some components of this domain nearly half of the preservice English language teachers reported that the training they received was adequate: "different test items/task types to test reading in English" (49.5%), "different test items/ task types to test listening in English" (47.5%), "testing reading in English" (45,5%), "testing listening in English" (43.6%), "different test items/task types to test speaking in English" (42.6%). As a consequence, nearly one fourth of them (26.7%, 24.8%, 28.7%, 30.7%, and 23.8%, respectively) did not wish to receive further training in these LTA aspects.

Drawing upon these, it can be concluded that the training needs of preservice English language teachers in the domain of *Content and concepts of LTA* established in the present study are in good agreement with the findings of the study by Vogt and Tsagari (2014) to a certain extent. In both studies, a need for training is clearly expressed for all aspects of this domain. However, most participants in the present study expressed a moderate need for training in the given areas of this domain, asking for basic training whereas those in Vogt and Tsagari's study tended to wish for more advanced training in these components. All in all, given that in general the preservice English language teachers find the training they received in this domain of LTA inadequate, it was surprising to find that they did not express a pronounced need for advanced training, which may result due to the fact that they may not be aware of the importance of having adequate training in LTA for their future assessment practices.

Overall, to answer the second research question of the present study, preservice English language teachers' responses to these 56 items were analyzed. When the overall mean score of these 56 items was calculated, it was found to be 1.14. Drawing upon the results of analyses regarding the amount of further training preservice English language teachers think they need in four areas of *LTA*, it can be concluded that they perceive the amount of training that they need further in these four domains to be basic. Although in general participants find the training they received in four domains of LTA inadequate, they did not express a pronounced need for advanced training in these domains.

Especially, it is important to note that of all four LTA domains included in the questionnaire, the domain in which need for further training is expressed less is *Knowledge of Testing and Assessment*, indicating that preservice language teachers perceive the level of training they need further in this domain to be less

than other three domains. It is critical to highlight that these LTA aspects that form an essential part of preservice language teachers' expertise, practice and LAL do not feature among the prioritized areas for further training.

All in all, preservice English language teachers expressed a need to receive training across the range of LTA aspects given in the study with varying priorities.In other words, drawing upon the findings, it can be concluded that there is a need for training across the board, despite different priorities for components of LTA or the extent of training expressed. In this regard, the results of Vogt & Tsagari (2014) and Hasselgreen et al. (2004) were corroborated. More specifically, in three domains of LTA included in the study of Vogt & Tsagari (2014) and Hasselgreen et al. (2004), foreign language teachers reported a need for further training in given areas of LTA with varying priorities, too. However, there are different priorities for the given aspects of LTA and the amount of training desired between these two studies and the present study.

The third research question aimed at investigating preservice language teachers' opinions and evaluations of the school experience course with regard to the English Language Testing and Evaluation (ELTE) course. More specifically, it aimed at exploring participants' opinions on whether and how the theory and practice of language assessment was covered besides whether and how they should be covered in school experience course. It also investigated participants' opinions on the usefulness of English Language Testing and Evaluation (ELTE) course on their school experience course in depth. To this end, semi-structured interviews were held with 25 randomly chosen participants. In the first interview question, the preservice English language teachers were asked whether their school experience course covered anything related to the theory of LTA, and if this were relevant, they were also required to list these topics covered in school experience course. All of them reported that their school experience course did not include anything related to the theory of LTA; as a consequence, the question related to the list of topics was not addressed. All in all, the analysis of the 25 preservice language teacher interviews showed that nothing in regard to the theory of LTA was covered in school experience course.

In the second interview question, the preservice English language teachers were first asked whether topics related to the theory of LTA should be covered in

school experience course, and in accordance with their answers, they were required to state the reason why they think in that way. All of them stated that the theory of LTA should be covered in school experience course due to its importance for their future profession, and the relationship between LTA and language teaching and learning processes. Specifically, the relation between LTA and language teaching and learning is emphasized as a reason, which may result because of the fact that participants as prospective teachers are aware of the fact that when they start teaching, they will be engaged in assessment-related activities most of the time. In other words, it can be concluded that they are aware of the fact that as foreign language teachers they will be dealing with assessment-related procedures like designing, developing and evaluating tests as part of their profession.

Moreover, some of them mentioned that another reason why the theoretical bases of LTA should be included is that they will have a chance to reflect on real assessment contexts within the scope of theoretical bases. Especially, it was indicated that discussing real assessment contexts with reference to the theory of LTA like they discussed real teaching contexts in school experience course would be meaningful and contribute to their theoretical understanding of LTA. Furthermore, one of the participants tried to explain the reason through making reference to the necessity of covering methodological knowledge in school experience course. She stated that without language teaching methodology, teachers would not be regarded as competent; likewise, without the theory of LTA, they cannot be competent since it is one crucial component of being an effective teacher. All in all, they pronounced the need for covering theory of LTA in school experience course for various reasons. Considering the amount of training received and needed further in LTA domains, and these reasons stated by the participants, it is not surprising to find that they state the necessity of covering theory of LTA in connection with real assessment contexts in school experience course.

In the third interview question, the preservice English language teachers were asked whether their school experience course/practicum gave them the opportunity to put what they learned in terms of LTA into practice, and depending on their answers, they were asked to name these practices. The analysis of the interviews revealed that most of them reported that they did not put their

theoretical knowledge into practice in that course. Besides, some of them stated that they did nothing related to LTA, but observed their mentor teachers during their practicum. On the other hand, some of them mentioned that they had a limited opportunity to practice their theoretical knowledge through preparing a short oral quiz and giving grades to pop-quiz of students. Besides, they emphasized a need for more practice in LTA like giving grades and evaluating students. All in all, the majority of the preservice English language teachers stated that they did not have the chance to put their theoretical knowledge into practice in school experience course/practicum; however, there were some who mentioned that they had a limited chance to apply their knowledge of LTA in school experience course and real classroom context.

In the fourth interview question, the prospective English language teachers were required to answer whether the practice in LTA should be realized in school experience course/practicum, and in accordance with their responses, they were asked to mention the reasons behind this idea. All of them stated that the practice of LTA should be realized in school experience course/practicum due to the importance of having experience in the practice of LTA, the relationship between the theory and practice of LTA, and its importance for their future practices and profession. Especially, the relationship between the theory and practice of LTA is highlighted in that if either one of these is lack, effective and accurate assessment cannot be mentioned, indicating that they are inseparable from each other. Furthermore, they indicated that to develop themselves in teaching profession, they need to develop themselves in LTA by realizing their weaknesses and strengths in assessment practice. Accordingly, it can be concluded that one of the reasons why they put emphasis on adequate practice in LTA is that they regard it as an essential part of their professional development. All in all, they expressed the necessity for practicing LTA aspects in school experience course/practicum for some reasons.

In the fifth interview question, the respondents were first asked whether they had the chance to reflect on issues related to LTA in their school experience course, then depending on their responses they were asked to mention the issues on which some reflections were made during the course. Most respondents stated that this course did not give them the chance to reflect on issues of LTA. On the other hand, there were some respondents who stated that they reflected on some

issues in regard to LTA. One of the observation tasks assigned to them in the course was observing their mentor teacher's way of assessment. While mentioning about this task in which they observed and discussed this issue during the lecture, they also highlighted how fruitful it was for them. Besides, some stated that they reflected on the LTA system that was used in the classrooms they visited during their practicum. She stated that most of the assessment was formal and included multiple choice tests as well as ignoring assessment of language skills, which indicates that this prospective teacher has a certain level of knowledge in and awareness of LTA, which enables her to make such a reflection.

In the sixth interview question, they were asked to state which topics of LTA should be covered in school experience course and how these topics would help them. All of them mentioned these topics in common: assessment types, testing young learners, validity and reliability. As for the reasons why these should be included in the course, they stated that these topics would be helpful for their future practices and profession specifically in adapting, analyzing, administering and evaluating tests and other types of assessment according to their contexts. Specifically, "testing young learners" is emphasized on the ground that the things that they learnt about assessment may not be valid for assessing young learners; in other words, they stated that they would need special techniques of assessing young learners rather than general techniques. This may result due to the fact that most of them did their practicum in pre-school and primary school settings where they had a chance to observe the characteristics of these young learners and to reflect on the issue of assessment.

As for the validity, which is among the keystones of any accurate and effective assessment, the prospective teachers mentioned that it was not just enough to cover it in testing and assessment course; therefore, it should be covered in school experience course, too. As to the rationale behind it, they stated that it would help them in preparing valid tests in terms of assessing what they aimed to assess and in analyzing existing tests in terms of validity and deciding whether to use these tests in their teaching contexts or not. Likewise, reliability was mentioned as one of the topics that should be covered not only in testing and assessment course but also in school experience course.  As to the rationale behind it, they mentioned that it would be helpful in analyzing existing tests in regard to reliability and determining whether to use it or not. Besides, it would help

them to reflect on students' scores in the exam in that low scores may result from the test itself or test administration environment, and to provide beneficial washback for students. Drawing upon these responses, it can be suggested that prospective teachers are aware of the importance of these two crucial aspects of assessment for their future assessment practices. All in all, it was not surprising to find that validity and reliability were listed among the topics that should be covered in school experience course as well as in testing and assessment course since both of them were among the keystones of any effective assessment.

In addition to these topics, one of the preservice English language teachers also mentioned that testing grammar and vocabulary should be covered in school experience course. As for the rationale, she stated that no matter what was claimed, they would assess students' grammar and vocabulary knowledge most of the time; therefore, having adequate knowledge in this aspect and in various ways of assessing these components is essential for them. Accordingly, she expressed a need for covering it in detail in school experience course. On the other hand, some mentioned that school experience course should cover testing pronunciation in that it would be helpful for them in the process of designing and preparing pronunciation tests. Especially, they stated that having adequate knowledge in designing and adapting pronunciation tests for different age groups, interests and proficiency levels is essential for language teachers. Therefore, they expressed a need for covering this aspect of LTA in a more detailed way in school experience course.

In the last interview question, the prospective English language teachers were first asked to answer whether taking English Language Testing and Evaluation (ELTE) course helped them during their school experience/ practicum, then depending on their responses, they were asked to mention in which ways it was useful for them. Most prospective English language teachers stated that ELTE course did not help them in their school experience course or practicum. In addition to this, some of the participants stated that they did not even see any connection between ELTE course and school experience course/practicum. Especially, one of them mentioned that if she had seen any connection between these two courses, she could have stated that it was useful. This finding is quite surprising and suggests that when LTA is not included and emphasized during school experience course or practicum, and the connection between the two is not

revealed explicitly, it is possible that prospective language teachers will fail to see any connection between them. As put forward by Alkharusi, Kazem and Al-Musawai (2011), this finding points to how crucial the connection between assessment course and field-based experience like practicum is considering the maximization of the development of LAL of prospective teachers. In this respect, as suggested by Lam (2015), this finding also highlights that when LAL is not included and evaluated during practicum, it is highly possible that preservice teachers will neglect its effect on teaching and learning processes. With regard to this, incorporating LAL into teaching practicum may increase teacher candidates' awareness and foster competence in LTA. Overall, the findings demonstrate that preservice English language teachers are aware of the contributions and importance of ELTE course to practicum and especially to their professional development and future practices.

**Summary of the Study**

This study was performed to investigate the LAL of preservice English as a foreign language teachers. To this end, the extent to which they have received training in both theoretical and practical terms of assessment in teacher education programs, and they perceive a need for training in language assessment were examined. Besides investigating the status quo of preservice language teachers' assessment literacy, it also targeted examining the extent to which assessment knowledge was put into practice in teaching practicum, and theory and practice of assessment was covered in school experience course and practicum. Furthermore, this research study attempted to find out whether and how they should be covered in school experience course and practicum as well as striving to see the usefulness of English Language Testing and Evaluation (ELTE) course on their school experience course and practicum experience. The results obtained could offer new insights for decision-makers, teacher trainers in teacher education programs in Turkey and could help them to rethink the significance of assessment in language education and the crucial need for training to develop LAL as well as encouraging them to take some measures to improve the current situation. It could also contribute to the development of policy of assessment training in teacher education programs.

This research study was conducted in a mixed methods research design in that both qualitative (interview) and quantitative (questionnaire) research methods were employed. It was carried out with 4th grade students who were studying in the ELT department of a state university. 101 students participated in the quantitative part of the study whereas 25 students were involved in the qualitative part of the study. For the quantitative part of the study, Vogt and Tsagari's (2014) Teachers' Questionnaire was adapted to form the 112-item instrument of the study. 4-point Likert scale was used in Part A whereas 3-point Likert scale was employed in Part B of the instrument. The Likert scale data gathered through the questionnaire was analyzed through the means, frequencies and percentages obtained for each item through SPSS Statistics 20.0. As for the qualitative part of the study, semi-structured interviews were employed. There were seven open-ended questions in the interviews and these were produced by the researcher. The qualitative data gathered through semi-structured interviews were analyzed through selective coding.

Main findings of the study are presented below:

1- The first research question aimed to investigate preservice language teachers' language assessment literacy in terms of the training that they have received in four domains of LTA which are C*lassroom-focused LTA, Knowledge of Testing and Assessment, Purposes of Testing, and Content and concepts of LTA*. It was found that the participants perceived the training that they received in these domains not to be sufficient. Especially, they reported the highest mean value for *Knowledge of Testing and Assessment*, followed by *Content and concepts of LTA*, *Classroom-focused LTA* and *Purposes of Testing*.

2- The second research question explored preservice language teachers' language assessment literacy in terms of the further training that they think they need in four domains of LTA. It was seen that preservice language teachers perceived the amount of training that they need further in these domains to be basic. Particularly, they stated the highest mean value for *Classroom-focused LTA,* followed by *Content and concepts of LTA*, *Purposes of Testing and Knowledge of Testing and Assessment*.

3- The target of the last research question was to explore preservice language teachers' opinions and evaluations of the school experience course with

regard to the English Language Testing and Evaluation (ELTE) course. The findings indicated that preservice English language teachers were aware of the importance and contributions of ELTE course to their practicum experience and school experience course. Besides, their responses revealed that they were aware of the critical contributions of this course to their professional development and future practices as language teachers.

**Pedagogical and Methodological Implications**

In this part, finding-based pedagogical and methodological implications will be presented to provide new insight to decision-makers, language educators, teacher trainers in teacher education programs, curriculum developers, assessment professionals and administrators, and researchers.

**Pedagogical Implications.** The findings of this particular study could be used for the betterment of the teacher training program and/or for revising the courses offered in the department of English Language Teaching at Middle East Technical University.

First of all, since language teachers' competence in LTA is considered to be an essential component of second/foreign language education, and since teacher education programs are pivotal in providing preservice language teachers with greater language assessment literacy competency (DeLuca & Klinger, 2010), preservice language teachers' stating the training received in the domains of LTA as not to be sufficient points out a significant issue to consider. Therefore, the courses offered in this program can be revised with regard to their content and methodology. These courses can be strengthened to provide adequate level of training in sound assessment knowledge base and practices and to foster LAL of preservice language teachers. To this end, ELTE offered in the last year of the program could be revised in a way that it should not only address sound foundations in LTA but also new developments in the field of language assessment so that language teachers can keep up with the innovations and employ these developments in their future practices. Besides, the requirements of this course could be revised in that practice component could be added to the grading system of the course. Apart from taking quizzes, mid-terms and finals, teacher candidates could be required to engage in designing and constructing

language tests and assessments as well as reflecting on the process in order to gain experience in and have deeper insights into LTA process. Furthermore, some workshops on language assessment could be held and these workshops could include some components such as experienced teachers' talks on their experience in real teaching contexts including the difficulties they had in practice and their suggestions for future practices, LTA experts' suggestions and the opportunity for designing some kind of language tests and assessments under the guidance of LTA professionals. By doing so, prospective language teachers' awareness of LTA could be increased and they could have a chance to reflect deeply on real assessment practices.

Besides, since having greater LAL requires both sound assessment knowledge base and practices (Paterno, 2001), and since school experience course and practicum experience are significant in bridging the gap between the theory and the practice (Lam, 2015), prospective teachers' stating the amount of practice and experience in LTA in school experience course and practicum as inadequate indicates an important issue to take into consideration. Considering that teaching practicum and experience are useful in preparing teachers in testing and assessment in terms of enhancing prospective teachers' assessment knowledge, skills and attitudes (Alkharusi, Kazem, & Al-Musawai, 2011), teacher education programs should connect authentic field experience with assessment courses in order to enhance the development of LAL of teacher candidates. In this regard, as proposed by Lam (2015), when language assessment is not included and evaluated during teaching practicum and school experience course, it is very possible for teacher candidates to ignore its importance and role in teaching and learning. Therefore, LTA should be incorporated into teaching practicum so that teacher candidates' awareness could be increased and higher competence in LTA could be obtained.

Given that the amount of practice and experience in LTA is regarded as inadequate by preservice English language teachers, school experience course and practicum experience should provide more chances of experience. To this end, school experience course and practicum experience could be revised in a way that teacher candidates could have more chance to put what they have learned in terms of the theory of LTA into practice in real teaching contexts and to reflect deeply on issues related to LTA in school experience through some ways

like small-group discussions or conducting research as an individual or pairs. In this regard, the requirements of school experience course could be revised in that some observation tasks could be assigned to teacher candidates, which could provide them with deeper insights into the real classroom assessment practices and increase their awareness of issues related to authentic assessment practices. Besides, this course could incorporate some components such as watching video recordings of real classroom contexts in which students' language progress and proficiency is aimed to be assessed and evaluated, and examining some case studies.

As for the practicum experience, it could be also revised in a way that it could require preservice language teachers to design, construct and administer some kind of language tests and assessments in a real classroom context as well as grading and evaluating these. It may be also suggested that after having experience in test preparation and administration, post-conferences with the mentor teacher could be held for teacher candidates to get constructive feedback. Another suggestion could be that teacher candidates could be required to videotape their assessment practices in practicum to assess their LAL. As put forward by Lam (2015), requiring them to have reflective journals that consist of their beliefs and experience in assessment of students in terms of knowledge, skills and practices covered in assessment course could be considered as a recommendation. To this end, the schools which teacher candidates visit for their practice teaching should be carefully chosen and the mentor teachers in these schools should be informed of the requirements and objectives of this course. In other words, with some revisions and changes in the requirements of these courses, preservice language teachers could be provided with more opportunities to develop their assessment practices, which will contribute to the maximization of the development of LAL. All in all, the results of the present study can serve as a needs analysis, which outlines the aspects in which teacher candidates need more training, practice and development, and guide the preservice teacher education programs accordingly.

**Methodological Implications.** This study was performed using both quantitative and qualitative research design, and questionnaires and interviews were employed to collect the data. For gaining an understanding of the current

level of preservice language teachers in LAL and identifying their training needs in that area, questionnaires were used. However, observations could also be employed to collect data on the in-class competencies of teacher candidates on LTA and to compare those with the quantitative results found in this study. In this regard, the methodology pursued could be improved by employing observation method and different results could be obtained as well. Besides, this study adopted a cross-sectional design and represented a snapshot of reality. The results obtained in this study with regard to teacher candidates' perceptions on the training received and the need for further training in LTA as well as their opinions and evaluations of the school experience course with regard to the English Language Testing and Evaluation (ELTE) course may be different from those of a longitudinal study that focus on these aspects over time.

In addition, since the study has focused only on the Middle East Technical University preservice language teachers, the findings obtained may not be generalizable to other preservice English language teacher education programs in that teacher candidate profile in terms of the competence in LTA and their perceptions on the training received and the need for furher training in LTA may be different from those of teacher candidates in the other teacher education programs. In this regard, more preservice English language teachers studying in different teacher education programs in Turkey could have attended the study and this study could also have produced different findings.

Last but not least, such studies should be conducted at regular intervals by the preservice English language teacher education programs and other departments of the Faculties of Education, and by the higher Education Council with the aim of fostering LAL and competency of teacher candidates and improving teacher training programs for the sake of the development of the education system in the country.

## Suggestions for Further Research

Based on the limitations of the study, some suggestions that may shed light on further research studies are presented below:

- Firstly, this study included both quantitative and qualitative research design, and questionnaires and interviews were employed to gather the data. With

the aim of getting a grasp of the level of language assessment literacy of teacher candidates and determining their training needs in this field, quantitative research method was used. Supporting the findings of the quantitative data with observation method could give a better picture of competencies of teacher candidates on LTA.

- Secondly, a further research could be conducted with many more preservice English language teachers who have taken English Language Testing and Evaluation (ELTE) course in ELT Departments of different universities in Turkey. It would help to gain a better grasp of language assessment literacy of preservice English language teachers.

- Lastly, the setting of the study may be different since it was only conducted in one state university. In other state or private universities, the results may yield very different owing to some factors such as their profile with regard to the competence in LTA and their perceptions on the training received and the need for further training in LTA, and the content, methodology and instruction of courses like ELTE, school experience and practice teaching.

**Conclusion**

The results emerged in this study showed that preservice English language teachers who have taken ELTE course perceived the training that they received in four domains of LTA as not sufficient. With regard to these domains, *"Knowledge of Testing and Assessment"* was reported to be the domain in which teacher candidates reported to have received more training, and it was followed by *"Content and concepts of LTA"*, *"Classroom-focused LTA" and "Purposes of Testing"*. Even though the majority of teacher candidates reported lack of adequate training in these domains, equally as many of them did not report a pronounced need for advanced training in these domains of LTA. All in all, the amount of training that they need further in these domains was reported to be basic.

This study also concluded that preservice English language teachers did not have the chance to put what they learned in LTA field into practice during teaching practicum, and the theory and practice of assessment was not covered

much in school experience course. They reported the need for covering these in school experience course and putting their theoretical knowledge into practice in an authentic context. Their responses also revealed that they were aware of the significant contributions of ELTE, school experience and practice teaching courses to their future practices as language teachers and their professional development.

**References**

Aksit, N. (2007). Educational reform in Turkey. *International Journal of Educational Development, 27*(2), 129-137. doi: 10.1016/j.ijedudev.2006.07.011

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education, 39*(2), 113-123.

Anderson, D. L., Mathys, H., & Mills, A. (2014). Pre-Service Teachers' Assessment of 7th Grade Students' Social Studies Learning. *Journal of Curriculum and Teaching, 3*(1). doi:10.5430/jct.v3n1p104

Andrade, H., & Cizek, G. (Eds.) (2010). *Handbook of formative assessment.* New York: Routledge.

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self Assessment. *Assessment and Evaluation in Higher Education, 32* (2), 159-181.

Bachman, L. F., & Palmer, A. (1996). Language testing in practice. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2000). *Language testing in practice.* Oxford: Oxford University Press.

Bailey, K. M., & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236–256). Clevedon, UK: Multilingual Matters.

Bardes, B., & Denton, J. (2001). Using the Grading Process for Departmental and Program Assessment. Denver: American Association for Higher Education Conference.
(Western and Northern Canadian Protocol for Collaboration in Education,2006). Rethinking Classroom Assessment with Purpose in Mind.

Berry, R. (2008). *Assessment for learning.* Aberdeen, Hong Kong: Hong Kong University Press.

Black, P., & Wiliam, D. (1998). Inside the black box. *Phi Delta Kappan, 80*(2), 139-148.

Brindley, G. (2001a). Assessment. In Ronald Carter & David Nunan (Eds.), *The Cambridge Guide to teaching English to speakers of other languages* (pp. 137-143). Cambridge: Cambridge University Press.

Brindley, G. (2001b). Language assessment and professional development. In C.

Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies*. Cambridge: Cambridge University Press, pp. 126–136.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson/Longman, 2004.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and Classroom Practices.* White Plains, NY: Pearson Education.

Brown, J. D., & Bailey, K. M. & (2008). Language testing courses: What are they in 2007? *Language Testing*, *25*(3), 349-383.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*, 653- 675.

Bryman, A. (2008). *Social research methods*. Oxford University Press.

Boston, C. (2002). The concept of formative assessment (Report No. EDO-TM-02 06). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

(ERIC Document Reproduction Service No. ED470206)

Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *New visions in action: National assessment summit papers* (pp. 18-23). Ames: Iowa State University.

Campbell, C., & Evans, J. A. (2000). Investigation of Preservice Teachers' Classroom Assessment Practices During Student Teaching. *The Journal of Educational Research*,*93*(6), 350-355.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, 360-389.

Cheng, L., Watanabe, Y. J., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum.

Cirit, N. C. (2015). Assessing ELT Pre-Service Teachers via Web 2.0 Tools: Perceptions toward Traditional, Online and Alternative Assessment. *The Turkish Online Journal of Educational Technology*, *14*(3), 9-19.

Chen, P. (2005). Teacher candidates' assessment literacy. *Academic Exchange Quarterly*, 62, 62-66.

Clapham, C. (2000). Assessment And Testing. *Annual Review of Applied Linguistics, 20,*147-161. doi:10.1017/s0267190500200093

Cohen, A. D. (2001). Second language assessment. In M. Celce-Murcia (Ed.). *Teaching English as a second or foreign language, 515-534* (3rd ed). Boston, MA: Heinle&Heinle.

Conley, M. W. (2005). *Connecting standards and assessment through literacy*. Boston: Pearson Education, Inc.

Coombe, C. A., Davidson, P., O'Sullivan, B., & Stoynoff, S. (2012). *The Cambridge guide to second language assessment*. New York: Cambridge University Press.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Qualitative Sociology, 13*(1), 3-21.

Council of Europe (2000). *European language portfolio (ELP): Principles and guidelines*, Document DGIV/EDU/LANG, (2000) 33, Strasbourg, Council of Europe.

Council of Europe and European Union. (2001). How you can learn languages. (European year of languages guide): Council of Europe and European Union.

Council of Europe (2011). European Language Portfolio. Retrieved from http://www.coe.int/t/dg4/education/elp

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, California: Sage Publications, Inc.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications Ltd.

Darling-Hammond, L. (2006): Powerful teacher education: lesson for exemplary programs. San Francisco: Jossey-Bass.

Davin, K. J., & Heineke, A. J. (2016). Preparing Teachers for Language Assessment: A Practice-Based Approach. *TESOL Journal, 7*(4), 921-938.

Davison, C., & Leung, C. (Eds.) (2009). Teacher-based assessment. *TESOL Quarterly, Special Issue, 43*(3), 389–562.

DeLuca, C., & Bellara, A. (2013). The Current State of Assessment Education Aligning Policy, Standards, and Teacher Education Curriculum. *Journal of Teacher Education, 64*(4),356-372. http://dx.doi.org/10.1177/0022487113488144

Deluca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice,17*(4), 419-438. doi:10.1080/0969594x.2010.516643

Dhindsa, H., Omar, K., & Waldrip, B. (2007, August 1). Upper Secondary Bruneian Science Students' Perceptions of Assessment. *International Journal of Science Education, 29*(10),1281-1280.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Oxford: Oxford University Press.

Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York, NY: Routledge.

Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in SLA research? In A. Mackey & S. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 74-94). Malden, MA: Wiley-Blackwell.

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence.' *Language Testing, 21*(3), 259–283.

Elshawa, N. R., Heng, C. S., Abdullah, A. N., & Rashid, S. M. (2016). Teachers' Assessment Literacy and Washback Effect of Assessment. *International Journal of Applied Linguistics and English Literature IJALEL, 5*(4). doi: 10.7575/aiac.ijalel.v.5n.4p.135

Falsgraf, C. (2005). Why a national assessment summit? New visions in action. *National Assessment Summit*, p. 6. Meeting conducted in Alexandria, Va. http://www.nflrc.iastate.edu/nva/worddocuments/assessment_2005/pdf/ns p_introduction.pdf

Fan, Y.-C., Wang, T.-H., & Wang, K.-H. (2011). A Web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education, 57*(2),1727-1740.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113–132.

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An advance Resource book.* Routledge Taylor and Francis Group. London and New York.

Gall, M. D., Gall, P. J., & Borg, W. R. (2007). *Educational Research an Introduction.* New York: Pearson.

Garrison, C., & Ehringhaus, M. (n.d.). Formative and Summative Assessments in
     The Classroom. *Association for Middle Level Education.*

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced Middle School Science
     Teachers'    Assessment Literacy: Investigating Knowledge of Students'
     Conceptions in Genetics   and Ways to Shape Instruction*. Journal of
     Science Teacher Education, 23*(5), 531-557. doi:10.1007/s10972-012-9278-
     z

Greenberg, J., & Walsh, K. (2012). What teacher preparation programs teach
     about K- 12  assessment: A review. Washington, DC: National Council on
     Teacher Quality.

Gronlund, N. E. (1998). *Assessment of student achievement.* Boston: Allyn and
     Bacon.

Gronlund, N. E., & Waugh, C.K. (2009). *Assessment of Student Achievement.*
     Upper Saddle River, NJ: Pearson Education.

Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the
     assessment practices of a teacher of English as a foreign language in a
     Chinese secondary school? *Assessment in Education: Principles, Policy &
     Practice, 21*(3), 286–305.

Gurski, L. F. (2009). *Secondary teachers' assessment and grading practices in
     Inclusive classrooms* (Unpublished master's thesis). Thesis / Dissertation
     ETD.

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

Hatipoglu, C. (2010). Summative Evaluation of an English Language Testing
     and Evaluation Course for Future English Language Teachers in Turkey.
     *English Language Teacher Education and Development (ELTED) Journal,
     13*, 40-51.

Hatipoglu, C. (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.

Heaton, J. B. (1975). *Writing English Language Tests.* London: Longman.

Herrera, L., & Macías, D. F. (2015). A Call for Language Assessment Literacy in the Education and Development of English Language Teachers. *Colombian Applied Linguistics Journal, 17*(2), 302-312.

Howitt, C. (2007). Pre-service elementary teachers' perceptions of factors in an holistic methods course influencing their confidence in teaching science. *Research in Science Education, 37* (1), 41–58.

Hughes, A. (2003). *Testing for language teachers.* Cambridge: Cambridge University Press.

Hyland, K. (2010). *Second language writing.* New York: Cambridge University Press.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*, 328–402.

Inbar-Lourie, O. (2013a). The special issue on language assessment literacy. *Language Testing*, *30*(3), 301-307.

Inbar-Lourie, O. (2013b). Guest editorial to the special issue on language assessment literacy. *Language Testing*, *30*(3), 301–307.

Inbar-Lourie, O., & Levi, T. (2015). Implementing Formative Classroom Assessment Initiatives: What Language Assessment Literacy Knowledge is Required? Paper presented at the 37th Language Testing and Research Colloquium. Toronto/Canada,16 to 20 March 2015.

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, *27* (4), 555-584.

Johnson, R. B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches.* Boston, MA: Pearson Education, Inc.

Johnson, K., & Johnson, H. (2001). Encyclopaedic Dictionary of Applied Linguistics. Oxford: Blackwell.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.

Kahl, S. R., Hofman, P., & Bryant, S. (2013). *Assessment literacy standards and performance measures for teacher candidates and practicing teachers.* Prepared for the Council for the Accreditation of Educator Preparation. Dover, NH: Measured Progress.

Karp, G. G., & Woods, M. L. (2008). Preservice Teachers' Perceptions about Assessment and Its Implementation. *Journal of Teaching in Physical Education, 27*(3), 327-346. doi:10.1123/jtpe.27.3.327

Kleinsasser, R.C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect, 20*, 77–102.

Lado, R. (1961). Language Testing. London: Longman.

Lam, R. (2014). Language assessment training in Hong Kong: Implications for Language assessment literacy. *Language Testing, 32*(2), 169-197.

Liljedahl, P. (2010). The four purposes of assessment. *Vector, 2010*(2), 4–12.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.) Upper Saddle River, NJ: Prentice Hall.

Little, D. (2005). The Common European Framework and the European Language Portfolio:Involving learners and their judgements in the assessment process. *Language Testing, 22*(3), 321-336.

López, A. A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile Issues in Teachers' Professional Development, 11*(2), 55–70.

Loughran, J., Mulhall, P., & Berry, A. (2008): Exploring pedagogical content knowledge inscience teacher education. *International Journal of Science Education, 30* (10). 1310- 1320.

Lynch B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4): 351–372.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education, 20*, 523–535.

Malone, M. (2008). Training in language assessment. In Shohamy, E. and Hornberger, N. (Eds.) Encyclopedia of Language and Education: Vol. *7, Language Testing and Assessment*. New York: Springer, 225 – 239.

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and   users. *Language Testing, 30*(3), 329-344.

Mark, M. M., & Shotland, R. L. (1987). Alternative models for the use of multiple methods. *New Directions for Program Evaluation, 35*, 95- 100.

Maxwell, G. S. (2001). *Teacher Observation in Student Assessment* (pp. 1-14).

(Discussion Paper). Retrieved from the University of Queensland, School of Education.

McCulloch, M. (2007). An Introduction to Assessment. Learning and Teaching Centre, University of Glasgow.1-36.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, *7*(11). Retrieved from http://pareonline.net/ getvn.asp?v=7&n=8.

McMillan, J. H. (2007). Classroom assessment: Principles and practice for effective standards-based instruction (Fourth Ed.). Boston, MA: Pearson Education, Inc. McNamara, T. (2000). *Language testing.* Oxford: Oxford University Press.

Mertler, C. A. (1998). *Classroom assessment practices of Ohio teachers.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL. ERIC Document Reproduction Service No. 428 085.

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, *120*, 285–296.

Mertler, C. A. (2002). Classroom assessment literacy inventory. (Adapted from the Teacher Assessment Literacy Questionnaire (1993), by Barbara S. Plake &James C. Impara, University of Nebraska-Lincoln, in cooperation with The National Council on Measurement in Education &the W. K. Kellogg Foundation.

Mertler, C.A. (2003). Pre-service versus in-service teachers' assessment literacy: does classroom experience make a difference? In Annual meeting of the Mid-Western Educational Research Association, Columbus.

Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the" Assessment Literacy Inventory". *Online Submission.*

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, *12*(2), 101–113.

Mikre, F. (2010). Review Article: The roles of assessment in curriculum practice and enhancement of learning. *Ethiopian Journal of Education and Sciences, 5*(2). doi:10.4314/ejesc. v5i2.65376

Mirici, I. H. (2000). "European Language Portfolio: A tool for a common language education policy in Europe". *Journal of Interdisciplinary Education, 6*(1): 161-166.

Morrow, K. (Ed.), (2004). *Insights from the Common European Framework*. Cambridge, UK: Cambridge University Press.

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*, 71–81.

Mousavi, S. A. (2002). *An encylopedic dictionary of language testing*. The third edition. Taiwan: Tung Hua Book Company.

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. SAGE Publications Ltd.

Mussawy, S. A. (2009). *Assessment Practices: Student's and Teachers' Perceptions of Classroom Assessment* (Unpublished master's thesis). University of Massachusetts.

Newfields, T. (2006). Teacher development and assessment literacy Authentic Communication: *Proceedings of the 5th Annual JALT Pan-SIG Conference*. (pp. 48- 73).

Nguyen, C., & Le, D. (n.d.). Communicative Language Testing: Do School Tests Measure Students' Communicative Competence?. 856-864.

Odo, D. M. (2016). An Investigation of the Development of Pre-service Teacher Assessment Literacy through Individualized Tutoring and Peer Debriefing. *Journal of Inquiry & Action in Education, 7*(2), 31-61.

Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice, *The Curriculum Journal, 25*:3, 344-371.

O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. University of Sydney Papers in TESOL, 1, 71-85.

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing, 30*(3), 363–380.

Otero, V. (2006). Moving beyond the "get it or don't" conception of formative assessment. *Journal of Teacher Education, 57*(3), 247-255.

Oz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy perceptions and practices. *ELT Research Journal, 6*(1), 25-44.

Paterno, J. (2001). Measuring success: A glossary of assessment terms. In *Building cathedrals: Compassion for the 21st century*. Retrieved September, 2016, from http://www.angelfire.com/wa2/buildingcathedrals/measuringsuccess.html

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381-402. doi:10.1177/0265532213480337

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, *6*(1), 21–27.

Pollock, K. (2010). *Occasional Teachers' Access to Professional Learning* [Scholarly project]. In http://edu.uwo.ca/faculty_profiles/cpels/pollock_katina/documents/Occasional_Teac er_Pollock_2010.pdf.

Popham, W. J. (2004). All about accountability / Why assessment illiteracy is professional suicide. *Educational Leadership*, *62*(1), 82-83.

Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership, 63*, 84-85.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*(1), 4-11.

Popham, W. J. (2010). Everything school leaders need to know about assessment. Thousand Oaks: Corwin Press.

Prapphal, K. (1990). *The Relevance of Language Testing Research in the Planning of Language Programmes* (Rep.). Bangkok.

Razavipur, K., Riazi, M., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy. *English Language Teaching*, *4*(1), 156-161.

Razavipour, K. (2013). Assessing Assessment Literacy: Insights From a High-Stakes Test. *RALS*, *4*(1), 111-131.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, *21*(3), 249–258.

Reeves, D. B. (2002a). *Making standards work*. Englewood, CO: Advanced Learning Press.

Richards, J.C., & Schmidt, R. (2002). Longman Dictionary of Language Teaching & Applied Linguistics (3rd edition). Essex: Pearson Education Limited.

Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, *23*, 497-519.

Roscoe, K. (2013). Enhancing assessment in teacher education courses. *The Canadian Journal for the Scholarship of Teaching and Learning*, *4*(1), 1-15. doi: http://dx.doi.org/10.5206/cjsotl-rcacea.2013.1.5

Sah, K. P. (2013). Assessment and Test in Teaching and Learning. *Academic Voices: A Multidisciplinary Journal, 2*(1). doi:10.3126/av.v2i1.8282

Samad, A. A. (2010). *Essentials of language testing for Malaysian teachers*. Selangor: Penerbit Universiti Putra Malaysia.

Sato, M., Wei, C. R., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal, 45*(3), 669e700

Scarino, A. (2013). Language assessment literacy as self awareness: Understanding the role of interpretation in assessment and in teacher learning. *Australia Language Testing 30*(3) 309–327. http://dx.doi.org/10.1177/0265532213480128

Schafer, W. D., & Lizzitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57e63

Schneider, G., & Lenz, P. (2001). *European language portfolio: Guide for developers*. Strasbourg: Council of Europe.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York, NY: Teachers College. Shepard, L. A. (2000). *The Role of Classroom Assessment in Teaching and Learning* (Rep.).

Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco: Jossey Bass.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education, 22*(4), 371–391. http://dx.doi.org/10.1007/s10972-011-9231-6

Smith, C. W., & Cumming, J. (2009). *Educational assessment in the 21st century.* NewYork: Springer B.V.

Spiller, D. (2012). *Assessment Matters: Self Assessment and Peer Assessment* (pp. 1-12).

Spolsky, B. (1995). *Measured words.* Oxford: Oxford University Press.

Stiggins, R. (1991). Assessment Literacy. *Phi Delta Kappan, 72(7*), 534 – 539.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappa, 77*(3), 238- 245.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758-65.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment FOR student learning: Doing it right—using it well.* Portland, OR: ETS Assessment Training Institute.

Stiggins, R. J. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324–328.

Stiggins, R. J. (2008). Assessment for learning, the achievement gap, and truly effective schools. Portland, OR: ETS Assessment Training Institute. 3–14.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265-269.

Stobart, G. (2012). Validity in formative assessment. In J. Gardner, (Ed.). *Assessment and learning, 233-242.* London: Sage Publications, Inc.

Talanquer, V., Tomanek, D., & Novodvorsky, I. (2012). Assessing students' understanding of inquiry: What do prospective science teachers notice? *Journal of Research in Science Teaching, 50*(2), 189-208. http://dx.doi.org/10.1002/tea.21074

Taras, M. (2005). Assessment-Summative and Formative--Some theoretical reflections. *British Journal of Educational Studies*, 466-478.

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The New Era of Mixed Methods. *Journal of Mixed Methods Research*, *1*(1), 3-7.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research.* Thousand Oaks: Sage.

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

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, *30*(3), 403– 412.

Thomas, J., Allman, C., & Beech, M. (2004). Assessment for the diverse classroom:   A handbook for teachers. Tallahassee, FL: Florida Department of Education, Bureau of Exceptional Education and Student Services. Retrieved from http://www.fldoe.org/ese/pdf/assess diverse.pdf

Tombari, M., & Borich, G. (1999). Authentic assessment in the classroom. New Jersey: Prentice-Hall, Inc.

Vogt, K., & Tsagari, D. (2014) Assessment Literacy of Foreign Language Teachers: Findings of a European Study, *Language Assessment Quarterly, 11*:4, 374-402, DOI: 10.1080/15434303.2014.960046

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.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 291–302). Dordrecht: Kluwer Academic.

Webb, N. (2002). Assessment literacy in Standards-based education setting. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, April1-5.Retrieved from:http://www.wcer.wisc.edu/publications/workingpapers/Working_Paper_ No_202_4.pdf

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*,   *37*(1), 3-14.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will  it take to make it work? In C. A. Dwyer (Ed.). *The future of assessment:*

*Shaping teaching and learning.* (pp. 53-82). New York: Lawrence Erlbaum Associates.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations.* Washington,

DC:    University of Washington, Center for the Study of Teaching and Policy.

Wojtczak, A. (2002). "Glossary of Medical Education Terms", Retrieved May 13, 2009 from http://www.iime.org/glossary.htm

Wood, D. L. (2007). *Assessment for learning: Connecting students to their learning* (Unpublished master's thesis).

Yetkin, C. (2015). *An Investigation on ELT Teacher Candidates' Assessment Literacy* (Unpublished master's thesis). Cag University.

YÖK (Yükseko ğretim Kurulu [The Council of Higher Education]). 2009. *Yükseköğretim Kataloğu 2009 [Higher Education Catalogue 2009].* (Retrieved September 21, 2017 from http://www.yok.gov.tr/documents/10279/49665/ingilizce_ogretmenligi.pdf/2 a01aab-772043e0-8215-749d3f5f8838).

# APPENDIX-A: Original Questionnaire

## Questions about training in LTA

### 1. Classroom-focused LTA

1.1. Please specify if you were trained in the following domains.

|  | Not at all | A little (1-2 days) | More advanced |
|---|---|---|---|
| a) Preparing classroom tests | ☐ | ☐ | ☐ |
| b) Using ready-made tests from textbook packages | ☐ | ☐ | ☐ |
| 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. Please specify if you need training in the following domains.

|  | None | Yes, basic training | Yes, more advanced training |
|---|---|---|---|
| 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. Please specify if you were trained in the following domains.

|  | Not at all | A little (1-2 days) | More |
|---|---|---|---|
| *advanced* |  |  |  |
| a) Giving grades | ☐ | ☐ | ☐ |
| b) Finding out what needs to be taught/learned | ☐ | ☐ | ☐ |
| c) Placing students onto courses, programs, etc. information from tests/assessment | ☐ | ☐ | ☐ |
| d) Awarding final certificates (from school/program; local, regional) | ☐ | ☐ | ☐ |

2.2. Please specify if you need training in the following domains.

|  | None | Yes, basic training | Yes, more advanced training |
|---|---|---|---|
| a) Giving grades | ☐ | ☐ | ☐ |
| b) Finding out what needs to be taught/learned | ☐ | ☐ | ☐ |
| c) Placing students onto courses, programs, etc. information from tests/assessment | ☐ | ☐ | ☐ |
| d) Awarding final certificates (from school/program; local, regional) | ☐ | ☐ | ☐ |

### 3. Content and concepts of LTA

3.1. Please specify if you were trained in the following domains.

|  | Not at all | A little (1-2 days) | More advanced |
|---|---|---|---|
| 1. Testing/Assessing: |  |  |  |
| a) Receptive skills (reading/listening) | ☐ | ☐ | ☐ |
| b) Productive skills (speaking/writing) | ☐ | ☐ | ☐ |
| c) Microlinguistic aspects (grammar/vocabulary) | ☐ | ☐ | ☐ |
| d) Integrated language skills | ☐ | ☐ | ☐ |
| e) Aspects of culture | ☐ | ☐ | ☐ |
| 2. Establishing reliability of tests/assessment | ☐ | ☐ | ☐ |
| 3. Establishing validity of tests/assessment | ☐ | ☐ | ☐ |
| 4. Using statistics to study the quality of tests/assessment | ☐ | ☐ | ☐ |

3.2. Please specify if you were trained in the following domains.

|  | None | Yes, basic training | Yes, more advanced training |
|---|---|---|---|
| 5. Testing/ Assessment | ☐ | ☐ | ☐ |
| f) Receptive skills (reading/listening) | ☐ | ☐ | ☐ |
| g) Productive skills (speaking/writing) | ☐ | ☐ | ☐ |
| h) Microlinguistic aspects (grammar/vocabulary) | ☐ | ☐ | ☐ |
| i) Integrated language skills | ☐ | ☐ | ☐ |
| j) Aspects of culture | ☐ | ☐ | ☐ |
| 6. Establishing reliability of tests/ assessment | ☐ | ☐ | ☐ |
| 7. Establishing validity of tests/ assessment | ☐ | ☐ | ☐ |
| 8. Using statistics to study the quality of tests/assessment | ☐ | ☐ | ☐ |

# APPENDIX-B: Language Assessment Literacy Survey

Dear Participant,

The purpose of the following survey is to find out your training in language testing and assessment and your training needs in this area. There are no 'right' or 'wrong' answers. Please be assured that all the information included in this survey is confidential.

Your answers will have a valuable contribution to the study.

Thank you very much for your participation.

Gamze Sarıyıldız

Middle East Technical University-Foreign Language Education Department

## Part I. General Information

1)         Age: .......

2)         Gender: Female ☐      Male ☐

3)         High school:

     Anatolian Teacher Training High School

     Anatolian High School

     Private High School

     Other (......................)

4) University: .........

5) GPA: ..........

6) Grade:  1 ☐     2 ☐     3 ☐     4 ☐

7) Which is your home country?  Turkey ☐     Other ☐ (.........................)

8) Which language is your first language? ..................

9) Have you ever taken any testing and evaluation course before?

     Yes ☐

     No ☐

10) Have you ever taken any English Language Testing and Evaluation (ELTE) courses before?

     Yes ☐

No ☐

11) Main course book of English Language Testing and Evaluation (ELTE) course:

......................................................................................................

....

12) School at which you are doing practicum:

State School ☐

Private School ☐

13) Level of education at which you are doing practicum:

Pre-primary School ☐

Primary School ☐

Secondary School ☐

High School ☐

University ☐

## Part II. Questions about Training in LTA

## II.I. Classroom-focused LTA

A. Please specify the amount of training you think you have received in the following domains by ticking the box.

| | Training | | | |
|---|---|---|---|---|
| | Received | | | |
| | None | Little (1-2 days) | Sufficient | Advanced |
| 1) Preparing classroom tests | | | | |
| 2) Preparing diagnostic tests | | | | |

| | | | | |
|---|---|---|---|---|
| 3) Preparing achievement tests | | | | |
| 4) Preparing proficiency tests | | | | |
| 5) Preparing placement tests | | | | |
| 6) Preparing progress tests | | | | |
| 7) Preparing language aptitude tests | | | | |
| 8) Using ready-made tests from textbook packages or from other sources | | | | |
| 9) Adapting ready-made tests for the needs of students | | | | |
| 10)Stages of language test construction(e.g. objectives, drawing up test specifications...) | | | | |
| 11)Scoring | | | | |
| 12)Grading | | | | |
| 13)Giving feedback to students based on information from tests/assessment | | | | |
| 14)Interpreting test scores | | | | |
| 15)Using self/peer assessment | | | | |
| 16)Using informal,non- | | | | |

| | test type of assessment (e.g. essays, presentations, homeworks) | | | | |
|---|---|---|---|---|---|
| 17)Using continuous type of assessment (e.g. quizzes) | | | | | |
| 18)Using European Language Portfolio | | | | | |

B. Please specify the amount of training you think you need in the following domains by ticking the box.

| | Training | | |
|---|---|---|---|
| | **Needed** | | |
| | None | Yes, basic training | Yes, more advanced training |
| 1) Preparing classroom tests | | | |
| 2) Preparing diagnostic tests | | | |
| 3) Preparing achievement tests | | | |
| 4) Preparing proficiency tests | | | |
| 5) Preparing placement tests | | | |
| 6) Preparing progress tests | | | |
| 7) Preparing language aptitude tests | | | |
| 8) Using ready-made tests from textbook packages or | | | |

| | | | |
|---|---|---|---|
| from other sources | | | |
| 9) Adapting ready-made tests for the needs of students | | | |
| 10) Stages of language test construction(e.g. objectives, drawing up test specifications...) | | | |
| 11) Scoring | | | |
| 12) Grading | | | |
| 13) Giving feedback to students based on information from tests/assessment | | | |
| 14) Interpreting test scores | | | |
| 15) Using self/peer assessment | | | |
| 16) Using informal,non-test type of assessment (e.g. essays, presentations, homeworks) | | | |
| 17) Using continuous type of assessment (e.g. quizzes) | | | |
| 18) Using European Language Portfolio | | | |

## II.II. Knowledge of Testing and Assessment

A. Please specify the amount of training you think you have received in the following domains by ticking the box.

| | Training Received | | | |
|---|---|---|---|---|
| | None | Little (1-2 days) | Sufficient | Advanced |
| 1) Informal/ Formal assessment | | | | |
| 2) Formative/ Summative assessment | | | | |
| 3) Norm /Criterion-referenced assessment | | | | |
| 4) Discrete point/Integrative testing | | | | |
| 5) Direct/Indirect testing | | | | |
| 6) Objective/Subjective testing | | | | |
| 7) Approaches to language testing (e.g. integrative, communicative, structuralist) | | | | |
| 8) Alternative assessment | | | | |
| 9) Computer-based testing | | | | |

B. Please specify the amount of training you think you need in the following domains by ticking the box.

| | Training Needed | | |
|---|---|---|---|
| | None | Yes, basic training | Yes, more advanced training |

| | | | |
|---|---|---|---|
| 1) Informal/ Formal assessment | | | |
| 2) Formative/ Summative assessment | | | |
| 3) Norm /Criterion-referenced assessment | | | |
| 4) Discrete point/Integrative testing | | | |
| 5) Direct/Indirect testing | | | |
| 6) Objective/Subjective testing | | | |
| 7) Approaches to language testing (e.g. integrative, communicative, structuralist) | | | |
| 8) Alternative assessment | | | |
| 9) Computer-based testing | | | |

## II.III. Purposes of Testing

A. Please specify the amount of training you think you have received in the following domains by ticking the box.

| | Training Received | | | |
|---|---|---|---|---|
| | None | Little (1-2 days) | Sufficient | Advanced |
| 1) Giving grades | | | | |
| 2) Finding out what needs to be learned/taught | | | | |
| 3) Placing students onto programs, courses, etc. | | | | |
| 4) Testing competence in a language | | | | |
| 5) Identifying what has been learned | | | | |

| | 6) Measuring general ability to learn a foreign language | | | | |
|---|---|---|---|---|---|
| | 7) Awarding final certificates (from school, program; local, regional/ national level) | | | | |

B. Please specify the amount of training you think you need in the following domains by ticking the box.

| | Training | | |
|---|---|---|---|
| | Needed | | |
| | None | Yes, basic training | Yes, more advanced training |
| 1) Giving grades | | | |
| 2) Finding out what needs to be learned/taught | | | |
| 3) Placing students onto programs, courses, etc. | | | |
| 4) Testing competence in a language | | | |
| 5) Identifying what has been learned | | | |
| 6) Measuring general ability to learn a foreign language | | | |
| 7) Awarding final certificates (from school, program; local, regional/ | | | |

| | national level) | | | |
|---|---|---|---|---|

## II.IV. Content and concepts of LTA

A. Please specify the amount of training you think you have received in the following domains by ticking the box.

| | **Training** | | | |
|---|---|---|---|---|
| | **Received** | | | |
| | None | Little (1-2 days) | Sufficient | Advanced |
| 1) Testing reading in English | | | | |
| 2) Different test items/task types to test reading in English | | | | |
| 3) Testing listening in English | | | | |
| 4) Different test items/task types to test listening in English | | | | |
| 5) Testing speaking in English | | | | |
| 6) Different test items/task types to test speaking in English | | | | |
| 7) Testing writing in English | | | | |
| 8) Different test items/task types to test writing in English | | | | |
| 9) Testing Grammar in English | | | | |
| 10) Different test items/task types to test grammar in English | | | | |
| 11) Testing Vocabulary in English | | | | |

| | | | | |
|---|---|---|---|---|
| 12)Different test items/task types to test vocabulary in English | | | | |
| 13)Testing integrated language skills | | | | |
| 14)Testing pronunciation in English | | | | |
| 15)Different test items/question types to test pronunciation in English | | | | |
| 16)Practicality | | | | |
| 17)Reliability (e.g. rater, test administration, test, student-related) | | | | |
| 18)Validity (face, construct, criterion, content) | | | | |
| 19)Authenticity | | | | |
| 20)Washback | | | | |
| 21)Using statistics to study the quality of tests / assessment | | | | |
| 22)Alternatives in assessment (portfolios,conferences, interviews, observations,self/peer assessment) | | | | |

B. Please specify the amount of training you think you need in the following domains by ticking the box.

| | Training | | |
|---|---|---|---|
| | **Needed** | | |
| | None | Yes, basic training | Yes, more advanced training |
| 1) Testing reading in English | | | |
| 2) Different test items/task types to | | | |

| | | | |
|---|---|---|---|
| test reading in English | | | |
| 3) Testing listening in English | | | |
| 4) Different test items/task types to test listening in English | | | |
| 5) Testing speaking in English | | | |
| 6) Different test items/task types to test speaking in English | | | |
| 7) Testing writing in English | | | |
| 8) Different test items/task types to test writing in English | | | |
| 9) Testing Grammar in English | | | |
| 10) Different test items/task types to test grammar in English | | | |
| 11) Testing Vocabulary in English | | | |
| 12) Different test items/task types to test vocabulary in English | | | |
| 13) Testing integrated language skills | | | |
| 14) Testing pronunciation in English | | | |
| 15) Different test items/question types to test pronunciation in English | | | |
| 16) Practicality | | | |
| 17) Reliability (e.g. rater, test administration, test, student-related) | | | |
| 18) Validity (face, construct, criterion, content) | | | |
| 19) Authenticity | | | |

| | | | |
|---|---|---|---|
| 20)Washback | | | |
| 21)Using statistics to study the quality of tests / assessment | | | |
| 22)Alternatives in assessment (portfolios,conferences, interviews, observations,self/peer assessment) | | | |

1) Do you think that taking ELTE course helped you during your school experience/practicum?

   Yes ☐

   No ☐

   a) If the answer is YES, please explain how it helped you during practicum.

   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   ------------.

2) Has your school experience course covered anything related to the theory of language testing and assessment?

   Yes ☐

   No ☐

   a) If the answer is YES, please list 5 topics related to the theory of language testing and assessment that you learnt during school experience course.

| 1) |
|---|
| 2) |
| 3) |
| 4) |
| 5) |

3) Do you think that topics related to the theory of language testing and assessment could be covered in school experience course? Explain your answer.

   -----------------------------------------------------------------------------------------------
   -----------------------------------------------------------------------------------------------
   -----------------------------------------------------------------------------------------------

------------------------------------------------------------------------------------------------

--------------------------------------------------------------------------------------------.

4) Has your school experience course given you the chance to put what you learned related to language testing and assessment into practice?

Yes ☐

No ☐

   a) If the answer is YES, could you list 5 practices related to language testing and assessment during your school experience course?

| |
|---|
| 1) |
| 2) |
| 3) |
| 4) |
| 5) |

5) Do you think that the practice related to the language testing and assessment could be realized in school experience course? Explain your answer.

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-----------------------------------------------------------------------------------------.

6) Did you have a chance to reflect on your experiences related to language testing and assessment in practicum?

Yes ☐

No ☐

7) Please, list 5 topics related to language testing and assessment, which in your opinion, must be covered in school experience course and explain how these will help you.

| Topics | How do they help? |
|---|---|
| 1) | |
| 2) | |
| 3) | |
| 4) | |
| 5) | |

8) Please, list 5 topics related to language testing and assessment which, in your opinion, are problematic during your observation in practicum.

| | |
|---|---|
| 1) | |
| 2) | |
| 3) | |
| 4) | |
| 5) | |

1) Yabancı dil öğretiminde ölçme ve değerlendirme dersini almanızın stajınızda size yardımcı olduğunu düşünüyor musunuz?

   Evet ☐

   Hayır ☐

   a) Eğer cevabınız **Evet** ise, lütfen bu dersi almanın size stajda nasıl yardımcı olduğunu açıklayınız.

   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   -------------------------------------------------------------------------------------------
   ------------.

2) Okul deneyimi dersinizde, dil ölçme ve değerlendirme alanında kuramsal herhangi bir şey öğrendiniz mi?

   Evet ☐

   Hayır ☐

   a) Eğer cevabınız **Evet** ise, lütfen dil ölçme ve değerlendirme alanında kuramsal olarak öğrendiğiniz 5 konuyu listeleyiniz.

   | 1) |
   |----|
   | 2) |
   | 3) |
   | 4) |
   | 5) |

3) Okul deneyimi dersinizde, dil ölçme ve değerlendirme alanında kuramsal konuların ele alınabileceğini düşünüyor musunuz? Cevabınızı açıklayınız.

4) Okul deneyimi dersiniz size dil ölçme ve değerlendirme alanında öğrendiklerinizi uygulamaya koyma şansı verdi mi?

Evet ☐

Hayır ☐

a) Eğer cevabınız **Evet** ise, dil ölçme ve değerlendirme ile ilgili okul deneyimi dersinizdeki 5 uygulamayı sıralar mısınız?

| | |
|---|---|
| 1) | |
| 2) | |
| 3) | |
| 4) | |
| 5) | |

5) Okul deneyimi dersinizde, dil ölçme ve değerlendirme alanıyla ilgili pratiğin gerçekleştirilebileceğini düşünüyor musunuz? Cevabınızı açıklayınız

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------

------------.

6) Stajınız süresinde dil ölçme ve değerlendirme alanıyla ilgili deneyimleriniz üzerine düşünme fırsatı buldunuz mu?

Evet ☐

Hayır ☐

7) Lütfen, okul deneyimi dersinizde ele alınması gerektiğini düşündüğünüz dil ölçme ve değerlendirme alanıyla ilgili 5 tane konu sıralayıp bunların size nasıl yardımcı olacağını açıklayınız.

| Konular | Nasıl yardımcı olur? |
|---|---|
| 1) | |
| 2) | |
| 3) | |

| | |
|---|---|
| 4) | |
| 5) | |

8) Lütfen, stajınızdaki gözleminiz süresinde dil ölçme ve değerlendirme alanıyla ilgili problemli olduğunu düşündüğünüz 5 konu sıralayınız.

| |
|---|
| 1) |
| 2) |
| 3) |
| 4) |
| 5) |

# APPENDIX-E: Language Assessment Literacy Questionnaire

## Consent Form

**INTRODUCTION:**

You are invited to consider participating in this research study. Please take as much time as you need to make your decision. Feel free to discuss your decision with whomever you want, but remember that the decision to participate, or not to participate, is yours. If you decide that you want to participate, please sign in your name in the space provided at the bottom of this page.

**PURPOSE:**

The purpose of this study is to investigate preservice English as a Foreign Language Teachers' language assessment literacy. You are being asked to participate in this study by completing one survey related to this topic.

**RISKS & BENEFITS:**

When you participate in this research study, there are no known risks greater than those encountered in everyday life. While the study may not provide any direct benefits to you, your participation will contribute to the professional knowledge base on English language teacher training.

**COMPENSATION FOR PARTICIPATION:**

There will be no monetary compensation to you. Nor will there be any costs to you for participating in the study.

**CONFIDENTIALITY:**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. However, it is impossible to guarantee absolute confidentiality. In order to keep information about you safe, computerized data will be kept in a password-protected file on the personal computer which only the researcher can access. Your name or other identifiable information will not be included in the final product that reports the research results. Please note that, even if your name is not used in publication, the researcher will still be able to connect you to the information gathered about you in this study.

**PARTICIPATION & WITHDRAWAL:**

Participation is completely voluntary. You can choose whether or not to be in this study. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind or loss of benefits to which you are otherwise entitled. You may also refuse to answer any questions you do not want to answer. There is no penalty if you withdraw from the study and you will not lose any benefits to which you are otherwise entitled.

**SUMMARY OF RESULTS:** A summary of the results of this research will be supplied to you, at no cost, upon your request.

**VOLUNTARY CONSENT (SIGNATURE):** _____

By completing and submitting this questionnaire to the researcher, you are voluntarily agreeing to participate. If you have any questions about the study, please contact the researcher and/or the thesis supervisor via email or phone.

| Gamze Sarıyıldız, MA Candidate | Hüseyin Öz, PhD – Thesis Supervisor |
|---|---|
| Middle East Technical University | Associate Professor |
| ELT Department | Hacettepe University |
| ODTÜ, Çankaya 06800, Ankara | ELT Department |
| Phone  : (507) 981-68-20 | Beytepe, Çankaya 06800, Ankara |
| E-mail  : gmzsryldz@gmail.com | Phone  : 312-297-8575 |
| | E-mail  : hoz@hacettepe.edu.tr |

# APPENDIX-F: Ethics Committee Approval

**T.C.**
**HACETTEPE ÜNİVERSİTESİ**
Rektörlük

Sayı : 35853172/ 433-32.21

27 Ekim 2016

## EĞİTİM BİLİMLERİ ENSTİTÜ MÜDÜRLÜĞÜNE

İlgi:   12.10.2016 tarih ve 2381 sayılı yazınız.

Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Bilim Dalı tezli yüksek lisans programı öğrencilerinden **Gamze SARIYILDIZ**'ın **Yrd. Doç. Dr. Hüseyin ÖZ** danışmanlığında yürüttüğü **"Türkiye'deki İngilizce Öğretmen Adaylarının Dil Değerlendirme Okuryazarlığı Üzerine Bir Araştırma/A Study Into Language Assessment Literacy Of Preservice English As a Foreign Language Teachers in Türkish Context"** başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun **18 Ekim 2016** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Rahime M. NOHUTCU
Rektör a.
Rektör Yardımcısı

# APPENDIX-G: Declaration of Ethical Conduct

I hereby declare that...

- I have prepared this thesis in accordance with the thesis writing guidelines of the Graduate School of Educational Sciences of Hacettepe University;

- all information and documents in the thesis/dissertation have been obtained in accordance with academic regulations;

- all audio visual and written information and results have been presented in compliance with scientific and ethical standards;

- in case of using other people's work, related studies have been cited in accordance with scientific and ethical standards;

- all cited studies have been fully and decently referenced and included in the list of References;

- I did not do any distortion and/or manipulation on the data set,

- and **NO** part of this work was presented as a part of any other thesis study at this or any other university.

04 /06/ 2018

Gamze Sarıyıldız

# APPENDIX-H: Thesis Originality Report

24/04/2018

HACETTEPE UNIVERSITY

Graduate School of Educational Sciences

To The Department of Foreign Language Education

Thesis Title: A STUDY INTO LANGUAGE ASSESSMENT LITERACY OF PRESERVICE ENGLISH AS A FOREIGN LANGUAGE TEACHERS IN TURKISH CONTEXT

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

| Time Submitted | Page Count | Character Count | Date of Thesis Defence | Similarity Index | Submission ID |
|---|---|---|---|---|---|
| 24/04/2018 | 149 | 282,642 | 04/06/2018 | 13% | 952557143 |

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

| | | |
|---|---|---|
| Name Lastname: | Gamze Sarıyıldız | |
| Student No.: | N14221507 | Signature |
| Department: | Foreign Language Education | |
| Program: | English Language Teaching | |
| Status: | ☒ Masters ☐ Ph.D. ☐ Integrated Ph.D. | |

**ADVISOR APPROVAL**

APPROVED

Assoc. Prof. Dr. Hüseyin ÖZ

# APPENDIX-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversite'ye verilen kullanım hakları dışındaki bütün fikrî mülkiyet haklarım bende kalacak, tezimin tamamının veya bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversite'ye teslim etmeyi taahhüt ederim.

☒ **Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirse bile, teziniz arama motorlarının ön belleklerinde kalmaya devam edebilecektir)

☐ **Tezimin/Raporumun ................. tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

☐ **Tezimin/Raporumun ................. tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

☐ **Serbest Seçenek/Yazarın Seçimi:**

...............................................................................................................
...............................................................................................................
...............................................................................................................
...

04/06/2018

Gamze Sarıyıldız