

**TC. KOCAELİ UNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI**

**EXPLORING THE LANGUAGE ASSESSMENT LITERACY OF
TURKISH IN-SERVICE EFL TEACHERS**

YÜKSEK LİSANS TEZİ

Sezgin BALLIDAĞ

KOCAELİ 2020

TC. KOCAELİ UNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI

**EXPLORING THE LANGUAGE ASSESSMENT LITERACY OF
TURKISH IN-SERVICE EFL TEACHERS**

YÜKSEK LİSANS TEZİ

Sezgin BALLIDAĞ

Assoc. Prof. Dr. Banu İNAN KARAGÜL

Tezin Kabul Edildiği Enstitü Yönetim Kurulu Karar ve No: 10.06.2020/ 13

KOCAELİ 2020

To my wife and my sons



ACKNOWLEDGEMENTS

First of all, I wish to express my sincere appreciation to my supervisor, Assoc. Prof. Dr. Banu İNAN KARAGÜL for her invaluable and constructive feedback, kindness and friendly support throughout this challenging journey.

I would also like to express my gratitude to Assoc. Prof. Dr. Dođan YÜKSEL and Asst. Prof. Dr. Mehmet ALTAY for their helpful contributions and generous support when I turned to them. I also thank Asst. Prof. Dr. Sezen ARSLAN for her participation to my thesis defense.

I am also thankful to my friends who live in various cities of our country since they helped me a lot through the data collection process, which was really precious for me. Without their support, this thesis would not have been possible.

Thanks also should go to my dear office-mates at school, Ali Çađrı, Emin and Aslı for continuously urging me to complete my master's degree as soon as possible.

I would like to extend my deepest gratitude to my family for their continuous and unrelenting support. They have always had a profound belief in my studies. I am especially grateful to my sister, Aylin for always being there for me whenever I needed help.

Finally, my deepest appreciation goes to my wife, Meltem for her endless patience, support and understanding during my master's education. She has made countless sacrifices to help me get to this point, thank you for being my greatest supporter. Along with her, I would like to acknowledge my sons, Arda and Mete. For the last two years, they have always seen their dad sitting in front of a computer whenever he has been at home. I am sorry boys for all those times I missed playing with you.

Sezgin BALLIDAĖ

Kocaeli, 2020

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	iii
LIST OF TABLES	iv
ABSTRACT	v
ÖZET	vi
CHAPTER I.....	1
1. INTRODUCTION	1
1.1. BACKGROUND OF THE STUDY	1
1.2. STATEMENT OF THE PROBLEM	2
1.3. SIGNIFICANCE OF THE STUDY.....	3
1.4. RESEARCH QUESTIONS	4
1.5. DEFINITIONS OF TERMS	4
1.6. LIMITATIONS OF THE STUDY.....	4
CHAPTER II.....	5
2. LITERATURE REVIEW	5
2.1 ASSESSMENT	5
2.1.1. Alternative Assessment	6
2.1.2. Dynamic Assessment	8
2.2. ASSESSMENT LITERACY	10
2.3 LANGUAGE ASSESSMENT LITERACY	12
2.4. STUDIES ON LANGUAGE ASSESSMENT LITERACY OF IN-SERVICE AND PRE-SERVICE TEACHERS ABROAD AND IN TURKEY	14
2.4.1 Research Conducted Abroad	14
2.4.2. Research Conducted in Turkey.....	21
CHAPTER III	26
3. METHODOLOGY	26
3.1. RESEARCH DESIGN	26
3.2. SETTING AND PARTICIPANTS	26
3.3. DATA COLLECTION INSTRUMENTS.....	28
3.4 DATA COLLECTION PROCEDURE AND DATA ANALYSIS.....	29

CHAPTER IV	32
4. FINDINGS.....	32
4.1 QUANTITATIVE DATA ANALYSIS	32
4.1.1 Results of Data Analysis for Research Question 1	32
4.1.2 Results of Data Analysis for Research Question 2	38
4.2 QUALITATIVE DATA ANALYSIS.....	43
CHAPTER V	54
5. DISCUSSION, CONCLUSION and SUGGESTIONS.....	54
5.1. THE OVERVIEW OF THE STUDY	54
5.2. DISCUSSION OF THE FINDINGS.....	55
5.3. PEDAGOGICAL IMPLICATIONS OF THE STUDY	66
5.4. LIMITATIONS OF THE STUDY AND SUGGESTIONS FOR FUTURE RESEARCH	66
REFERENCES.....	68
APPENDICES	75
APPENDIX A. In-Service Language Teachers’ Questionnaire	75
APPENDIX B. Interview Questions	78
APPENDIX C. Ethics Committee Approval.....	80
CURRICULUM VITAE.....	80

LIST OF ABBREVIATIONS

- LAL: Language Assessment Literacy**
LTA: Language Testing and Assessment
ELTE: English Language Testing and Evaluation
AL: Assessment Literacy
DA: Dynamic Assessment
EFL: English as a Foreign Language
ELT: English Language Teaching
SCT: Sociocultural Theory
ZPD: Zone of Proximal Development
AID: Action-Impact-Desired Outcome
LAKS: Language Assessment Knowledge Scale

LIST OF TABLES

TABLE 1: The Profile of the Participants	27
TABLE 2: The Profile of the Interviewed Participants.....	28
TABLE 3: The Participants' Background in LTA Course.....	32
TABLE 4: Perceived Training Levels of Participants in Classroom-Focused LTA	33
TABLE 5: Perceived Training Levels of Participants in Purposes of Testing.....	34
TABLE 6: Perceived Training Levels of Participants in Content and Concepts of LTA.....	36
TABLE 7: Perceived Training Levels of Participants in All Domains of LTA.....	37
TABLE 8: Perceived Training Needs of Participants in Classroom-Focused LTA	38
TABLE 9: Perceived Training Needs of Participants in Purposes of Testing.....	40
TABLE 10: Perceived Training Needs of Participants in Content and Concepts of LTA.....	41
TABLE 11: Perceived Training Needs of the Participants in All Domains of LTA	43
TABLE 12: The Assessment Tools Used by the Interviewees	50

ABSTRACT

This study aimed to explore the language assessment literacy (LAL) of English language teachers working at state elementary, middle, and high schools in Turkey. 101 teachers working in various cities participated in this study. A mixed-methods research design was employed in the study, so the data were collected through both quantitative (questionnaires) and qualitative (interviews) instruments. The data for the quantitative part of the study were collected by means of the “Teachers Questionnaire” by Vogt and Tsagari (2014) under three domains; classroom-focused language testing and assessment (LTA), purposes of testing and content and concepts of LTA. The data from questionnaires were analyzed via SPSS in terms of means, percentages, and frequencies to find out the training levels and needs of the respondents in language assessment. In the qualitative phase, a total of 18 teachers volunteered to take part in the interviews. The qualitative data that was gathered through semi-structured interviews were analyzed one by one in terms of their similarities and differences, and selective coding was utilized. The overall results from the questionnaires indicated that EFL teachers working at state elementary, middle, and high schools in Turkey perceived their training levels in all three domains of LTA insufficient, and they were also in need of further basic training in those domains. On the other hand, qualitative findings also supported the findings from the questionnaires by also providing insights into the reasons for participants’ insufficiency of LAL levels.

Keywords: EFL, language assessment literacy, language testing and assessment,

ÖZET

Bu çalışma Türkiye'deki ilkokul, ortaokul ve lise kademesindeki devlet okullarında çalışan İngilizce öğretmenlerinin dil değerlendirme okuryazarlığını incelemeyi hedeflemiştir. Bu çalışmaya, farklı şehirlerde çalışan 101 öğretmen katılmıştır. Çalışmada, yöntem olarak karma yöntemden yararlanılmıştır, bu yüzden veri, hem nicel (anket) hem de nitel araştırma (mülakat) araçları aracılığıyla toplanmıştır. Araştırmanın nicel bölümünün verisi Vogt ve Tsagari'nin (2014) Öğretmen Anketi vasıtasıyla üç alan altında toplanmıştır; sınıf odaklı dilde ölçme ve değerlendirme, ölçmenin amaçları ve dilde ölçme ve değerlendirmenin içerik ve kavramları. Anketlerden toplanan veri, katılımcıların eğitim seviyeleri ve ihtiyaçlarını belirlemek için ortalamalar, yüzdeler ve sıklıklar bakımından SPSS aracılığıyla analiz edilmiştir. Nitel aşamada, toplam 18 öğretmen mülakatlara katılmaya gönüllü olmuştur. Yarı-yapılandırılmış anketler vasıtasıyla toplanan veri benzerlik ve farklılıkları bakımından tek tek analiz edilmiş ve seçici kodlama yapılmıştır. Anketlerden elde edilen genel sonuçlar, devlet ilkokulu, ortaokulu ve lisesinde çalışan İngilizce öğretmenlerinin, yabancı dilde ölçme değerlendirmenin tüm üç alanında da eğitim seviyelerini yetersiz olarak gördüklerini ve aynı zamanda bu alanlarda temel bir eğitime ihtiyaç duyduklarını göstermiştir. Diğer taraftan, nitel bulgular da anketlerden elde edilen veriyi, katılımcıların dil değerlendirme okuryazarlığının yetersizliğinin sebepleri konusunda fikir de vererek desteklemektedir.

Anahtar Kelimeler: dil değerlendirme okuryazarlığı, dilde ölçme ve değerlendirme, yabancı dil olarak İngilizce

CHAPTER I

1. INTRODUCTION

1.1. BACKGROUND OF THE STUDY

Improving the teachers' professional knowledge has been an area of interest for many decades (Tellez and Mosqueda, 2015: p. 87); thus, there has been a process towards professionalism in the field of language testing and assessment (Farhady, 2019: p. 4). This process to improve the ability of teachers in testing-related activities brought about a new term, *language assessment literacy* (LAL). Various definitions of the term have been suggested in the literature. O'Loughlin (2013) defines the term as having "a range of skills related to test production, test score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within education and society" (p. 363). A further definition is given by Pill and Harding (2013) who describe LAL as "a repertoire of competences that enable an individual to understand, evaluate and, in some cases, create language tests and analyze test data" (p. 382).

LAL has become an important area of research in the field of language teaching and learning, and the last decade has witnessed an increasing amount of literature that seeks to define LAL conceptually, and how to learn it and develop it productively (Lam, 2015: p. 170). The research conducted regarding LAL has a focus on basically three categories; the effectiveness of the assessment courses at pre-service education, the extent of teachers' knowledge on assessment, and their self-expressed needs and beliefs in assessment issues. Pre-service education years are of utmost importance in equipping prospective teachers with necessary assessment knowledge and practice. Despite its significance, there is an insufficient amount of literature that has been published to investigate the potency of the English language testing and evaluation (ELTE) courses offered at pre-service education. The content of the course is crucial. However, what needs to be included in the ELTE courses is a concern that administrators and teachers are confronted with (Berger, 2012: p. 65). Since most of the teacher educators focus on the theoretical aspects of the language

assessment (Jeong, 2013: p. 353), teachers may not reflect their theoretical knowledge in actual classrooms when they start work. Most of the research conducted on the efficacy of ELTE courses revealed that students need more training in language assessment (Hatipoğlu, 2010: p.124; Volante and Fazio, 2007: p. 764). In the face of these facts, not enough attention has been devoted to the ELTE courses taught at education programs (Hatipoğlu, 2015: p. 116; Jin, 2010: p. 555; Orr, 2010: p. 4).

How much background knowledge teachers have and need on assessment issues is another concern of the conducted research in the field. Several attempts have been made to investigate the LAL knowledge of teachers (Alderson, 2005: p. 4; Fulcher, 2012: p. 113; Köksal, 2004: p. 1; Muhammad and Bardakçı, 2019: p. 431; Sarıçoban, 2011: p. 398; Tzagari and Vogt, 2017: p. 41). The results of various research indicate that the LAL of the teachers was underdeveloped. A recent study by Hakim (2015: p. 42) demonstrated that the amount of LAL correlated with the participants' experience. In an attempt to reveal the LAL, not only did the researchers evaluate the exams produced by teachers or ask assessment related questions, but they also inquired participants' perceived levels of assessment knowledge. Several studies have found out that (Hasselgreen et al., 2004: p. 11; Ölmezer-Öztürk and Aydın, 2019: p. 614) the participants believed that they lacked the assessment skills; therefore, the participants expressed the need for more training. However, not all the studies reported low levels of perceived knowledge; for instance, the participants in Shim's (2009: p. 5) study believed that they were assessment literate.

Although some research has been carried out on LAL in the Turkish context (Büyükkarcı, 2016: p. 333; Hatipoğlu, 2010: p. 40; Mede and Atay, 2017: p. 43; Ölmezer-Öztürk and Aydın, 2019: p. 602; Şahin, 2019: p. 14), conducting more research is suggested by the researchers due to some limitations in the variety and size of the sample.

1.2. STATEMENT OF THE PROBLEM

In the teaching and learning process, the role of assessing is momentous (Lam, 2015: p. 196), and teachers are one of the most crucial elements for evaluating students (Vogt and Tzagari, 2014: p. 375). Therefore, more attention has started to be paid for the professional development of language teachers (Farhady, 2019: p. 2). However, the results

from a considerable amount of research worldwide indicated a low level of LAL (Jannati, 2015: p. 35; Semiz and Odabaş, 2016: p. 66; Xu and Brown, 2017: p. 133). Since LAL is a relatively new field of interest, there is a lack of research in the field, and in Turkish setting, as well. The studies conducted in Turkey have some limitations. The main limitation of the studies is the small number of participants (Tamerer, 2019: p. 17; Yetkin, 2015: p. 19). Some studies had larger samples, yet they were only carried out at only universities throughout Turkey (Mede and Atay, 2017: p.48; Ölmezer-Öztürk, 2018: p. 66), or only at one foundation university (Öz and Atay, 2017: p. 47; Yastıbaş and Takkaç, 2018: p. 178). Therefore, more research is definitely needed due to the aforementioned limitations. This study aims to explore the language assessment literacy of English language teachers working at state elementary, middle, and high schools.

1.3. SIGNIFICANCE OF THE STUDY

Language assessment literacy is relatively a new term, and limited research has been conducted to reveal the assessment literacy of language teachers in the Turkish context. A detailed review of the literature revealed that although some research was conducted with pre-service (Sarıyıldız, 2018: p. 73; Tamerer, 2019: p. 18) and university instructors (Mede and Atay, 2017: p. 48; Ölmezer-Öztürk and Aydın, 2019: p. 602), the studies that have been implemented to investigate LAL of teachers working at government elementary, middle or high schools in Turkish context is very limited. This study aims to close this gap and contribute to the growing research area by exploring the training levels and needs of language teachers working at the aforementioned state schools.

The findings of this study are expected to make essential contributions to the field of LAL, specifically to teacher training institutions to revise their curriculum. It will also give a chance to teachers around Turkey to reflect on their LAL, which may motivate them for professional development courses.

1.4. RESEARCH QUESTIONS

Since there is a lack of research regarding LAL in Turkish context, further research is necessary to explore the LAL of EFL teachers. To this end, this study aimed to address the following research questions:

1. What level of assessment training do the teachers of English working at state schools in Turkey believe they received during their pre-service years at university?
2. What level of assessment training do the teachers of English working at state schools in Turkey think they need?

1.5. DEFINITIONS OF TERMS

Assessment

Assessment is a process of gathering and discussing information from various sources to develop a deep understanding of what students know, understand, and can do with their knowledge as a result of educational experiences (Huba and Freed, 2000: p. 8).

Assessment Literacy (AL)

Assessment literacy is defined as "an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (Popham, 2011: p. 267).

Language Assessment Literacy (LAL)

Language assessment literacy is "the ability to design, develop and critically evaluate language tests and other assessment procedures, as well as the ability to monitor, grade and score assessments on the basis of theoretical knowledge" (Vogt and Tsagari, 2014: p.377).

1.6. LIMITATIONS OF THE STUDY

The major limitation of the current study is the sample size. The data for this study were collected from 101 English teachers who work at state elementary, middle, and high schools in Turkey.

CHAPTER II

2. LITERATURE REVIEW

2.1 ASSESSMENT

Assessment is a fundamental part of instruction since it gives feedback to teachers about the effectiveness of their teaching. Since it reveals if the teaching is effective or not in terms of students' learning or achieving the objectives of the lessons, assessment improves and also guides teachers to provide better learning environments for students. Therefore, an accurate assessment will motivate students in return and facilitate the learning environment for both students and teachers. Such an important term, assessment, is among the most common and also misunderstood terms in the educational context. It can be confusing for many people due to other related terms; testing, evaluation, and measurement. To start with, assessment is an ongoing process; whenever a student produces something, written or oral, teachers make assessments about the students' performance incidentally or intentionally (Brown, 2003: p. 4). It is also defined by Thomas et al. (2004: p. 2) as data collection for the decision-making process. Tests that include a set of questions and administered at some time intervals, on the other hand, are only one form of assessment device thanks to which we check if our assessments are reliable (Douglas, 2014: p. 1). Another confusing term, measurement, is defined as "the systematic classification of observations of student performance" (Brady and Kennedy, 2014: p. 171), and assigning numbers to students' assessment results (Miller et al., 2008: p.28). The final term, evaluation, on the other hand, involves making judgments based on the information we received through assessment (Brady and Kennedy, 2014: p. 171). To make the difference between evaluation and measurement, imagine that a student receives a six from the IELTS exam, which is measurement, and then if that grade is thought to be enough to become a pilot, it is evaluation.

Assessment plays a prominent role in the learning and teaching process (Lam, 2015: p. 1). Mertler and Campbell (2005: p. 2) point out that one of the most crucial

responsibilities of teachers is assessing the performance of students, and this responsibility, which is placed upon teachers, has increased dramatically (Fulcher, 2012: p. 113). So as to emphasize the importance of assessment, DiRanna et al. (2008) state that "assessment and instruction are two sides of the same coin" (p. 22). Therefore, teachers should regard assessment as an integral part of their instruction (Ölmezer-Öztürk and Aydın, 2019: p. 603). We assess to "gain insights into learners' level of knowledge or ability" (McNamara, 2004: p. 765), so assessment is a powerful tool for teachers as it gives them a chance to comprehend and track the flow of the teaching and make the essential changes so as to facilitate the students' learning process (Öz and Atay, 2017: p. 26). Without effective assessment techniques, teachers cannot assist students in improving themselves; therefore, teachers should keep themselves up to date. To do this, Djoub (2017) highlights the need for educators to deal with "the changing and challenging demands of the society, which necessitates more flexibility in assessment to support learning" (p.9). Assessment does not only entail outcome measurement; it also creates more opportunities to learn in L2 (Djoub, 2017: p. 9). With the emerging various assessment techniques such as alternative assessment and dynamic assessment, teachers need to be even more equipped with the necessary knowledge on assessment in order to gather and interpret the data from classroom assessment (Büyükkaracı, 2016: p. 334).

2.1.1. Alternative Assessment

Most traditional assessment is used to determine what students know or do not know about what they have learned in the lessons. Since students' capacity to recall is usually tested, they do not give students the opportunity to show their ability or perform a meaningful task. In other words, students do not use their thinking skills in authentic situations. That's why alternative assessment has become common today. Gummer and Shepardson (2001: p. 54) define alternative assessment as a non-traditional approach, which lets the students know what they know and what they can do, determine what they can understand, and evaluate students' performance. Thanks to the authenticity and meaningful context they have (Brown, 2003: p. 13), alternative assessment techniques such as portfolios or presentations let the students realize and evaluate themselves

objectively (Gözüyeşil and Tanrıseven, 2017: p. 38; Pekbay and Koray, 2020: p. 51). Not only do alternative assessment techniques let the teachers assess their students from various dimensions (MoNE, 2019: p. 6), they also make the learning environments more motivating by taking students' own learning process into consideration (Greenstein, 2010: p. 16). Some of the alternative assessment techniques are explained below.

Portfolios: Portfolios are collections of student work that is collected over time to reveal their learning process. They are not "either the arbitrary collections or observation of student's works to be filled haphazardly"; it is crucial that the portfolio collections should be purposeful, systematic, have the determined evaluation criteria, and take a period of time" (Birgin and Baki, 2007, p. 78). They offer benefits both for teachers and learners. In the first place, they facilitate the responsibility of learners in their learning process. They also guide teachers in shaping their instruction since they reveal the weaknesses and needs of the learners.

Self and Peer-Assessments: Self-assessment, in which learners make decisions about their own learning, is a functional tool for learning and measurement. As a collaborative learning technique, it helps learners to be aware of their strengths and weaknesses, which will increase their intrinsic motivation and eventually make them autonomous learners, which is one of the foundation stones of successful learning (Brown, 2003: p. 26). Peer-assessment also benefits students by making them accountable and responsible for the work they do and also assessing the others (Caspary and Boothe, 2017: p. 110); namely, students are involved in both the learning and assessment process. With precise monitoring and clear guidelines, teachers can make the utmost use of self and peer-assessment.

Observations: As teachers, whether we are aware or not, we observe our students constantly. Brown (2003) states that "experienced teachers are so good at this subliminal process of assessment that their estimates of a student's competence are often highly correlated with actual independently administered test scores" (p. 267). The type of observations as an alternative assessment involves systematic and planned procedures, which include verbal and non-verbal behaviors. Observing and recording student behaviors in real-time has some challenges; therefore, the purpose of a specific

observation has to be defined at the outset of the lesson, and teachers should have checklists to ease the data collection process.

Conferences and Interviews: Although its use is not limited to the drafts of written work, conferences as an alternative assessment technique have been a part of writing lessons for years. It is a process in which teachers as facilitators, try to ease the improvement of the written work (Brown, 2003: p. 265). Conferences include one-to-one meetings between a learner and a teacher with the aim of reviewing the student's learning generally or giving feedback on the previous assignment (Douglas, 2014: p. 74). With regard to interviews as an assessment technique, they are not activities in which students conduct interviews to gather data. Interviews involve one specific conference in which teachers talk to students about a designated assessment purpose, such as assessing oral production, ascertaining students' needs, or discovering their learning styles (Brown, 2003: p. 265).

Journals: Journals are pieces of written work that are written with no focus on the form and, therefore, without the fear of being judged. They can be about one's feelings, thoughts, ideas, or process towards an ultimate aim, and they are utilized to get detailed information about the learners' experience (Ölmezer-Öztürk, 2018: p. 49). There are numerous categories or purposes of journals such as grammar journals, language-learning logs, response to reading, self-assessment reflections (Brown, 2003: p. 260), and they provide the learners an excellent opportunity to exhibit their learning by using their own words.

In educational environments today, mainly used testing methods are usually product-oriented; that is, students are tested on how much they have improved in terms of objectives of the lesson by ignoring the process that the learners go through. At this point, sociocultural theory by Vygotsky offers a novel insight into the field of assessment by integrating assessment and instruction, which are inseparable from one another.

2.1.2. Dynamic Assessment

According to Lantolf and Poehner (2004), the term dynamic assessment (DA) can be defined as follows;

"DA integrates assessment and instruction into a seamless, unified activity aimed at promoting learner development through appropriate forms of mediation that are sensitive to the individual's (or in some cases a group's) current abilities. In essence, DA is a procedure for simultaneously assessing and promoting development that takes account of the individuals (or group's) zone of proximal development." (p.50)

As can be understood from the definition, DA assesses each student individually by paying attention to his or her pace. Due to the importance that it attaches to the guidance of teachers and the focus on the process instead of the product, DA is different from traditional assessment, which, according to Lidz and Gindis (2003: p. 99), causes dissatisfaction. To underscore the difference of DA from other assessments, Poehner (2008) also states that DA "challenges conventional views on teaching and assessment by arguing that these should not be seen as separate activities but should instead be fully integrated" (p.5).

Dynamic assessment "consists, minimally, of an initial assessment, an intervention or *mediation*, followed by a second assessment to obtain a *difference score*, reflecting the change resulting from the mediation" (Douglas, 2014: p. 79). There are two approaches of DA proposed by Lantolf and Poehner (2004: p. 239): *interventionist* and *interactionist*. The major difference lies in the way we approach the mediation process. In the interventionist approach, there is a standardized mediation process to get a quantitative point. The interactionist approach, on the other hand, is sensitive to learners ZPD (Poehner, 2008: p. 18). The integration of instruction and assessment takes its roots from Vygotsky's sociocultural theory (SCT), which emphasizes the interaction between human beings in the learning process. While observing children, Vygotsky came to realize that the problem-solving activities they were dealing with revealed nothing about their abilities, which were in the progress of developing. To him, this means that "the scope of individuals' abilities can only be revealed when various forms of support are offered as they struggle with difficult tasks" (Poehner, 2008: p.5); namely, there is a difference between learners' independent and mediated performance, and this difference is named as Zone of Proximal Development (ZPD) by Vygotsky. The social assistance provided by a knowledgeable person to help a less knowledgeable person within ZPD is called scaffolding.

2.2. ASSESSMENT LITERACY

When the assessment knowledge of teachers gained importance, the National Council on Measurement in Education, the National Education Association and The American Federation of Teachers (1990) tried to provide a framework and these institutions developed the following standards for an assessment literate teacher;

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

The increasing professional demand on teachers to assess their students precisely brought about the terms assessment literacy (AL), which was coined by Stiggins in 1991. There seems to be no consensus about the definition of the term (Fulcher, 2012: p. 115). According to Purpura (2016), assessment literacy is "teachers' understandings of assessment and assessment processes related to the identification and narrowing of learning gaps in instruction through formative assessment" (p.201). Mertler and Campbell (2005) refer to the term as "teachers' knowledge and abilities to apply assessment concepts and techniques to inform decision making and guiding practice" (p. 16). Popham (2011) also provides a definition of the term as "an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (p. 265). These decisions teachers make about the assessment tools they utilize determine the quality of education; thus, teachers' expertise in assessment has a profound impact on the quality of education (Malone, 2013: p. 330). The correlation between the

quality of education and teachers' expertise is also addressed by Stanford and Reeves (2005: p. 18); they noted that to attain a quality education, what is needed is the assessment literate teachers who are conscious about how, what, and why they are utilizing certain assessment practices. Popham (2009: p. 4) reiterates the importance of AL by pointing out that it is a 'sine qua non for a proficient educator today. Popham adds that unless assessment-literate teachers graduate from teacher education programs, this omission in the professional capabilities of prospective teachers has to be rectified by professional developers later in teachers' careers. AL is also needed since the teachers today are also valued by their students' scores on tests (Popham, 2011: p. 269). Therefore, teachers are required to be knowledgeable about high stakes tests and instruments which their students take.

Besides these, a number of reasons were suggested by Coombe et al. (2012: p. 2) to highlight the importance of the assessment literacy of teachers. To start with, since teachers spend most of their time assessing their learners, they are supposed to have theoretical and practical knowledge in assessment. Secondly, thanks to their knowledge in assessment, teachers will be able to both comprehend the results of their assessment better and also negotiate these results with their stakeholders precisely, which will make the tests more reliable, valid and transparent. Finally, teachers with a comprehensive assessment knowledge are able to integrate it into their teaching, which will also make their instruction more effective and qualified, as also suggested by Malone (2013: p. 330). Popham (2009) also highlights the significance of assessment knowledge by arguing that a lack of assessment knowledge may "cripple the quality of education" (p.4).

As can be understood from the definitions, assessment literacy entails the knowledge regarding assessment and its application in assessment practices in all fields in education. Namely, assessment literate teachers should be aware of how to teach and what to teach (Stanford and Reeves, 2005: p. 18). However, despite the rising need for assessment literacy, extensive research on the assessment literacy of teachers has reported that teachers are not prepared enough to assess and evaluate student learning (Alderson, 2005: p.4; Çalışkan and Kaşıkçı, 2010: p. 4155; Farhady, 2019: p. 6; Volante and Fazio, 2007: p. 749). According to Poehner (2008: p. 4), teachers are only armed with a repertoire

of practices such as portfolios, cloze tests, dictations, quizzes but do not have the theoretical background to lead them.

2.3 LANGUAGE ASSESSMENT LITERACY

Besides the development of standards for assessment knowledge in general education, there has also been a process of raising professionalism in the field of language assessment (Farhady, 2019: p. 4), and a new field, language assessment literacy (LAL) was born.

According to Coombe et al. (2020: p. 11), language assessment literacy is a multi-faceted concept and it is a major challenge to define it clearly. For Purpura (2016), language assessment literacy is “a broad term referring to a systematic procedure for eliciting test and non-test data (e.g., a teacher checklist of student performance) for the purpose of making inferences or claims about certain language-related characteristics of an individual” (p. 191). This new field, LAL, covers more competencies than AL does (Inbar-Lourie, 2017: p. 261). To define the term, an array of definitions have been suggested; however, since the LAL is quite a new field, a comprehensive definition of it for the groups of people teaching various levels of language in diverse settings is yet to emerge. In broad terms, Malone (2013: p. 329) uses the term, LAL, to refer to the teacher’s knowledge of testing definitions and the ability to use this knowledge in class for assessment. Although different definitions of the term exist in the literature, there appears to be some agreement that LAL has to comply with realities of the context where a specific language is being taught (Coombe et al., 2020: p. 11; Inbar-Laurie, 2017: p. 258; O’Loughlin, 2006: p. 71; Scarino, 2013: p. 311; Vogt and Tsagari, 2014: p. 392). Namely, the expected LAL of the teachers may differ depending on such factors as the proficiency level of students, the exam-orientedness of the country or the expectation of the educational institutions; therefore, attempting to increase the LAL of the diverse groups is problematic due to the uncertainty of the knowledge to help them make good decisions about the tests and test scores (Harding and Phil, 2013: p. 382; Hatipoğlu, 2015: p. 112).

The need for teachers to be language-assessment literate was stated by a number of researchers. According to Scarino (2013), thanks to LAL, teachers are able to:

“explore and evaluate their own preconceptions, understand the interpretive nature of the phenomenon of assessment and become increasingly aware of their own dynamic framework of knowledge, understanding, practices, and values, which shape their conceptualizations, interpretations, judgments, and decisions in assessment and their students’ second language learning. Through these processes, they will gradually develop self-awareness as assessors, an integral part of their language assessment literacy” (p. 311).

O’Loughlin (2006: p. 71) also states that LAL has become a vital mastery for language teachers. According to Fulcher (2012: p. 113), due to the rising responsibilities of teachers dramatically in the 21st century, and there are three major reasons for this increase. To start with, more and more tests, both externally obligated and locally developed, have begun to be administered. The second reason is that the use of language tests were expanded as a part of immigration policy as a component of citizenship tests. Finally, the use of assessment as a tool that leads to learning has gained a focus.

In order for teachers to be knowledgeable in assessment in their classroom practices, it is crucial that they are provided with sufficient teacher training in assessment (Jeong, 2013: p. 346). Since language teachers are not born testers (Jin, 2010: p. 556), they need training in assessment concepts, strategies, skills and knowledge: hence, teacher training has a crucial role (Odo, 2016: p. 40). Most of the teachers learn the basic assessment knowledge at pre-service years at teacher education departments; however, Poehner (2008: p. 4) argues that teachers are not competent in capturing learners’ abilities due to the attention devoted to assessment in most teacher education programs. Besides the lack of interest in teacher education programs, teacher trainers in the teacher training institutes are in a quandary about what to include in English Language Testing and Evaluation (ELTE) courses (Hatipoğlu, 2015: p. 112). Because of the importance attached to summative assessment (Şahin, 2019: p. 194), a large body of research indicated that teacher trainers focus more on theoretical aspects of assessment by ignoring the social

aspects of assessment (Büyükkarcı, 2016: p. 339, Hatipoğlu, 2015: p. 125; Jeong, 2013: p. 348; Jin, 2010: p. 567).

Despite the variation of LAL requirements of the teachers across the different regional contexts, there is a consensus among researchers about the inadequacy of LAL of teachers. In the last two decades, a growing body of research has investigated the LAL of the teachers or pre-service teachers, and a recent review of the literature on the LAL suggests that teachers' LAL is not sufficient (Berry et al., 2019: p. 121; Büyükkarcı, 2016: p. 338; Mede and Atay, 2017: p. 57; Muhammad and Bardakçı, 2019: p. 431; Ölmezer-Öztürk and Aydın, 2019: p. 614; Sariçoban, 2011: p. 405; Vogt and Tzagari, 2014: p. 391; Xu and Brown, 2017: p. 133).

2.4. STUDIES ON LANGUAGE ASSESSMENT LITERACY OF IN-SERVICE AND PRE-SERVICE TEACHERS ABROAD AND IN TURKEY

This chapter reviews the research conducted on LAL with pre-service and in-service teachers abroad as well as in Turkey. The LAL research conducted around the world focused on three main aspects: the effectiveness of assessment courses offered at teacher training programs at universities, training needs and training knowledge of teachers and assessment practices and beliefs of teachers. Under these aspects, research conducted on LAL, both abroad and in Turkey, are presented below, respectively.

2.4.1 Research Conducted Abroad

The effectiveness of assessment courses offered at universities has been examined by a number of researchers. To start with, O'Loughlin (2006) conducted research that lasted over 12 weeks to examine a postgraduate course on second language assessment. During 36 hours of face to face instruction, the researcher tried to develop a good understanding of important concepts in language assessment, the ability to evaluate the assessment instruments they already use critically, and develop students' capacity to adapt or design a test instrument for a particular context. Besides face to face instruction, an

online forum was offered to learners. Although they were not assessed on what they wrote in the forum, at least one contribution to the forum a week was obligatory. The researcher indicated that the first two aims were achieved broadly; however, the participants differed in their capacity and eagerness to adopt new ideas in language assessment. According to the researcher, this could result from participants' cultural and professional backgrounds as well as the input quality they got in the classes (p. 71).

Viengsang (2006) carried out a study to investigate how pre-service teachers understand assessment literacy, their previous training, their practices and problems in the practicum, and problems and needs regarding assessment. The sample was 46 pre-service practicum students in Thailand, and the data were collected through a survey questionnaire and a semi-structured interview. The findings revealed that pre-service teachers had a satisfactory level of information about concepts such as reliability and validity. They were also aware of the importance of relevance between the objectives of the course and contents of the tests; however, the study showed that the learners' knowledge about the types of the tests were unsatisfactory. Finally, it was clear both in the questionnaire and surveys that the courses on the assessment literacy were not sufficient; therefore, they were not confident about designing tests (p. 432).

Another study was run to investigate the training of university teachers with a focus on Language Testing and Assessment (LTA) courses in China by Jin (2010). The researcher made use of a survey that inquired students' thoughts about the necessity and efficiency of those courses. The survey was conducted with 86 teachers teaching LTA courses. The study showed that the necessary aspects of theory and practice of language testing were covered in the lessons. On the other hand, psychological and educational measurement and student classroom practice received considerably less attention. Also, suggestions were made to focus on under-addressed prospects of teaching contents and to create a network of testers to exchange experiences, skills, and knowledge (p. 555).

In the same vein, Lam (2015) conducted comprehensive research to examine the language assessment training programs in five language teacher education schools in Hong Kong. To gather the data, the researcher examined programs and government

documents, conducted interviews with instructors and pre-service teachers, and evaluated the assessment tasks and courses. The research yielded the following results; the programs were insufficient in developing LAL, there was a limited application of LAL in the classrooms, the training in administrating large scale and classroom-based assessment was insufficient which resulted in a gap between theory and practice, and finally, exam-oriented culture that they had made their perceptions about assessment examination oriented, as well. All in all, findings indicated an inadequate level of assessment training in Hong Kong, and the bridge between theory and practice could not be established in the assessment context (p. 169).

Deluca and Bellara (2013) carried out research to investigate adjustment between professional standards for assessment practice, pre-service teacher education policies, and course curricula, which aims at improving teacher competency in the assessment. To analyze the agreement between the components mentioned above, a moderate complexity alignment method was utilized. A comprehensive analysis of policy documents revealed both alignments and misalignments. The major trend in the alignment data was in the course content. Namely, themes such as assessment fairness, assessment processes, and measurement theory were evident in the content of the lessons, which somewhat matched the standards. Another finding was the insufficient amount of time for assessment courses; since the courses lasted typically for one semester, left teacher candidates only a little time to practice their strong theoretical knowledge. They also suggested that preservice assessment education research should be supported in order to enhance teacher learning (p. 356).

In a recent study, Davin and Heineke (2016) also conducted a comprehensive study in order to see the effectiveness of the assessment course through a practice-based approach within a course in a teacher preparation model in the state of Illinois. The focus of the study was on the preparation of pre-service students to assess their learners. All the learning experiences in their study were meticulously designed to help students prepare two performance tasks. The four critical components of this practice-based approach included the selection of core practices, instructional activities, the structure of the learning cycle, and finding an authentic setting for practice. They concluded that many

participants assured their desire to become teachers thanks to the interaction with students. Another important finding of the study was that pre-service teachers had the feeling that the situated nature of the course enabled them to understand the assets that the learners bring to the classroom by learning crucial aspects of language education (p. 921).

Another concern of the researchers regarding LAL is the teachers' knowledge and training needs and a number of studies have been implemented to reveal these. One of the earliest studies to investigate the training needs and LTA literacy of teachers was conducted by Hasselgreen et al. (2004). A questionnaire was employed to determine the training needs of teachers, test designers, and language teacher trainers. The sample consisted of 914 people, and 361 of them were teachers. The questionnaire had two parts. The first part inquired about their professional background. The second part of the questionnaire included questions under three categories as classroom-focused activities, assessment purpose, and concepts and contents of assessment. Although the majority of the items were similar in teachers' and teacher trainers' questions, teacher trainers had different items in their questionnaires, as well. The findings of the study revealed that teachers received no training in using ready tests and continuous assessment and also giving feedback. For the assessment purposes, participants stated that they received no training in placing students and awarding them certificates. Finally, for the contents and concepts of the assessment, they reported they needed training in all aspects under this group.

Fulcher (2012) also developed a survey instrument to investigate the assessment needs of language teachers with the explicit purpose of developing and writing a new text for teachers about language testing which will be available online. After piloting it with 24 teachers, some alterations were made. The questionnaire was sent online, and 278 teachers from various continents such as New Zealand, South America, North America, Europa, Far East, and Australia completed it. The results of this comprehensive study revealed that the participants were aware of different assessment needs, and they asserted that historical and social context should be taken into account while dealing with the principles and practices of assessment. Finally, the researchers recommended that

classroom-based assessment and large-scale assessment are supposed to be treated in a more balanced way (p. 113).

Another comprehensive study about the training needs and levels of teachers was conducted by Vogt and Tsagari (2014). The research was implemented in seven countries in Europe (Cyprus, Macedonia, Germany, Greece, Italy, Poland, Turkey). To collect the data, a mixed-method design was adopted. The data were collected via questionnaires (n=753) and interviews (n=63). The results showed that the teachers' expertise, which did not vary from country to country, was developed in only certain elements. The participants indicated that they gained experience after they started working through teaching materials. They also indicated a need for training for the LTA features stated in the survey. In 2017, a very similar study was implemented by Tsagari and Vogt (2017) to relate the findings from their previous study. The data were collected in three countries this time (Cyprus, Germany, and Greece), and the focus was on the qualitative findings. The findings revealed low levels of LAL. It was observed that the foreign language teachers learned about the LTA generally from colleagues or assessment materials they used, which indicated the insufficiency of the teacher education programs in equipping teachers with necessary LTA knowledge. Finally, participants tended to use the traditional ways of assessment, which means alternative assessment forms were not in the mainstream assessment practices of teachers (p. 374).

Muhammad and Bardakçı (2019) also tried to explore LAL of Iraqi teachers. Classroom Assessment Inventory prepared by Mertler (2003) was used to collect the data. 101 teachers working at secondary and preparatory schools participated in the study. They found out that although 77% of the participants stated that they had been trained enough for assessment, the results showed that they had low assessment literacy. Finally, they recommended revisions in teacher training programs and professional training courses preparation (p. 431).

Recently, Xie and Tan (2019) investigated the language assessment literacy needs of both pre-service and in-service primary school teachers in Hong Kong. The data were collected through interviews and questionnaires created by Vogt and Tsagari (2014). The

sample consisted of eleven English teachers from eleven different primary schools. As for the results, in terms of assessment design, participants expressed a need to adopt various informal and formal assessment methods. It was also reported by all the teachers that they followed a marking scheme prepared by their schools, and they thought they were proficient in using them; however, they showed quite low confidence in designing them. When they were asked about the ethical issues, the participants also expressed their concern and the challenge they had about protecting the assessment results of their students from keen parents. Another important finding was that the participant teachers were less enthusiastic about theory or concept of assessment than conducting the assessment, and issues of validity and reliability were stated to be nonessential by many. Finally, even though they were not confident in interpreting their assessment score, the participants thought themselves proficient in giving feedback to learners and parents (p. 653).

Assessment practices and beliefs of teachers have also been explored by many scholars. To start with, Campbell and Evans (2010) investigated the assessment practices of teachers who have taken LTA courses. Through a review of lesson plans of 65 pre-service teachers, they found out that pre-service teachers did not make use of necessary criteria to evaluate students' learning. Especially concepts of validity and reliability were not incorporated into their assessment in spite of the recent training they had on those concepts. The researchers indicated that completing the course did not necessarily mean that the pre-service students could show the knowledge in their practice. Item and test construction knowledge was also found inadequate, and researchers concluded that pre-service students ignored most of the content they studied in LTA courses (p. 350).

Jannati (2015) investigated the practices and viewpoints of 18 Iranian teachers of the English language through semi-structured interviews. The sample was divided into three according to participants' teaching experience. It was reported that teaching experience did not make any significant difference in teachers' perceptions about assessment. The analysis of the data also revealed that teachers within all three groups were knowledgeable of the basic concepts and assessment terminology. Furthermore,

despite being assessment literate, the participants lacked the skills to put this knowledge into practice (p. 26).

Giraldo Aristizabal (2018) is another researcher to examine the assessment practices and beliefs of English language teachers. He conducted action research to improve the LAL of English language teachers at a language institute in Colombia, and 60 EFL teachers participated in the study. The data were collected through questionnaires, interviews, and document analysis. Priori coding system was implemented to analyze the data. The analysis of the data yielded that the participants were of the opinion that tests should abide by the four basic principles, which are reliability, authenticity, validity, and positive washback. However, not all the qualities were observed to be reflected in their practices due to not having a unified approach towards rubrics designing or administrative constraints (p. 25).

To investigate the attitudes of teachers towards assessment and assessment practices, Berry et al. (2019) also conducted research with a sample from France, United Kingdom, and Spain. Classroom observations with follow-up interviews, focus group discussions and interviews were utilized. It was found out that even though teachers did not feel confident about their knowledge, they made use of a variety of assessment techniques successfully. The participants also expressed negative views towards testing and grading, and they used the terms ‘assessment’ and ‘testing’ interchangeably. Another finding of the study was that the participants tended to rely on ready-made tests or external testing companies due to the lack of confidence to create their own materials (p. 113).

In a recent study, Giraldo (2019) set out to determine the kinds of assessment instruments and beliefs of five Colombian English language teachers with the help of semi-structured interviews and reflective journals. The findings demonstrated that teachers were able to use traditional and alternative assessment techniques. They were also found to assess language as well as non-language constructs and make use of the information gained from assessing to improve students’ learning. As for the beliefs of teachers, two main beliefs were expressed. Firstly, the failure or success of students was

connected to their previous learning. Secondly, the participants asserted that teaching and assessment fed each other (p. 35).

2.4.2. Research Conducted in Turkey

An analysis of the literature revealed that there is a dearth of research conducted in Turkey about LAL, especially related to assessment courses offered at universities. In one of the earliest studies, Hatipoğlu (2010) focused on how ELTE courses were taught at Middle East Technical University, teacher education program, students' perceptions about those courses were asked. The data were collected through questionnaires and interviews, and 81 students participated in the study. The questionnaire had two parts, in the first part of the questionnaire, students had to provide personal information, and in the second part, they were supposed to answer two questions. Firstly, they had to list five topics that were studied in the lesson, and they thought they would be helpful in the future and why. Secondly, they were required to say five things they thought needed to change to make it more fruitful and state their reason. One of the major findings was that students stated that reliability, validity, and testing skills were among the most studied topics in the ELTE courses. They also expressed that they did not have enough time to practice, and highlighted the need for more testing courses; namely, there was not balance between theory and practice in the course. Crowded classrooms were another problem they mentioned by the participants because it inhibited classroom discussion. All in all, Hatipoğlu (2010) found the data she collected valuable and concluded to arrange her classroom in the light of students' suggestions (p. 40).

In her PhD thesis, Şahin (2019) conducted a comprehensive analysis of the status quo of ELTE courses offered by English Language Teacher Education programs in Turkey with a focus on LAL of the pre-service teachers. In order to collect the data, syllabi from 36 ELTE courses were examined, interviews were conducted by 21 ELTE instructors, and finally, 846 pre-service students completed questionnaires. The findings showed that only one LTA course was not sufficient to improve students' LAL effectively in terms of both theoretical and practical knowledge. The findings also revealed that traditional testing

tools were emphasized over alternative assessment, and practical aspects were not attached the same amount of importance due to time constraints. Finally, students who participated in the research expressed that although, in general, they were satisfied with the ELTE training they received, they did not find education in alternative assessment forms, assessing productive skills and formative assessment satisfactory (p. 10).

In a recent study, Kavaklı and Arslan (2019) aimed to find out how Action-Impact-Desired Outcome (AID) model affected the professional development of candidate English teachers. The AID model gives primacy to giving feedback to the most crucial issues; therefore, the value of feedback is more appreciated. Data were collected from 18 4th-year students at a state university in Turkey. For data collection, peer feedback forms are examined together with teachers' observation reports. Finally, semi-structured interviews were conducted to explore how AID model of feedback affected students' development during a 14-week course. The results obtained revealed that theory-orientedness of English language teacher education programs in Turkey failed to educate preservice teachers for real-life classrooms; hence, preservice teachers expressed the need for more practical content in ELTE courses. AID model also proved to offer more practical information about theoretical knowledge in practice. The researchers suggested that preservice teachers should take more active roles in their practicum years, which offer real-life experiences (p. 622).

In order to investigate the training needs and level of English language teachers, a number of studies have been conducted recently. Mede and Atay (2017) undertook a study to find out LAL levels of English language teachers who work at both state and foundation universities in Turkey. In their study, they collected the data by means of the questionnaire by Vgot and Tsagari (2004) and focused group interviews. The sample consisted of 350 English language teachers who graduated from ELT departments and had a minimum of 5 years of experience. The results from the questionnaire revealed that the participants lacked training in classroom-focused activities such as preparing tests and giving feedback. The research also indicated that half of the participants believed that they were competent in testing vocabulary and grammar; however, preparing skill-based tests were hard for them. Another problem that the majority of the participants had was giving

feedback on the assessment, and also over 60% of the participants expressed a need for an advanced training in receptive, productive, and integrated skills (p. 1).

Another important study regarding LAL in Turkey was conducted by Ölmezer-Öztürk (2018). In the PhD dissertation, the researcher developed and validated a scale called Language Assessment Knowledge Scale (LAKS). LAKS included 60 items regarding the assessment of four skills. A total of 542 teachers working at universities participated in the study. For the qualitative part of the study, 11 teachers were interviewed through open-ended questions to get detailed data about what teachers thought about language assessment knowledge. It was reported that the teachers were the most informed about assessing reading while they had the least literacy in assessing listening. Another important finding of the study was that pre-service and in-service education in Turkey was not sufficient; therefore, more training was required on assessing the four skills (p. 12).

In her master's thesis, Sarıyıldız (2018) investigated language assessment literacy levels and needs of preservice language teachers. The study employed a mixed design, so both quantitative and qualitative data were collected. In the quantitative part of the study, 101 4th grade students from Middle East Technical University participated, and 25 students volunteered in participating in the qualitative part. The results of the study indicated that although the participants received some education in language testing and assessment, more training was needed in these domains. It was also found out thanks to the qualitative data that the participants could not put the theoretical knowledge regarding testing and assessment into practice in the practicum. Finally, the importance of the ELTE courses for their career was also acknowledged by the participants (p. 7).

In a recent study, Tamerer (2019) was also concerned with the assessment needs and knowledge of the 4th-grade students studying at the English Language Teaching (ELT) department at a state university in Turkey. The primary data collection instrument was a questionnaire, and it was completed by 30 pre-service teachers. For the qualitative part, the researchers conducted interviews with ten volunteer participants. The research yielded that more than half of the participants were only a little experienced in developing classroom tests. Another important finding was that the majority of the participants had

no training in portfolio assessment and little training in self or peer assessment, and more than half of the participants reported the need for advanced training in informal assessment. On the other hand, the participants perceived they had a higher level of vocabulary and grammar assessment literacy. Finally, 90% of the teacher candidates expressed an interest in benefiting from online sources to improve their LTA after school (p. 4).

There were also some Turkish researchers who were concerned about the assessment practices and beliefs of pre-service or in-service English language teachers. To start with, in his master's thesis, Yetkin (2015) examined the perceptions of preservice teachers about the implementation of various assessment strategies. The sample included 30 4th-year students who were doing their practicum at elementary schools. Data collection instruments were questionnaires, which were conducted with 30 teacher candidates, and five open-ended questions, which were asked to 5 teacher candidates. The findings of the study regarding their perceptions of assessment purposes showed that classroom assessment was thought to be a means for observing student development and teacher success. Participants also assessed learners to increase their motivation and alter their teaching strategies accordingly to facilitate learning. It was also stated by the participants that the ELT program they were studying helped them increase their theoretical knowledge and their assessment literacy. Finally, the results indicated that preservice teachers relied on observation more than the other assessment techniques (p. 4).

In a current master's thesis, Dođru (2020) aimed to find out the assessment literacy needs of English language teachers working at an aviation school in Turkey. She also focused on their assessment practices, assessment beliefs, and discrepancies between their assessment applications and beliefs. The data were collected by means of both quantitative and qualitative instruments. For the quantitative part, questionnaires were completed by 32 teachers, and 7 of the participants also participated in the semi-structured interviews in the qualitative phase. The results revealed that although the participants received training in language testing and assessment, they reported that their knowledge is not sufficient. Also, the findings from the interviews showed that the assessment beliefs and practices of

the participants did not overlap. Finally, the participants had the opinion that the assessment of the aviation students ought to focus on listening and speaking (p. 3).



CHAPTER III

3. METHODOLOGY

3.1. RESEARCH DESIGN

This study employed both quantitative and qualitative data collection instruments. A mixed-methods design was utilized since “the uses of both quantitative and qualitative methods, in combination, provide a better understanding of the research problem and question than either method by itself” (Creswell, 2012, p.535), and contribute to the reliability of the findings. In the quantitative phase, a questionnaire was employed since it enables the researcher to reach a high number of participants at once, and the anonymous data collection increases the reliability of the data. Since the data are collected at one time, this research adopts a cross-sectional design in terms of data collection time frame. In the qualitative part of the study, an interview was utilized in order to get an in-depth understanding of the participants about the issues which might not be clear or mentioned in the questionnaires.

3.2. SETTING AND PARTICIPANTS

The present study was conducted online in the 2019-2020 academic year to have an understanding of language assessment training and needs of teachers working at state schools (elementary, middle, and high) in Turkey. The sample of the study consisted of 101 teachers. The participants of the study were selected through convenience sampling. Convenience sampling is selecting the participants that are suitable and convenient for the researcher (Dörnyei and Taguchi, 2009: p. 61; Mackey and Gass, 2005: p. 122), and in the current study, it was employed to be able to reach the teachers working at different parts of the country.

The sample included teachers who worked in 26 cities in Turkey. Regarding the gender distribution of the study, more females (62.4%) participated in the questionnaires

than males (37.6%). The majority of the participants had 11-15 years of experience (41.6%), which was followed by more than 15 years (25.7%), 6-10 (24.8%), 2-5 years (5.9%), and 0-1years (2%). Most of the participants graduated from English Language Teaching departments (86.1%) while 13.9% were graduates from departments. With respect to their highest qualification, 73 (78.2%) of the participants did not continue their education after the university while 16.8% had an MA degree, and 5% completed a PhD program. Regarding the schools they worked at, 49.5% worked at high schools, which was followed by middle schools (36.6%) and elementary schools (13.9%).

Table 1
Profile of the Participants (N=101)

		N	%
Gender	Male	38	37.6
	Female	63	62.4
Years of Experience	0-1	2	2
	2-5	6	5.9
	6-10	25	24.8
	11-15	42	41.6
	More than 15 years	26	25.7
Graduated BA Program	English Language Teaching (ELT)	87	86.1
	Non-ELT	14	13.9
Highest Qualification	BA	79	78.2
	MA	17	16.8
	PhD	5	5.0
School Level	Elementary School	14	13.9
	Middle School	37	36.6
	High School	50	49.5

In the qualitative part of the study, a total of 18 teachers volunteered to answer the written interview questions. Table 2 below demonstrates the profile of the interviewed participants.

Table 2
Profile of the Interviewed Participants (N=18)

		N	%
Gender	Male	10	55.6
	Female	8	44.4
Years of Experience	2-5	1	5.6
	6-10	5	27.8
	11-15	9	50
	More than 15 years	3	16.7
Highest Qualification	BA	16	88.9
	MA	2	11.1
School Level	Elementary School	4	22.2
	Middle School	8	44.4
	High School	6	33.3

3.3. DATA COLLECTION INSTRUMENTS

In this study, a mixed-methods design was employed, so both quantitative and qualitative data collection instruments were employed to be able to address the research questions of the study. The primary data collection tool was a questionnaire. The questionnaire that was employed in this study included two main parts. Part I included a total of eight questions. In this part, the participants were required to provide some demographic information such as gender, years of experience, the name of the university and program they graduated from, their highest qualification, the city they work in, the level of the institution they worked at (elementary, middle, high school), and whether they took a testing and assessment course at university or not.

In the second part, the Teachers' Questionnaire, which was developed by Vogt and Tsagari's (2014) was adopted due to the parallelism between the aims of the original study with the current one (See Appendix A for the questionnaire). This part of the questionnaire consisted of 3 sub-parts; a) classroom-focused LTA, b) purposes of testing, c) content and

concepts of LTA, and there were a total of 36 items (12, 8, 16 items respectively). In each sub-part, the participants were asked if they received any training, or need any training in the same domains. In the first part where they were asked about their current training level, they had to rate their training from a 3-point Likert scale (not at all, a little (1 - 2 days), more advanced). In the second part under each sub-part, the respondents were required to rate their need for training for the same domains from a 3-point Likert scale (none, yes, basic training, yes more advanced training).

As the second data collection instrument, interviews were conducted to complement the qualitative data in order to get more reliable, detailed, in-depth information about the perceptions and needs of the participants about their language testing and assessment knowledge. Interviews make great data collection tools since they are concerned about the respondents' experiences, identities, beliefs and more (Talmy and Richards, 2010: p. 1). In this regard, interviews enable the researcher to get more thorough insights, which would not be possible through questionnaires solely. The interview questions, which were initially used by Tsagari and Vogt (2014), were adapted in accordance with the aim and sample of the study. The interview questions that would be used were decided with the help of the thesis advisor. The questions were also translated into Turkish to make the participants feel more stress-free while answering, and therefore, provide more personal data. Five open-ended questions (See Appendix B for interview questions) were asked to have a deeper understanding of teachers working for state schools in terms of their current level, need, and implementation of language assessment tools.

3.4 DATA COLLECTION PROCEDURE AND DATA ANALYSIS

Prior to the data collection process, the researcher applied for the ethics committee approval for the research to Kocaeli University Ethics Commission. After it was approved by the committee (See Appendix C), data collection instruments were utilized to collect the data.

The data for the quantitative part of the study was collected by means of the “Teachers Questionnaire” by Tsagari and Vogt (2014). First of all, the questionnaire was created online by Google Forms. After the review and approval by the researcher’s advisor, the questionnaire was ready to be sent. The respondents of the study were selected through convenience sampling. The researcher assured them that no personal data such as their names or the names of their schools would be requested, or none of the data collected through these questionnaires would be shared by other institutions. This way, a total of 101 English teachers from various cities in Turkey filled out the questionnaire.

After collecting the quantitative data, 18 teachers volunteered to participate in the interviews. The interview questions were prepared through Google Forms and sent to volunteer teachers. The written interview consisted of five questions, and they were both in Turkish and English. The participants were informed that they could answer the questions in either language. The respondents were also requested to elaborate on their ideas to give more details while answering the questions.

In order to analyze the quantitative data, the data collected through questionnaires were entered into the Statistical Package for Social Sciences (SPSS). Initially, descriptive statistics were run to calculate percentages and frequencies of participants’ gender, years of experience, the university they graduated from, highest qualifications, the city they work in, the level of the school they were working at and whether they had a testing and assessment course at the university. Then, the second part of the questionnaire, which included the Likert scale was analyzed via SPSS in terms of means, percentages, and frequencies to find out the training levels and needs of the respondents in language assessment.

Finally, in order to analyze the qualitative data, grounded theory principles were adopted, and all the items were examined one by one in terms of their similarities and differences. After classifying the data, selective coding, which is a coding process in which “all categories are unified around a "core" category” to find out the main analytic idea in the research (Corbin and Strauss, 1990: p.14) was utilized.

The qualitative data of the study were re-coded by the same researcher three weeks after the initial coding so as to ensure intra-rater reliability. Intra-rater reliability is the measure of an individual's consistency at measuring the same phenomenon at different time intervals. Since there is not too much data to re-code, all of the interview data were coded again, and the results were compared and contrasted. It was found out that intra-rater reliability was excellent since new codes correlated with the initial ones 100%.



CHAPTER IV

4. FINDINGS

4.1 QUANTITATIVE DATA ANALYSIS

In order to examine the LAL of the teachers who work at government schools in Turkey regarding the training they received and also find out their training needs in LTA, questionnaires were conducted with 101 respondents. All the collected data were analyzed using SPSS descriptive statistics. In this section, findings from the quantitative data are presented by addressing the research questions of the study.

4.1.1 Results of Data Analysis for Research Question 1

The first research question of the study sought to explore the perceived training levels of teachers who work at state schools regarding their training levels in LTA.

In order to find out if the participants took any LTA courses before, they were asked to state it in Part I of the questionnaire. (Q7)

Table. 3

The Participants' Background in LTA Course (N=101)

		N	%
Took LTA Course	Yes	63	62.4
	No	38	37.6

Table 3 displays the results for Q7 in Part I of the questionnaire. 62.4 % of the participants indicated that they took an LTA course, while 37.6% did not have a LTA course before.

Part II of the questionnaire included three parts, and each part had two sub-parts, too. The first part of it had questions to investigate the classroom-focused LTA, and

there were six questions to find out the perceived training levels of the teachers working at state schools in Turkey.

Table. 4

Perceived Training Levels of Participants in Classroom-Focused LTA (N=101)

	N	Mean	SD	0	1	2
Preparing classroom tests	101	1.36	.687	12	41	48
Using ready-made tests from textbook packages or from other sources	101	1.36	.672	11	43	47
Giving feedback to students based on information from tests/assessments	101	1.55	.608	6	33	62
Using self-or peer-assessment	101	1.34	.667	11	45	45
Using informal, continuous, non-test type of assessment	101	1.15	.713	19	48	34
Using the European Language Portfolio, an adaptation of it or some other portfolio	101	.86	.775	38	39	24

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

Table 4 displays the perceived training levels of teachers working at state schools in terms of classroom-focused LTA. As illustrated in Table 4, the participants had the highest mean value in “giving feedback to students based on information from tests/assessments” ($M=1.55$, $SD=.608$), which indicated that the participants had the most training in this area. 61.4% of the participants indicated that they had advanced training in “giving feedback to students based on information from tests/assessments,” which was followed by a little (1-2 days) and not at all (32.7% and 5.9%, respectively). The second highest mean values were in “preparing classroom tests” ($M=1.36$, $SD=.687$) and “using ready-made tests from textbook packages or from other sources” ($M= 1.36$, $SD=.672$). While 47.5% of the participants indicated that they had advanced training in “preparing

classroom tests,” 40.6% received a little training, and 11.9% stated they had no training at all. Regarding “using ready-made tests from textbook packages or from other sources,” 46.5% of the participants received advanced training, indicating that almost half of the participants thought that they had advanced training in this component. It was followed by a little (42.6%) and not at all (10.9%). With a very close mean value of 1.34, “using self-or peer-assessment” had the third highest mean value ($SD=.667$). 44.6% of the respondents stated that they had a little or advanced training in this field while 10.9% received no training at all. “Using an informal, continuous, non-test type of assessment” had a relatively low mean score ($M=1.15$, $SD=.713$). In this area, most of the participants (47.5%) received a little training, which was followed by advanced training (33.7%), and no training (18.8%). The lowest mean value was found for “using the European Language Portfolio, an adaptation of it or some other portfolio” ($M=.86$, $SD=.775$) indicating that the participants received the least training in this component. With a slight difference in their percentages, the majority of the participants received a little and no training (38.6%, 37.6%, respectively), while 23.8% had advanced training in this area. As demonstrated in Table 4, even though the participants varied in their training level across different components, their perceived training levels are basic in this domain.

The second part of Part II consisted of four questions in or to reveal the perceived training levels of the teachers working at state schools in terms of purposes of testing.

Table. 5

Perceived Training Levels of Participants in Purposes of Testing (N=101)

	N	Mean	SD	0	1	2
Giving grades	101	1.47	.672	10	34	57
Finding out what needs to be taught/learned	101	1.49	.642	8	36	57
Placing students onto courses, programs etc	101	1.12	.739	22	45	34

Awarding students final certificates (from school/program: local, regional or national level)	101	.98	.735	28	47	26
---	-----	-----	------	----	----	----

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

As Table 5 indicates, teachers who work at state schools in Turkey appear to have the highest mean score in “finding out what needs to be taught/learned” ($M=1.49$, $SD=.642$), suggesting that the teachers working at state schools in Turkey are of the opinion that they received the highest training in identifying their learners’ needs. More than half of the participants (57.4%) perceived that they had advanced training in this field, while 35.6% of the participants stated that they had a little training, and 7.9% had no training at all, which indicates that more importance is attached to training the participants in these components. The second highest mean value ($M= 1.47$, $SD=.672$) was found for “giving grades” because the majority of the participants (56.4%) indicated the training they had in this field was advanced. It was followed by a little (33.7%) and not at all (9.9%). “Placing students onto courses, programs, etc.” had the third highest mean value ($M=1.12$, $SD=.739$). The majority of the participants stated that they had a little training (44.6%). 33.7% stated that they had advanced training in “placing students onto courses, programs, etc.” while 21.8% indicated that they had no training at all. The lowest mean score ($M=.98$, $SD=.735$) was observed for “awarding students final certificates (from school/program: local, regional or national level)” since the majority of the participants reported that they received a little training or no training at all in this field (46.5%, 27.7%, respectively), demonstrating that the majority of the participants lack advanced training in this component since only about a quarter of the participants (25.7%) indicated that they received advanced training in this field.

In the final section in Part II, the data were collected regarding the perceived training levels of the English teachers working at state schools in terms of content and concepts of LTA, and Table 6 demonstrates the findings from this domain.

Table. 6

Perceived Training Levels of Participants in Content and Concepts of LTA (N=101)

	N	Mean	SD	0	1	2
1. Testing and assessing						
Receptive skills (reading/listening)	101	1.39	.648	9	44	48
Productive skills (speaking/writing)	101	1.39	.663	10	42	49
Microlinguistic aspects (grammar/vocabulary)	101	1.41	.710	13	34	54
Integrated language skills	101	1.28	.709	15	43	43
Aspects of culture	101	1.08	.744	24	45	32
2. Establishing reliability of tests/assessment	101	1.32	.692	13	43	45
3. Establishing validity of tests/assessment	101	1.33	.694	13	42	46
4. Using statistics to study the quality of tests/assessment	101	1.21	.697	16	48	37

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

The results, as can be seen in Table 6, demonstrated that the highest mean score was found for “testing and assessing microlinguistic aspects (grammar/vocabulary)” ($M=1.41$, $SD= .710$) More than half of the participants (53.5%) reported that they had advanced training in “testing and assessing microlinguistic aspects and 33.7% stated that they received a little training while 12.9% of them indicated that they received no training at all. It was followed by testing and assessing receptive skills (reading/listening) and productive skills (speaking/writing) with the same mean value ($M=1.39$, $SD= .648$ and $.663$, respectively), suggesting that perceived overall training of the participants are the same in this component since in both components the majority of the participants perceived an advanced training. For both receptive and productive skills (with the same mean scores), nearly half of the participants (47.5%, 48.5%, respectively) stated that they had advanced training in this field. It was followed by a little training in both fields (43.6%, 41.6%, respectively) and not at all (8.9%, 9.9%, respectively). The third highest

mean score was found for “establishing the validity of tests/assessment” ($M=1.33$, $SD=.694$). 45.5% of the participants indicated that they received advanced training in establishing the validity of tests/assessment, and 41.6% of them indicated that they had a little training while 12.9% of the respondents stated that they had no training in this field at all. With a slightly lower mean score, it was followed by “establishing the reliability of tests/assessment” ($M=1.32$, $SD=.692$). The majority of the participants reported that they had advanced training (44.6%), which was followed by a little training (42.6%) and not, all (12.9%). The following highest mean value belonged to “teaching and assessing integrated language skills” ($M=1.28$, $SD=.709$). In this field, the same number of participants reported advanced and a little training (42.6%), while 14.9% of them stated that they received no training at all. For another aspect under this domain, “using statistics to study the quality of tests/assessment,” the mean value was 1.21 ($SD=.697$). Nearly half of the participants (47.5%) indicated that they received a little training, which was followed by advanced training and no training at all (36.6%, 15.8%, respectively). The lowest mean score was found in “teaching and assessing aspects of culture” ($M=1.08$, $SD=.744$). 44.6% of the participants reported that they had little training, and 31.7% of them received advanced training. 23.8% of them expressed that they had no training at all in this field, suggesting that almost a quarter of the participants thought that they had no training in this component.

Finally, the responses of teachers who work at state schools to 18 items were analyzed in terms of mean values in all three domains of LTA. Table 7 illustrates the findings.

Table. 7

Perceived Training Levels of Participants in All Domains of LTA (N=101)

	Mean
Classroom-focused LTA	1.27
Purposes of testing	1.27
Contents and concepts of LTA	1.30
Overall	1.28

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

As shown in Table 7, English teachers who work at state schools had the highest mean value in “contents and concepts of LTA” ($M=1.30$). The lowest mean values were found for classroom-focused LTA and purposes of testing ($M=1.27$). Finally, overall mean values for all three domains were found to be 1.28. As can be seen in the table, even though the participants’ perceived training levels were the highest for the content and concepts of LTA, the overall mean scores of the other domains are only slightly lower.

4.1.2 Results of Data Analysis for Research Question 2

The second research question in this study sought to explore the perceived training needs of English teachers who work at state schools in language testing and assessment (LTA).

Part II of the questionnaire also included questions to reveal the perceived training needs of the teachers who work at state schools in Turkey. The classroom-focused LTA, purposes of testing, and content and concepts of LTA were explored in 3 sub-parts of Part II. In the first part, teachers’ perceived training needs in classroom-focused LTA were investigated through six questions.

Table. 8

Perceived Training Needs of Participants in Classroom-Focused LTA (N=101)

	N	Mean	SD	0	1	2
Preparing classroom tests	101	.88	.816	40	33	28
Using ready-made tests from textbook packages or from other sources	101	.70	.794	51	29	21
Giving feedback to students based on information from tests/assessments	101	.85	.792	40	36	25
Using self-or peer-assessment	101	.99	.794	32	38	31

Using informal, continuous, non-test type of assessment	101	.95	.753	31	44	26
Using the European Language Portfolio, an adaptation of it or some other portfolio	101	1.12	.765	24	41	36

0= None, 1= Yes, basic training, 2= Yes, more advanced training

Table 8 illustrates the perceived training needs of the teachers who work at state schools in Turkey. The highest mean value in this domain was observed in “using the European Language Portfolio, an adaptation of it or some other portfolio” ($M=1.12$, $SD=.765$), suggesting that the participants needed the most training in this component. 40.6% of the participants reported that they needed basic training, and 35.6% of them indicated that they needed more advanced training. 23.8% of the participants stated that they needed no training in this field. The second highest mean score was found for “using self-or peer-assessment” ($M=.99$, $SD=.794$). The majority of the participants (37.6%) reported that they needed basic training in this field, and 30.7% of them indicated an advanced need for using peer-assessment, while 31.7% of the participants stated that they needed no training in this field. It was followed by “using an informal, continuous, non-test type of assessment” ($M=.95$, $SD=.753$). In this field, 43.6% of the participants expressed a need for basic training, while 30.7% of them reported no need for training. About a quarter of the participants (25.7%) stated that they needed advanced training in this field. Preparing classroom tests had the next highest mean score ($M=.88$, $SD=.816$). 39.6% of the respondents stated that they needed no training in this field, while 32.7% of the reported a little need for training, and 27.7% expressed an advanced need in this field. With a slightly lower mean value, it was followed by “giving feedback to students based on information from tests/assessments” ($M=.85$, $SD=.792$). 39.6% of the participants stated that they needed no training in this field. While 35.6% of them expressed a need for basic training, almost a quarter of the participants (24.8%) reported the need for advanced training in giving feedback to students. The lowest mean score was found for “using ready-made tests from textbook packages or from other sources” ($M=.70$, $SD=.794$) as

more than half of the participants (51%) indicated no need for further education. 29% of the participants reported a need for basic training, while 21% of them reported a need for more advanced training, which suggests that the majority of the participants believed that they were proficient in using ready-made tests.

The second part in part II investigated the perceived needs of teachers who work at state school in terms of purposes of testing, and the findings are illustrated in Table 9.

Table. 9

Perceived Training Needs of Participants in Purposes of Testing (N=101)

	N	Mean	SD	0	1	2
Giving grades	101	.77	.768	45	34	22
Finding out what needs to be taught/learned	101	.86	.800	40	35	26
Placing students onto courses, programs etc	101	1.01	.755	28	44	29
Awarding students final certificates (from school/program: local, regional or national level)	101	1.12	.711	20	49	32

0= None, 1= Yes, basic training, 2= Yes, more advanced training

As illustrated in Table 9, the highest mean score was found for “awarding students final certificates (from school/program: local, regional or national level)” ($M=1.12$, $SD=.711$), meaning that this is the component in which the teachers who work at state schools needed the most training. Almost half of the participants (48.5%) expressed a need for basic training. 31.7% of them stated that they needed advanced training, while 19.8% reported no need for this domain. The second highest mean value was observed in “placing students onto courses, programs, etc.” ($M= 1.01$, $SD= .755$). For this field, 43.6% of the participants indicated a need for basic training, and 28.7% of them stated that they needed more advanced training, while with a slightly lower percentage, 27.7% of them expressed

no need for training in this field. It was followed by “finding out what needs to be taught/learned” ($M=.86$, $SD=.800$). 39.6% of the participants reported no further need in this field, and 34.7% of them expressed a need for basic training. About a quarter of the participants (25.7%) stated that they needed advanced training in finding out what needs to be taught or learned. The lowest mean score was observed in “giving grades” ($M=.77$, $SD=.768$), which indicates that the overall perceived needs of the participants are less than basic. Nearly half of the participants (44.6%) stated that they needed no training in this field, while 33.7% of them expressed a need for basic training, and a need for advanced training was expressed by 21.8% of the participants.

The last part in part II included 8 items to investigate the perceived training needs of English teachers who work at state schools in the content and concepts of LTA.

Table 10

Perceived Training Needs of Participants in Content and Concepts of LTA (N=101)

	N	Mean	SD	0	1	2
1. Testing and assessing						
Receptive skills (reading/listening)	101	.90	.819	39	33	29
Productive skills (speaking/writing)	101	.93	.816	37	34	30
Microlinguistic aspects (grammar/vocabulary)	101	.82	.817	44	31	26
Integrated language skills	101	.95	.792	34	38	29
Aspects of culture	101	1.06	.732	24	47	30
2. Establishing reliability of tests/assessment	101	1.00	.800	32	37	32
3. Establishing validity of tests/assessment	101	.97	.768	31	42	28
4. Using statistics to study the quality of tests/assessment	101	1.04	.774	28	41	32

0= None, 1= Yes, basic training, 2= Yes, more advanced training

The findings of the training needs of the participant teachers who work at state schools are presented in Table 10. This highest mean score in this domain was in “testing and assessing aspects of culture” ($M=1.06$, $SD=.732$), suggesting that the participants in this study thought that they needed the most training in the component. Nearly half of the participants (46.5%) reported a need for basic training in this field. While 29.7% of them expressed a need for more advanced training, 23.8% of the reported no need for further training. It was followed by “using statistics to study the quality of tests/assessment” ($M=1.04$, $SD=.774$) with a slightly lower mean value. 40.6% of the participants stated that they needed basic training in this field, and 31.7% of them expressed a need for more advanced training, while 27.7% reported no need for training in using statistics to study the quality of tests or assessment. The third highest mean value was in “establishing the reliability of tests/assessment” ($M=1.00$, $SD=.800$). In this field, the majority of the participants (36.6%) reported a need for basic training, while the percentages of the respondents who indicated a need for more advanced training (31.7%) and no training (31.7%) were observed the same. It was followed by “establishing the validity of tests/assessment” ($M=.97$, $SD=.768$). The majority of the participants (41.6%) expressed a need for basic training in this field. 30.7% of them reported no further training, while 27.7% of the participants stated that they needed more advanced training in establishing the validity of tests/assessment. “Testing and assessing the integrated language skills” follows with a mean value of .95 ($SD=.792$). Most of the participants (37.6%) reported a need for basic training in this field, while 33.7% of them did not indicate a need for further training, and 28.7% of them expressed a need for more advanced training in the field. It was followed by “testing and assessing productive skills” ($M=.93$, $SD=.816$). In this field, the need for no training and basic training yielded close percentages (36.6%, 33.7%, respectively), while 29.7% of the participants stated that they needed more advanced training. The second lowest mean value was found for “testing and assessing receptive skills” ($M=.90$, $SD=.819$). The majority of the participants indicated that they did not need further training in this field, while 32.7% of them expressed a need for basic training, and 28.7% reported that they needed more advanced training. The lowest mean score was found in “testing and assessing microlinguistic aspects” ($M=.82$, $SD=.817$), which demonstrated that in this domain, the participants felt the most proficient in testing and

assessing grammar and vocabulary. 43.6% of the participants stated that they did not need any further training, while 30.7% of them reported a need for basic training. About a quarter of the participants (25.7%) indicated that they needed more advanced training.

Finally, responses of teachers who work at state schools to these 18 items in the questionnaire were analyzed to reveal the perceived training needs of the teachers who work at state schools in all the domains of LTA, and the results are displayed in Table 11.

Table 11

Perceived Training Needs of the Participants in All Domains of LTA (N=101)

	Mean
Classroom-focused LTA	.91
Purposes of testing	.94
Content and concepts of LTA	.95
Overall	.93

0= None, 1= Yes, basic training, 2= Yes, more advanced training

As the Table 11 illustrates the findings for the stated needs of teachers who work at state schools in all three domains of LTA, the overall mean value of the 18 items in the questionnaire was found to be 0.93, suggesting that the overall perceived needs of the participants are less than basic. The highest mean value was found in the domain, “content and concepts of LTA” ($M=.95$), which was followed by “purposes of testing” with a slightly lower mean value ($M=.94$). The lowest mean score was observed in the classroom-focused LTA domain ($M=.91$).

4.2 QUALITATIVE DATA ANALYSIS

In order to complement the quantitative data, written interviews were conducted to get more reliable, detailed, in-depth information about the perceptions and needs of the participants about their language testing and assessment knowledge. 18 English teachers who work at state schools in Turkey participated in the interview part, which included 5

open-ended questions. In order to maintain anonymity, rather than using their names, the participants were mentioned as T (teacher), followed by a number. All the data collected via written interviews were analyzed item by item by means of selective coding.

Interview question 1: Did your school experience course you took at the university give you the chance to put what you learned related to language testing and assessment into practice and contribute to your professional development? Please explain.

Many of the participants appreciated the help of school experience courses and practicum experience they had in the final year of the university with some limitations. When the responses were analyzed, some common points were realized. The findings revealed that although teachers learn basics at the universities, they keep learning after they start teaching, and the first theme emerged as “*learning by doing*”. In this vein, one individual said:

“Actually, the school experience course we took at the university surely helped us; however, when we found ourselves in a real teaching atmosphere, we had different experiences through learning by doing. Teaching as a profession is not static; therefore, we may get new experiences every day or even every class. All in all, the education we had at university establishes our background, and this process continues when we start teaching.” (T4)

Another theme which was observed through the analysis of interviews was “*the lack of involvement of preservice teachers in assessment process*”. It was found out that the attitude of the advisory teachers at practicum schools may have adverse impact on the development of preservice teachers as one of the participants made the following comment:

“Advisory teacher at the practicum school never involved us in testing and assessment process; therefore, we did not have any change to put what learned at the university into practice.”(T18)

This comment is significant as it reveals that not all the advisory teachers at practicum schools are cooperative, which hinders the professional development of

preservice teachers. The final theme which was identified in the interview data was *“limited time at practicum schools”*. It was found out that the preservice teachers did not benefit from school experience courses due to insufficient time they spent practicing at practicum schools. Regarding this theme, a participant made this comment:

“School experience course was not very helpful. Since the amount of time we participated in the lessons was very limited, we did not have much time to practice what we learned.”(T5)

However, not all the problems stemmed from issues related to limited time or advisory teachers at practicum schools, moreover, one interviewee narrated a negative experience resulted from his supervisor for school experience course at the university by making the following comment:

“... our supervisor teacher for school experience course was someone out the teacher education field (his undergraduate degree was in translation), so the only thing he suggested to us is to teach in English. Nothing about what to teach, how to teach or what kind of strategies to use in classroom discourse in terms of motivating students, promoting their involvement or maintaining discipline, etc. (T15)

As can be understood from the comment above, not all of the professors at universities are qualified enough to equip preservice teachers with fundamental assessment skills. On the other hand, some of the interviewees were totally content with the school experience course they took at universities and the way it enabled them to put what they learned into practice as one interviewee put it:

“In the school experience course we had at the university, we had a chance to practice what we learned at the LTA course. In the practicum school, in which we worked together with teachers and students, we had a chance to observe and reflect our observations, and we had experience in teaching skills. We participated in teaching activities actively in as many different levels as possible and gained experience in LTA.” (T8)

Although the overall attitude of the interviewees towards LTA courses was positive, one individual made the striking comment below about the importance of testing and assessment courses:

“The teaching atmosphere of the newly appointed teachers at schools where they fulfill their compulsory service requires more prior skills. Teachers from totally different parts of the country, sometimes even with a different mother tongue, might need more basic skills such as classroom management, human relations, etc. Therefore, personally, since the teachers are not in a position to do testing and assessment, they should be given such training as effective classroom communication and leadership skills.” (T11)

The comment above highlights the importance of the schools or cities where teachers work at on the opportunities to use the skills they have, suggesting that not all the learned skills could be practiced due to the diversity of cultures or languages within the country.

Interview question 2: Did you feel prepared enough to assess your students’ language learning when you finished university? If not, what kind of difficulties did you have in terms of assessing language learning, and how did you overcome them?

The majority of the interviewees indicated that they did not feel prepared enough to assess their students’ language learning right after they finished university, and the participants were asked to mention the strategies they made use of to tackle with this insufficiency. Analysis of how they overcame it yielded some common points. To start with, *“learning from colleagues”* was voiced by many of the participants, to illustrate, T7 made the following comment regarding this theme:

“No. I felt ready in assessing grammar, reading, and a little bit of writing; however, I did not feel ready for listening and speaking assessment because I did not do any assessment in those two skills. When I had difficulty in testing and assessment, I turned to the Internet or my colleagues for help.” (T7)

It is clear from the comment above that novice teachers improve themselves thanks to the experienced teachers who work at the same schools. The second theme that was

identified was “*learning in the course of time*”. Besides the help of colleagues, some of the participants stated that as the time past they also gained experienced, and the following comment was made related to this issue:

“No, when I first started working, I tried to teach everything in detail. But, in the course of time, I learned to present the information according to the levels of my learners, so while assessing them, I focused on the learning environment and goals of teaching rather than everything related to the topic.” (T12)

It could be understood from the comment above that, novice teachers may fail to prioritize the goals of the lesson and teach accordingly. Another theme that emerged through interviews was “*being away from realities of the country*”. Some participants stated that the education the preservice teachers receive at universities do not overlap with the facts of the country. Regarding this theme, the following comment was made:

“No, I did not feel ready because the education at given at the universities does not correspond to the realities of the country. The school to which I was appointed for my compulsory service had its all internal dynamics where we could not make an assessment through commonplace knowledge” (T11)

The comment above is significant since it demonstrates that theoretical knowledge cannot be applied at some schools, and this situation is even more difficult for novice teachers since they are not trained for real life situations at universities. In addition to these themes, one of the interviewees mentioned the insufficiency of the ELTE courses offered at universities due to the focus on pen and paper tests rather than improving speaking proficiency by making the following comment:

“Whether the students filled out the gaps correctly, or they answered multiple-choice questions correctly was not sufficient for me to assess their language learning. The ELTE course I took did not have the logic that language is for learning and speaking. Moreover, the schools we work at do not have a curriculum to help us improve these skills; therefore, we are pushed to test our learners through gap-filling.” (T10)

The comment by T10 also suggests that teachers who work at state schools are not required to boost speaking skills of the students; therefore, even though they might want to develop the students' oral skills, they do not have a chance. Another interviewee, when asked if she was ready to assess their learners' language learning after graduating from university, stated that how he felt about his ability in LTA changed upon starting to work at a different school. She made the following comment:

“I first started to work at a regular high school. When I started to work there, I taught my lessons quite easily. Sometime later, I started to work at Anatolian high school. There, I realized that I was not done with learning since my students had a good background in English and expected more from me; therefore, I had to prepare for my lessons, which helped me to improve myself, too.” (T4)

The comment above indicates that the language education at Anatolian high schools was better than regular high schools. On the other hand, some interviewees reported a strong sense of efficacy in LTA after graduating from universities. In this regard, one of the interviewees made the following comment:

“When I finished university, I did not feel that I was inadequate. Since I knew which skills they needed to have in the levels I taught, I did not have difficulty in assessing them. Moreover, although, at that time, we were not expected to assess the learners in four skills, I assessed them in listening and writing. Sometime later, the Ministry of National Education integrated a production exam on the evaluation scale.” (T14)

Interview question 3: After you finished university, have you received any training in language testing and assessment (LTA)? If yes, what was the focus of this training? If no, in which domain would you like to receive one?

When the interviewees were asked if they participated in any training in LTA after university education, only 2 of them gave a positive answer. As for the details of the course, one of the interviewees stated that he participated in a TESOL program. The other one mentioned that she completed a training program which is organized by the Ministry of National Education, and it included a variety of topics. The training lasted for one week, and according to the interviewee, it was a lot of fun.

The majority of the interviewees did not receive any training after the university. When they were asked what kind of training they wanted to have, a number of topics were identified. The types of trainings they wished to have included “assessing writing, assessing listening and speaking, using SPSS in testing, test preparation techniques, alternative assessment and preparing and using rubrics”. This outcome is significant since it indicates how varied the participants’ needs were. The types of training the participants wanted to have also revealed that the participants were not only interested in traditional assessment, they also wished to improve themselves in alternative assessment and test preparation procedures.

Regarding training wishes of the participants, especially one interviewee made the following comment about using rubrics by underscoring the importance of it for all the teachers:

“If I could, I would love to learn how to use rubrics to assign grades. Actually, I would like to learn how to grade the students objectively since I tend to give 5 to the best student in my classroom even though we use evaluating rubrics. Maybe, that student would get a 3 if she were in a different class. Therefore, I think all the teachers should be provided with training to overcome this tendency.” (T18)

This comment is noteworthy since it displays the significance of objectivity for some teachers who work at state schools. On the other hand, surprisingly, some interviewees who did not receive any training in LTA reported no desire to have any kind of training.

Interview question 4: How do you usually assess your student’s language knowledge in the classroom? What kind of assessment instruments do you use?

In response to this questions, all of the participants reported some instruments they made use of while assessing their learners. The instruments they stated that they used were analyzed and categorized. The responses of the participants revealed that both traditional and alternative assessment tools were utilized, so the tools that were mentioned were listed under these two categories, and they were presented in Table 12 below.

Table. 12

The assessment tools used by the interviewees

Traditional assessment	Alternative assessment
Written exams	Portfolios
Vocabulary quizzes	Peer-assessment
Question and answer	Presentations
Singing songs	Self-assessment
Weekly writing assignments	Writing diaries
Vocabulary matching activities	

When the interviewees were asked to state which instruments they make use of when assessing their learners, it was realized that they benefitted from a wide range of both traditional and alternative assessment instruments. A recurrent instrument in the interviews was written exam, which has to be administered twice a semester. This finding is not surprising since conducting a written exam is compulsory at schools in Turkey. Regarding the written exams, some of the participants stated that they aimed to assess the four skills in the written exams they conducted, which suggests that improving productive skills were also the aim of some teachers. While the majority of the participants did not specify what was included in their written exams, the exams they prepared were mainly composed of activities to test reading comprehension and lexical knowledge. In this regard, one individual said:

“First of all, I administer two written exams in one semester. In those exams, reading comprehension, and vocabulary knowledge are tested through vocabulary matching, true-false items in an open-ended format, or through multiple-choice items.”
(T12)

Moreover, some of the interviewees reported that they utilized speaking activities in classes for various purposes such as motivating their students and boosting their communicative abilities. Regarding this issue, one of the participants stated:

“Question and answer activities give me a helping hand to warm up and motivate the students. They also offer me a chance to assess and improve their speaking skill.” (T9)

This comment is important as it reveals that speaking activities are not seen the ultimate goals, they are also used to raise the motivation of the learners. Besides traditional assessment tools, a number of alternative assessment tools were voiced in the interviews, too. The alternative assessment tools the participants mentioned included portfolios, peer-assessment, presentations, self-assessment, and writing diaries, and the following comments were made regarding interviewees’ choice of instruments:

“... toward the end of the unit, I want my students to prepare presentations related to the topic of the day and present them to their classmates. Firstly, I make them reflect on their presentations themselves, and then the presentations are assessed by the presenter’s peers. I want them to put everything we did during the semester in their portfolio and collect them at the end of the semester.” (T12)

“... besides traditional ways, I assign my students group projects and assess them individually in their groups. We have assessment charts in which they assess their own group in terms of their performance. This type of assessment also helps avoid piggybacking of some lazy students.”(T15)

The above comments indicate that some of the participants were well-informed about combining various alternative assessment tools together and also strengths of them in involving students in lessons.

Interview question 5: Do you know about more recent Language Testing and Assessment (LTA) methods, e.g., portfolio assessment, self- or peer-assessment? Have you ever tried them? How often do you use them?

All of the participants stated that they had heard the recent alternative assessment tools. When they were asked whether they had used any of them so far, all the interviewees but 3 indicated that they had used it with different frequencies. While some of the respondents said that they made use of a type of alternative assessment tool every week, the use of them was minimal for some other interviewees due to some reasons. The most

common reason indicated by the participants for the disusing alternative assessment tools was “insufficient time” in their curriculum. In this sense, the following comments were made:

“I have information about the recent LTA methods, but I have never used them. In 3 hours a week, we try to catch up with the curriculum, improve the students’ test-taking skills, have them memorize vocabulary items, etc.; therefore, these methods are ignored.” (T10)

“I know them, but I have never used them. We have limited time for the courses; 5 hours for 9th graders and 2 hours for the others. It may even be impossible to complete all the subjects in the curriculum in that much time.” (T17)

The comments above show that the curriculum is so tight that there is no room for benefitting from alternative assessment tools and assess the learning process of the students. Besides insufficient time for covering the subjects in the curriculum, one individual highlighted the importance of the school level where teachers work in terms of applicability of alternative assessment tools at elementary schools by making the following comment:

“I have heard and used peer-assessment and self-assessment before. When I worked at a high school, I used to use them often. However, at the elementary school where I now work, they are not easily applicable tools, so I cannot use them.” (T1)

The comment above suggests that peer- or self-assessment techniques are not very common at elementary schools. Furthermore, when one interviewee was asked if he was knowledgeable about alternative assessment tools, he stated that he did not have information by blaming the Ministry of National Education with the following comments:

“... but I do not have information about portfolio assessment. Although the Ministry of National Education featured portfolios, as far as I know, it has not held training seminars or training sessions on that. Maybe they did have the training, but apparently, the training was far from reaching the whole country.” (T18)

The comment above reveals that the Ministry of National Education fails to offer in-service training on alternative assessment tools throughout the country even though teachers are expected to make use of them.



CHAPTER V

5. DISCUSSION, CONCLUSION and SUGGESTIONS

In this chapter, an overview of the findings will be provided, and the findings will be discussed in relation to the studies in the literature. It will be concluded with the pedagogical implications, limitations of the study, and suggestions for future research.

5.1. THE OVERVIEW OF THE STUDY

The study sought to explore the language assessment literacy of English language teachers working at state elementary, middle, and high schools in Turkey, regarding their perceived training levels and needs under such domains as classroom-focused LTA, purposes of testing, and content and concepts of LTA. The participants of the study were 101 English teachers who worked in various cities throughout the country. The data were collected through both quantitative and qualitative instruments. For the quantitative data collection, the “Teachers Questionnaire” by Vogt and Tsagari (2014) was adapted and conducted. The questionnaire included three sub-parts and a total of 36 items. In the qualitative phase of the study, written interviews, which included 5 open-ended questions, were held with 18 of the teachers who filled out the questionnaires. Both quantitative and qualitative data were collected through online forms.

The data collected through questionnaires were analyzed through the Statistical Package for Social Sciences (SPSS). Descriptive statistics were run to calculate percentages, means, and frequencies of the items. In order to analyze the interview responses, grounded theory principles were adopted, and all the items were examined one by one in terms of their similarities and differences. After classifying the data, selective coding was utilized.

5.2. DISCUSSION OF THE FINDINGS

The findings in this study were examined under four sections; classroom-focused LTA, purposes of testing, and content and concepts of LTA and overall comparison of the three domains. The findings from the study were discussed in relation to the research questions of the study.

The first research question sought to investigate the perceived training levels of the English teachers who worked at state schools in classroom-focused LTA. Firstly, it intended to determine the perceived level of training in this domain through six items. The overall mean score of these six items was found to be 1.27. It can be concluded from this result that the perceived training levels of teachers in classroom-focused LTA are not sufficient. This finding is consistent with those of Mede and Atay (2017: p. 49), Sarıyıldız (2018: p. 111), Tamerer (2019: p. 35), and Vogt and Tsagari (2014: p. 384), who also uncovered inadequate levels of training in this domain in their research. On the other hand, when the components of this domain are compared to the ones in the literature, some consistencies, as well as differences, were detected in the findings. To illustrate, the highest mean value for perceived training levels in the current study was found for “giving feedback to students,” for which 61.7% of the participants stated that they received advanced training, which means that the participants thought that they had the highest training in this component. This finding supports the previous study by Xie and Tan (2019: p. 663), in which the participants reported proficiency in giving feedback to learners. However, it differs from the studies by Mede and Atay (2017: p. 49), Tamerer (2019: p. 23), and Vogt and Tsagari (2014: p. 385), in which the percentage of the participants who indicated that they received advanced training in this component was 37%, 43.3%, and 40.5%, respectively. On the other hand, the present findings of some other components under this domain seem to be consistent with the research mentioned above; especially, having the lowest mean score in the present study, using portfolios, was also studied by the minority of the participants in the aforementioned studies. This finding indicates that teaching how to use portfolios is not paid sufficient attention by teacher training institutions both in Turkey and in many European countries. It is also crucial to note that

the majority of the respondents reported that they received little or advanced training in self- or peer-assessment (44.6% each). This finding can be said to contradict the previous finding in the literature by Şahin (2019: p. 193), who found out in her analysis of ELTE courses in Turkey that traditional testing tools were emphasized over alternative assessment tools. Based on these findings, it is evident that LAL of teachers in classroom-focused LTA who work at state schools in Turkey is insufficient.

LAL in the “purposes of testing” was the second focus of the current study. The findings from this part of the questionnaire uncovered that the highest perceived training level was in “finding out what needs to be taught/learned” whereas “awarding final certificates” was found to have the lowest mean value. These findings of the current study are consistent with those of Sarıyıldız’s (2019: p. 91) and Tamerer’s (2019: p. 24), who also reported the same components as having the highest and lowest mean scores, respectively. However, the findings of awarding final certificates in this study have not confirmed the previous research by Vogt and Tsagari (2014: p. 383) regarding the percentage of the participants who received no training in this component. That is, while over half the respondents in their study (58.2%) indicated that they had no training, the percentage of the participants in the current study who reported no training in this component was a lot lower, only 25.7%. Despite the fact that there was some inconsistency in the distribution of the ratings within the component, it was rather expected that awarding final certificates had the lowest mean value in this domain. The possible explanation of this might be that awarding certificates is not common in the Turkish education system; hence, it is not a concern of the teacher training institutions to train preservice teachers in this component. On the other hand, surprisingly, in “placing students onto courses and programs, etc.”, another component in this domain, which is not also a commonplace practice at state schools or focus at training institutions in Turkey, relatively more participants rated their training as advanced.

The next part of the questionnaire examined the perceptions of the participants regarding their training levels in “the content and concepts of LTA.” The results indicated that more participants tended to rate their training in this domain as “advanced” than as “a little”; therefore, the highest mean score (1.30) was found for this domain. Despite having

the highest mean value, this study presents vital evidence for the insufficiency of participants' knowledge in the content and concepts of LTA. The participants in this study, English teachers who work at state schools, were of the opinion that the training they had in this domain was low. This finding is in good agreement with the previous findings of the studies by Mede and Atay (2017: p. 52), Sarıyıldız (2018: p. 114), Tamerer (2019: p. 36), and Vogt Tsagari (2014: p.383), in which all the researchers reported insufficiency of LAL in this domain. When the components of the content and the concepts of LTA were analyzed individually, especially, “testing and assessing aspects of culture” and “using statistics to study the quality of tests/assessment” seemed to have been neglected. It is, therefore, unlikely that teachers will attempt to assess any cultural objects in the assessment tools they utilize. In the same domain, more than half of the participants rated their perceived training level for “testing and assessing grammar and vocabulary” as advanced, which was not a surprising finding considering the importance that is attached to grammar and vocabulary learning in Turkey. On the other hand, some of the findings of the current study are in contradiction with previous findings in the literature. To exemplify, while in this study, the perceived training level of the teachers was found insufficient in the concepts of reliability and validity, the preservice teachers in Viengsang's (2006: p. 437) study reported that they had adequate knowledge of such issues. Although it is difficult to explain this rather contradictory result, it may be explained by the fact that the participants in that study (preservice teachers) still retained the knowledge about those concepts since they studied such concepts not long ago.

Finally, when the results from all three domains were compared, it was uncovered that the participants in the current study had the most knowledge in the content and concepts of LTA. This result matches some of the earlier studies by Şahin (2019: p. 251), Vogt and Tsagari (2014: p. 385), and Tamerer (2019: p. 27). Although the percentages of participants' rating their training level are relatively high in this domain, the overall LAL of the teachers was found to be unsatisfactory. This result corroborates the findings of a great deal of previous studies which also investigated the amount of LAL of teachers (Hasselgreen et al., 2004: p. 10; Mede and Atay, 2017: p.52; Muhammad and Bardakçı, 2019: p. 431; Ölmezer-Öztürk, 2018: p. 150; Sarıyıldız, 2018: p. 115; Şahin, 2019: p. 251;

Tamerer, 2019: p. 36; Viengsang, 2006: p. 440; Vogt and Tsagari, 2014: p. 385). Drawing upon the results of the aforementioned studies, it can be stated that the training at teacher training institutions is not sufficient. It is critical to highlight that, most of the time, the ELTE courses that are offered at universities are the only chances for teachers to be language assessment literate. However, it was found out by numerous researchers (Deluca and Bellara, 2013: p. 356; Hatipoğlu, 2010: p. 49; Lam, 2005: p. 1; Ölmezer-Öztürk, 2018: p. 153) that those courses are far from equipping the preservice teachers with the necessary skills to assess their learners adequately due to various reasons such as insufficiency of classroom time or lack of balance between theory and practice. Therefore, it can be concluded that enhancing the content of ELTE courses will yield to higher LAL.

The second research question in the current study sought to examine the perceived training needs of the English teachers who work at state schools in Turkey in LAL under such subparts as classroom-focused LTA, the purposes of testing and the content and concepts of LTA. To this end, first of all, 6 items in classroom-focused were analyzed. The results revealed that the component which needed the most training was “using the European language portfolio or other portfolios,” while the participants reported the least need for training in “using ready-made tests.” Using portfolios was observed to be the component that is needed the most training in this domain, and this finding agrees with the results from other studies such as Tamerer (2019: p. 28) and Vogt and Tsagari (2014: p. 381). It is significant at this point to note that this finding is also in line with findings from the current study in terms of perceived learning levels of the participants, in which they reported the least knowledge in using portfolios. In the same vein, it should also be noted that more than half of the participants (51%) expressed no need for further training in “using ready-made tests.” This latter finding may be explained by the over-reliance of teachers on the traditional assessment techniques in the classrooms, which was also found out by numerous researchers (Berry et al., 2019: p. 121; Şahin, 2019: p. 252). Another possible explanation of this might be the lack of teachers’ skills in preparing their own tests, just like the participants in Viengsang’s (2006: p. 438) study reported. All in all, when all the items were examined, the overall mean score was found to be .93, which indicated a basic perceived need in this domain.

The next part of the questionnaire investigated the training needs of the participants in the domain of purposes of testing. Therefore, 4 items in this domain were examined, and the overall mean value was found to be .93, which suggests that the LAL of English teachers who work at state schools is not sufficient. This finding is consistent with those of Sarıyıldız (2018: p. 118) and Tamerer (2019: p. 38), in which the participants also indicated a basic need in this domain. Moreover, the highest score belonged to “awarding students final certificates.” This finding is also in line with the results that emerged from the first research question of the current study, which revealed that the participants received the lowest training in this component. On the other hand, the lowest mean value was found for “giving grades,” which suggests that almost half of the participants (45%) do not need further training in this component. This finding could be explained by exam-oriented education in Turkey. As also found out by Lam (2005: p. 182), this fact makes teachers’ perceptions about assessment exam-oriented, and giving grades becomes their strongest suit and, therefore, an inevitable routine in their career. It is also crucial to note that the participants indicated a basic need in this domain even though their overall training level was found to be insufficient. This contradictory result may be attributed to the fact that the participants do not need advanced skills at the schools where they work due to the low proficiency levels of their students or time constraints.

Lastly, the study aimed to find out how much further training the participants thought they needed in the content and concepts of LTA. When the overall mean value was calculated, it was found to be .95, which indicated that the participants in the study perceived basic training in this domain. This finding is in agreement with the previous ones by Sarıyıldız (2018: p. 119) and Tamerer (2019: p. 38), who also uncovered a basic need of participants in this domain. When the components of the content and concepts of LTA were examined individually, it was realized that while some of the results were in line with the previous research, some others were not consistent with them. To illustrate, the lowest perceived need was identified in the component of “testing and assessing grammar and vocabulary,” which was not surprising considering the importance attached to the grammar and over-reliance on the use of traditional assessment tools in Turkey. When the distribution of the participants’ ratings in this component was examined, some

discrepancies were observed. For example, while in the current study, the majority of the participants (43.6%) reported no further need in testing and assessing grammar and vocabulary, the majority of the participants (44.4%) in Vogt and Tsagari's (2014: p. 384) study indicated that they needed more advanced training in this component. This rather contradictory result could be due to the importance attached to grammar and vocabulary teaching in Turkey. In the same vein, the finding in this study related to testing and assessing grammar and vocabulary supports the findings by Mede and Atay (2017: p. 55), who reported that 68% of their participants who worked in Turkey also expressed no need for further training in this component. Finally, It is also significant to note that a similar number of people rated their need as none, basic or more advanced (37%, 34%, 30%, respectively) in testing and assessing productive skills. These varying needs in training could be attributed to the opportunity of the participants to assess productive skills at the schools where they work.

All in all, when 18 items in all three domains in LTA regarding the perceived needs of the participants for further training were analyzed, the mean value was found to be .93. Drawing upon this finding, it can be concluded that teachers who work at state schools in Turkey perceive a need for basic training. This finding is also in line with the findings of the studies by Sarıyıldız (2018: p. 120) and Tamerer (2019: p. 38), in which the participants also expressed a need for basic training. Moreover, the overall findings from the current study indicated that there is a need for training across all the components with varying percentages. In this regard, the findings from the current study corroborate the previous findings in the literature (Büyükkarcı, 2016: p. 338; Doğru, 2020: p. 60; Hasselgreen et al., 2004: p. 11; Jannati, 2015: p. 35; Kavaklı and Arslan, 2019: p. 636; Lam, 2015: p. 169; Mede and Atay, 2017: p. 58; Ölmezer-Öztürk and Aydın, 2019: p. 617; Semiz and Odabaş, 2016; Sarıçoban, 2011: p. 405; 2014; Vogt and Tsagari, 2014: p. 385; Xu and Brown, 2017: p. 152)

Finally, in order to support the findings from the questionnaire, written interviews were conducted with 18 of the teachers who filled out the questionnaires. Five open-ended questions were asked so as to collect more in-depth data about the training levels and needs of the participants. The collected data were analyzed one by one.

In the first interview question, in order to shed light on the reasons why the participants had insufficient LAL, the participants were asked if the school experience course they took at the university gave them a chance to put what they learned at university into practice and contribute to their professional development. In response to this question, the majority of the interviewees voiced some limitations. To start with, many of the participants indicated that the school experience course was not sufficient. Looking at the one of the participant's response *“actually, the school experience course we took at the university surely helped us; however, when we found ourselves in a real teaching atmosphere...”* (T4), it can be concluded that the practice at practicum schools are far from preparing preservice teachers for real life-like learning and teaching atmosphere. Another reason why the school experience course was not of help for the preservice teachers was due to the advisory teachers at practicum schools. As an interviewee put *“advisory teacher at the practicum school never involved us in testing and assessment process; therefore, we did not have any chance to put what learned at the university into practice.”* (T18), lack of involvement in assessment processes at the practicum schools takes away the only chance that preservice teachers have before getting into the actual classrooms. Hence, it can be concluded that active involvement, as also highlighted by Kavaklı and Arslan (2019: p. 622), is of crucial importance. Furthermore, the limited time at the practicum schools was also mentioned as a reason for the ineffectiveness of the school experience course, as voiced by one of the interviewees *“school experience course was not very helpful. Since the amount of time we participated in the lessons was very limited, we did not have much time to practice what we learned.”* (T5) Hence, it could conceivably be hypothesized that if more time was allocated for practice at practicum schools, school experience courses would be more effective. Talking about this issue, it is also crucial to refer to the answer provided by one of the interviewees as he criticized his supervisor at university for the ineffectiveness of the course as he put it: *“... our supervisor teacher for the school experience course was someone out of the teacher education field (his undergraduate degree was in translation), so the only thing he suggested to us is to teach in English. Nothing about what to teach, how to teach, or what kind of strategies to use in classroom discourse in terms of motivating students, promoting their involvement or maintaining discipline, etc.* (T15). Drawing upon this answer, it can be concluded that

even teacher trainers may be weak in teaching assessment skills. All in all, in an attempt to answer the first research question, the factors mentioned above such as the artificial practice environment, exclusion of preservice teachers in the assessment process and inefficacy of teacher trainers may explain and also support the findings from the current study and other studies in the literature (Hasselgreen et al., 2004: p. 11; Mede and Atay, 2017: p. 57; Ölmezer-Öztürk, 2018: p. 104), which reported that the LAL that teachers had was insufficient.

In order to come up with more in-depth knowledge regarding the perceived training levels of the participants (first research question), the second question in the interview investigated the readiness of the teachers in assessing students' learning when they graduated from university. The majority of the interviewees voiced that they were not prepared enough, and some common points emerged from interviews as to how they coped with this insufficiency. First of all, the majority of the interviewees stated that they received help from their colleagues; thus, it can be suggested that there is a good network of help, also suggested by Jin (2010: p. 555), among the English teachers who work at state schools in Turkey. Another common point that was identified in the responses of the interviewees was the fact that they learned testing and assessment in the course of time. This could possibly be explained due to the inefficiency of ELTE courses at university. In this vein, as also stated by another interviewee, "*whether the students filled out the gaps correctly, or they answered multiple-choice questions correctly was not sufficient for me to assess their language learning. The ELTE course I took did not have the logic that language is for learning and speaking.*" (T10), it can be concluded that ELTE courses offered at universities fail to facilitate preservice teachers' assessment skills. This perceived inadequacy of the ELTE courses could result from the limited time to practice due to the crowdedness of the classrooms at university (Deluca and Bellara, 2013: p. 366; Hatipoğlu, 2010: p. 210), or lack of balance between theory and practice (Şahin, 2019: p. 226). This finding from the interviews also confirms the quantitative findings for the first research question of the study, in which participants reported inadequate LAL levels. It is also crucial to mention the comment made by another interviewee regarding the teacher education institutions in Turkey as "... *I did not feel ready because the education given at*

the universities does not correspond to the realities of the country. The school to which I was appointed for my compulsory service had its all internal dynamics where we could not make an assessment through common knowledge” (T11). This finding could also explain the findings from the quantitative part of the study regarding why some teachers did not express a further need for training in some domains even though they had insufficient knowledge. It is probable that some teachers do not even have the opportunity at the schools where they work to make use of their assessment skills. The comment made by another interviewee, *“I first started to work at a regular high school. When I started to work there, I taught my lessons quite easily. Sometime later, I started to work at Anatolian high school. There, I realized that I was not done with learning since my students had a good background in English and expected more from me; therefore, I had to prepare for my lessons, which helped me to improve myself, too.” (T4)* also supports findings from the questionnaire as to why teachers do not wish further education in some domains. Drawing upon these findings, it can be hypothesized that not all the assessment skills are required at all schools; therefore, even though, in response to the first research question, some teachers reported a basic training level, they did not rate a need more advanced training while their perceived needs were examined in the second research question.

The next question in the interview aimed to serve both the first and second research questions by finding out whether the participants had received any training in language testing and assessment (LTA) after they finished university and if they wanted to have any LTA courses. The answers of the interviewees revealed that only two of them had received an LTA course before; one of them was a TESOL training, and the other one was a training held by the Ministry of National Education. This finding supports the findings for the first research question in the quantitative part of the current research, and some other research in the literature (Sarıyıldız, 2018: p. 115; Tamerer, 2019: p. 36), in which the participants did not think that they received more advanced training in all three domains. In other words, although the participants thought that they did not receive sufficient training in LTA, they did not intend to improve themselves by attending professional development courses. There are two likely causes for this fact; they either did not want to receive any

training, or they did not have a chance to do it. Moreover, the types of courses that the participants wished to have also corroborated the findings from the quantitative part of the current study. To illustrate, interviewees reported a desire for courses on SPSS and portfolio assessment, which were also found to have the highest mean scores ($M=1.04$, $M=1.12$, respectively) with respect to the perceived needs of the participants (second research question) in the questionnaire. Another course that the interviewees wanted to have was on test preparations. This finding is also in line with the finding from the study by Dođru (2020: p. 36), in which over half of the participants (%59.4) wished to learn about designing tests. A possible explanation for this might be a lack of training and confidence of teachers in this component, as also stated by Campbell and Evans (2010: p. 354) and Viengsang (2006: p. 418). On the other hand, this finding related to test preparation can be said to contradict the results that emerged from the questionnaire data, which showed that almost half of the participants (47.5%) perceived that they had advanced training in preparing classroom tests.

The fourth question in the interview investigated how the participants assessed their learners. The findings from this question provided insights for the perceived training levels and needs of the participants. It was revealed that the most common tool that was used was written exams. This finding was not surprising since it is obligatory to conduct at least two written exams for one semester at all the state schools in Turkey, which explains why the participants perceived the lowest needs in preparing classroom tests or using ready-made tests in the classrooms since they practice this skill quite often. On the other hand, it is significant to note that alternative assessment tools were also reported as they were used by the interviewees. One unanticipated finding was that portfolios were utilized by many of the participants. This finding is in contradiction with the findings from the quantitative part of this study, which indicated that the participants perceived the lowest training level ($M=0.86$) in this component. On the other hand, other alternative assessment tools, which were stated in the interviews, such as self- and peer-assessment, could be said to be in agreement with quantitative findings, which indicated that the participants had basic training in using these tools. Drawing upon these findings, it can be argued that even though alternative assessment tools, especially portfolios, are common

tools that are utilized at state schools, the effectiveness of those is questionable due to the lack of training the teachers had in them.

Finally, in order to get a more comprehensive opinion about the classroom-focused LTA of the participants in terms of the recent assessment tools such as portfolio or self-assessment and how often they were utilized by the participant teachers, a final question was addressed to participants. All of the participants stated that they made use of alternative assessment tools with varying frequencies. As for the reasons why some of the use of the alternative assessment tools was minimal for participants, insufficient time in the curriculum emerged as an important factor, which can be realized by the following comments:

“I have information about the recent LTA methods, but I have never used them. In 3 hours a week, we try to catch up with the curriculum, improve the students’ test-taking skills, have them memorize vocabulary items, etc.; therefore, these methods are ignored.” (T10)

“I know them, but I have never used them. We have limited time for the courses; 5 hours for 9th graders and 2 hours for the others. It may even be impossible to complete all the subjects in the curriculum in that much time.” (T17)

These comments are significant since it may be concluded that teachers working at state schools race against time; therefore, due to limited time, it is quite unlikely to expect them to pay attention to improving productive skills, too. This finding also agrees with the finding from the quantitative data, which indicated that the majority of the participants (36.6%) did not indicate further need for testing and assessing the productive skills. Another reason for the limited use of alternative assessment tools was not having the opportunity to use them as one interviewee put: *“I have heard and used peer-assessment and self-assessment before. When I worked at a high school, I used to use them often. However, at the elementary school where I now work, they are not easily applicable tools, so I cannot use them.” (T1)*

This comment by T(1) is in line with the comment made by T (4) in response to the second question, in which he explained how altering the schools he worked at affected

the skills he required. Responses from both participants suggest that what is required from a teacher and what a teacher can do differ depending on the schools where they work.

5.3. PEDAGOGICAL IMPLICATIONS OF THE STUDY

The findings of this current study could have implications for the Council of Higher Education, specifically teacher training institutions as well as the General Directorate of Teacher Training and Education and the Board of Education within the Ministry of Education in Turkey.

The overall findings of this study revealed that English language teachers who work at state schools in Turkey had insufficient LAL, which is a vital mastery (O'loughlin, 2006: p. 71). Therefore, first of all, the courses offered at universities in Turkey should be revised in terms of their content and number. When more assessment courses are added to the curriculum with a more focus on practice, it is likely to increase the LAL of the teachers. Also, it was found out in this research that school experience courses are not taken very seriously in Turkey, which hampers the likelihood of turning the theoretical knowledge into practice. In this vein, the Council of Higher Education and The Ministry of Education has to work coordinately, and some standards for the school experience course content and requirements of preservice teachers at practicum schools such as more active participation, the minimum amount of teaching practice, etc. should be officialized by the two parties. Finally, given that the perceived levels and needs of the participants are not satisfactory, English language teachers working at state schools are in a definite need of in-service training which needs to be held nationwide.

5.4. LIMITATIONS OF THE STUDY AND SUGGESTIONS FOR FUTURE RESEARCH

This study aimed to investigate the perceived training levels and needs of English teachers working at state schools in Turkey. The most significant limitation of this study is the sample size. The quantitative data were collected from 101 teachers, and 18 of them volunteered to participate in the interview.

Based on the findings and limitations of this study, further research may be conducted with more participants from various cities in order to make the results more generalizable. Furthermore, considering the statements of the participants in the interviews who complained about the lack of opportunity to apply assessment skills at some cities or levels, a similar study could also be conducted in a specific city or level in order to shed more light on the LAL of teachers at particular settings. A final suggestion which also came from this research would be to investigate the way teachers utilize the alternative assessment methods, especially portfolios, as it was revealed in the current study that although the participants perceived the least training in using them, the findings from interviews indicated that portfolios are used by the majority of teachers frequently.

REFERENCES

1. Books

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Berger, A. (2012). "Creating language assessment literacy: A model for teacher education." In J.I. Hüttner, B. Mehlmauer-Larcher, S. Reichl, B. Schiftner (Eds.), *Theory and practice in EFL teacher education: Bridging the gap*. Bristol: Multilingual Matters, 57-82.
- Brady, L., Kennedy, K. (2014). *Curriculum construction (5th ed.)*. Frenchs Forest, NSW: Pearson Australia.
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Pearson Education.
- Caspary, M., Boothe, D. (2017). "Assessment of EFL through the process of problem-based learning." In R. Al-Mahrooqi, C. Combee, F. Al-Maamari, V. Thakur (Eds), *Revisiting EFL assessment*. Cham, Switzerland: Springer, 9-27
- Coombe, C., Troudi, S., Al-Hamly, M. (2012). "Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations." In C. Coombe, P. Davidson, B. O'Sullivan, S. Stoyhoff. (Eds.). *The Cambridge guide to second language assessment*. Cambridge University Press, USA, 20-29.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Boston: Pearson Education.
- DiRanna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K., Carnahan, D., Strang, C. (2008). *Assessment-centered teaching: A Reflective Practice*. Thousand Oaks, CA: Corwin Press.
- Djoub, Z. (2017). "Assessment literacy: Beyond teacher practice." In R. Al-Mahrooqi, C. Combee, F. Al-Maamari, V. Thakur (Eds), *Revisiting EFL assessment*. Cham, Switzerland: Springer, 9-27.
- Dörnyei, Z., Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.
- Douglas, D. (2014). *Understanding language testing (2nd ed.)*. New York, NY: Routledge.
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. Alexandria, Virginia USA: ASCD.
- Gummer, S., Shepardson, D.P. (2001). "Facilitating change in classroom assessment practice: Issues for professional development." In Shepardson D.P. (ed.), *Assessment in science*. Springer, Dordrecht, 53-66.

- Huba, M. E., Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning* (1st ed.). Pearson.
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, L. Or, S. May (Eds.), *Language testing and assessment* (pp. 257-270). Cham, Switzerland: Springer
- Kavaklı, N., Arslan, S. (2019). “When student-teachers’ benefits are not enough: The critical perspective of action-impact-desired outcome model for the evaluation of instruction.” In B. Tuncsiper, F. Sayın, B. Herguner, F. Irem Aydin (Eds.). *Selected discussions on social science research*. Frontpage Publications: London, 622-640.
- Lantolf, J. P., Poehner, M. E. (2008) “Dynamic Assessment.” In E. Shohamy, N. H. Hornberger (eds.), *Language testing and assessment* (2nd Ed., Volume 7, pp. 273-284).
- Lidz, C.S., Gindis, B. (2003). “Dynamic assessment of the evolving cognitive functions in children.” In A. Kozulin, B. Gindis, V. S. Ageyev, S. M. Miller (Eds.), *Vygotsky’s educational theory in cultural context*. Cambridge: Cambridge University Press, 99-118.
- Mackey, A., Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McNamara, T. (2004). “Language testing.” In A. Davies, C. Elder (Eds.). *The Handbook of applied linguistics*. Malden, MA: Blackwell, 763-783.
- Miller, M. D., Linn, R. L., Gronlund, N. E. (2008). *Measurement and assessment in teaching* (10th ed.). New Jersey: Pearson.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer Science & Business Media.

2. Articles, Papers, Other Printed Publications

- Berry, V., Sheehan, S., Munro, S. (2019, April). “What does language assessment literacy mean to teachers?” *ELT Journal*, 73(2), 113-123. doi:10.1093/elt/ccy055
- Birgin, O., Baki, A. (2007) “The Use of portfolio to assess student’s performance.” *Journal of Turkish Science Education*, 4(2), 75-90.
- Büyükkarcı, K. (2016). “Identifying the areas for English language teacher development: A study of assessment literacy.” *Pegem Eğitim ve Öğretim Dergisi*, 6(3), 333-346. <http://dx.doi.org/10.14527/pegegog.2016.017>
- 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.

- Coombe, C., Vafadar, H., Mohebbi, H. (2020). "Language assessment literacy: What do we need to learn, unlearn, and relearn?" *Language Testing in Asia*, 10 (3), 1-16. <https://doi.org/10.1186/s40468-020-00101-6>
- Corbin, J., Strauss, A. (1990). "Grounded theory research: Procedures, canons, and evaluative criteria." *Qualitative Sociology*, 13(1), 3-21.
- Çalışkan, H., Kaşıkçı, Y. (2010). "The application of traditional and alternative assessment and evaluation tools by teachers in social studies." *Procedia-Social and Behavioral Sciences*, 2, 4152-4156.
- Davin, K. J., Heineke, A. J. (2016). "Preparing teachers for language assessment: A practice-based approach." *TESOL Journal*, 7(4), 921-938.
- 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>
- Doğru, D. (2020). Assessment literacy conceptions and practices EFL teachers in higher education. Master's thesis, Bahçeşehir University, The Graduate School of Educational Sciences, İstanbul.
- Farhady, H. (2019). "A cross-contextual perspective on EFL teachers' assessment knowledge." *EDU7*, 8(10), 1-19.
- Fulcher, G. (2012). "Assessment literacy for the language classroom." *Language Assessment Quarterly*, 9(2), 113–132.
- Giraldo Aristizabal, F. G. (2018). "A diagnostic study on teachers' beliefs and practices in foreign language assessment." *İkala: Revista de Lenguaje y Cultura*, 23(1), 25-44. doi:10.17533/udea.ikala.v23n01a04
- Giraldo, F. (2019). "Language assessment practices and beliefs: Implications for language assessment literacy." *HOW*, 26(1), 35-61. <https://doi.org/10.19183/how.26.1.481>
- Gözüyeşil, E., Tanrıseven, I. (2017). "A meta-analysis of the effectiveness of alternative assessment techniques." *Eurasian Journal of Educational Research*, 70, 37-56, DOI: <http://dx.doi.org/10.14689/ejer.2017.70.3>
- Hakim, B. (2015). "English language teachers' ideology of ELT assessment literacy." *International Journal of Education & Literacy Studies*, 3(4), 42-48.
- Harding, L., Phil, J. (2013). "Defining the language assessment literacy gap: Evidence from a parliamentary inquiry." *Language Testing*, 30(3), 381–402.
DOI: 10.1177/0265532213480337
- Hatipoğlu, 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.

- Hatipoğlu, 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.
- Jannati, S. (2015). "ELT teachers' language assessment literacy: Perceptions and practices." *The International Journal of Research in Teacher Education*, 6(2), 26-37.
- Jeong, H. (2013). "Defining assessment literacy: Is it different for language testers and non-language testers?" *Language Testing*, 30 (3), 345-362.
- Jin, Y. (2010). "The place of language testing and assessment in the professional preparation of foreign language teachers in China." *Language Testing*, 27 (4), 555-584.
- Köksal, D. (2004). "Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net." *Turkish Online Journal of Distance Education (TOJDE)*, 5(1), 1-11.
- Lam, R. (2015). "Language assessment training in Hong Kong: Implications for language assessment literacy." *Language Testing*, 32(2), 169-197.
- Lantolf, J. P., M. E. Poehner. (2004). "Dynamic assessment: Bringing the past into the future." *Journal of Applied Linguistics* 1, 49-74.
- Malone, M. E. (2013). "The essentials of assessment literacy: Contrasts between testers and users." *Language Testing*, 30(3), 329-344.
- Mede, E., Atay, D. (2017). "English language teachers' assessment literacy: The Turkish context." *Dil Dergisi*, 168 (1), 1-5.
- Mertler, C. A., Campbell, C. (2005). "Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "assessment literacy inventory"." Online Submission.
- Muhammad, F. H. N., Bardakç1, M. (2019). "Iraqi EFL Teachers' Assessment Literacy: Perceptions and Practices." *Arab World English Journal*, 10 (2) 420-430. doi: <https://dx.doi.org/10.24093/awej/vol10no2.33>
- 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.
- 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.

- Orr, M. (2010). "Evaluating the testing course in an MA in ELT." *English Language Teacher Education and Development (ELTED)*, 13, 4-14.
- Ölmezer-Öztürk, E. (2018). Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers. Doctoral thesis, Anadolu University, Graduate School of Educational Sciences, Eskişehir.
- Ölmezer-Öztürk, E., Aydın, B. (2019). "Investigating language assessment knowledge of EFL teachers." *Hacettepe University Journal of Education*, 34(3), 602-620. doi:10.16986/HUJE.2018043465
- Öz, S., Atay, D. (2017). "Turkish EFL instructors' in-class language assessment literacy perceptions and practices." *ELT Research Journal*, 6(1), 25-44.
- Pekbay, C., Koray, O. (2020). "Alternative assessment tools based on a feedback process: Perceptions of pre-service science teachers." *Malaysian Online Journal of Educational Sciences*, 8(1), 50-60.
- Pill, J., Harding, L. (2013). "Defining the language assessment literacy gap: Evidence from a parliamentary inquiry." *Language Testing*, 30(3), 381-402.
- Popham, W. J. (2009). "Assessment literacy for teachers: Faddish or fundamental?" *Theory Into Practice*, 48, 4-11.
- Popham, W. J. (2011). "Assessment literacy overlooked: A teacher educator's confession." *The Teacher Educator*, 46(4), 265-273.
- Purpura, J. E. (2016). "Second and foreign language assessment." *The Modern Language Journal*, 100, 190-208.
- Sarıçoban, A. (2011). "A study on the English language teachers' preparation of tests." *Hacettepe University Journal of Education*, 41, 398-410.
- Sarıyıldız, G. (2018). A study into language assessment literacy of preservice English as a foreign language teachers in Turkish context. Master's thesis, Institute of Social Sciences. Hacettepe University, Ankara, Turkey.
- Scarino, A. (2013). "Language assessment literacy as self awareness: Understanding the role of interpretation in assessment and in teacher learning." *Language Testing*, 30(3), 309-327. <http://dx.doi.org/10.1177/0265532213480128>
- Shim, K. N. (2009). An investigation into teachers' perceptions of classroom-based assessment of English as a foreign language in Korean primary education. Doctoral dissertation, University of Exeter, Exeter.
- Stanford, P., Reeves, S. (2005). "Assessment that drives instruction." *Teaching Exceptional Children*, 37(4), 18-22.
- Şahin, S. (2019). An analysis of English language testing and evaluation course in English language teacher education programs in Turkey: Developing language assessment

- literacy of pre-service EFL teachers. Doctoral thesis, Middle East Technical University, The Graduate School of Social Sciences, Ankara, Turkey.
- Talmy, S., Richards, K. (2010). "Theorizing qualitative research interviews in applied linguistics." *Applied linguistics*, 32(1), 1-5.
- Tamerer, R. (2019). An investigation of Turkish pre-service EFL teachers' language assessment literacy. Master's thesis, Kocaeli University, Institute of Social Sciences, Kocaeli.
- Téllez, K., Mosqueda, E. (2015). "Developing teachers' knowledge and skills at the intersection of English language learners and language assessment." *Review of Research in Education*, 39, 87–121. doi: 10.3102/0091732X14554552
- Tsagari, D., Vogt, K. (2017). "Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects." *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Viengsang, R. (2016). "Exploring pre-service English teachers' language assessment literacy." *Modern Journal of Language Teaching Methods (MJLTM)*, 6(5), 432-442.
- Vogt, K., Tsagari, D. (2014). "Assessment literacy of foreign language teachers: Findings of a European study." *Language Assessment Quarterly*, 11(4), 374-392.
- 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.
- Xie, Q., Tan, S. (2019). "Preparing primary English teachers in Hong Kong: Focusing on language assessment literacy." *The Journal of Asia TEFL*, 16(2), 653-673.
- Xu, Y., Brown, G. T. L. (2017). "University English teacher assessment literacy: A survey-test report from China." *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Yastıbaş, A. E., Takkaç, M. (2018). "Understanding language assessment literacy: Developing language assessments." *Journal of Language and Linguistic Studies*, 14(1), 178-193.
- Yetkin, C. (2015). An investigation on ELT teacher candidates' assessment literacy. Master's thesis, Cag University, Institute of Social Sciences, Mersin.

3. Electronic Sources

- American Federation of Teachers. (1990). Standards for teacher competence in educational assessment of students. <https://buros.org/standards-teacher-competence-educational-assessment-students> / May, 2020.

- Hasselgreen, A., Carlsen, C., Helness, H. (2004). European survey of language testing and assessment needs. Report: Part one – general findings. Retrieved from <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf> / May, 2020.
- MEB (Milli Egitim Bakanligi) (2019). Ilkogretim Turkce dersi ogretim programi. [The curriculum of primary school Turkish course] Ankara: MEB Yayinlari. Retrieved from <http://mufredat.meb.gov.tr/Dosyalar/20195716392253-02-Türkçe%20Öğretim%20Programı%202019.pdf> / March, 2020.
- Semiz, Ö., Odabaş, K. (2016, June). Turkish EFL teachers' familiarity of and perceived needs for language testing and assessment literacy. Proceeding of the 3rd international linguistics and language studies conference, İstanbul, Turkey. Retrieved from https://www.academia.edu/34827097/Turkish_EFL_Teachers_Familiarity_of_and_Perceived_Needs_for_Language_Testing_and_Assesment_Literacy / April, 2020.
- 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 / April, 2020.

APPENDICES

APPENDIX A. In-Service Language Teachers' Questionnaire

Dear Participant,

This survey aims to find out your TRAINING LEVEL and your TRAINING NEEDS in Language Testing and Assessment (LTA).

Please be assured that there are no "right" or "wrong" answers.

You do not need to give your name or the name of the institution you work at.

Your participation in this study is completely voluntary, and all your information in this survey is confidential.

Thank you very much for your time and support.

If you agree to participate in this research, please click Next/Sonraki

Sezgin Ballıdağ

Kocaeli University English Language Teaching Department

PART I. GENERAL INFORMATION

1. What is your gender? Male Female
2. How many years of teaching experience do you have?
0-1 2-5 6-10 11-15 More than 15
3. Which university did you graduate from?
4. Which BA program did you graduate from?
English Language Teaching (ELT) Non-ELT
5. Which city do you work in?
6. Which level are you teaching at?
Elementary School Middle School High School
7. Did you have a separate testing and assessment course when you were at university? Yes No

PART II. 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 or from other sources			
c) Giving feedback to students based on information from tests/assessment			
d) Using self- or peer-assessment			
e) Using informal, continuous, non-test type of assessment			
f) Using the European Language Portfolio, an adaptation of it or some other portfolio			

1.2. 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			
d) Awarding final certificates (from school/program; local, regional or national level)			

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			
d) Awarding final certificates (from school/program; local, regional or national level)			

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 and 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 need training in the following domains.

	None	Yes, basic training	Yes, more advanced training
1. Testing and 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			

APPENDIX B. Interview Questions

Gender: Male Female

Where do you work: Elementary School

Middle School

High School

Which university did you graduate from?

Which city do you work in?.....

Years of Experience: 0-1 2-5 6-10 11-15 More than 15 years

Highest Qualification: BA MA PhD

Have you taken LTA course at university: Yes No

There are 5 questions and you can answer the questions in Turkish or English.

Please try to elaborate on your ideas and give as many details as possible.

1. Did your school experience course you took at the university give you the chance to put what you learned related to language testing and assessment into practice, and contribute to your professional development? Please explain.

1. Üniversitede aldığımız okul deneyimi dersi size Dilde Ölçme ve Değerlendirme ile ilgili öğrendiğiniz bilgileri uygulamaya koyma fırsatı verdi ve sizin mesleki gelişiminize katkıda bulundu mu? Lütfen açıklayınız.

.....

2. Did you feel prepared enough to assess your students' language learning when you finished university? If not, what kind of difficulties did you have in terms of assessing language learning, and how did you overcome them?

2. Üniversiteyi bitirdiğiniz zaman, kendinizi öğrencilerinizin dil öğrenme seviyelerini değerlendirmek için yeteri kadar hazır hissediyor muydunuz? Eğer cevabınız “hayır” ise, dil öğrenimini değerlendirme açısından ne gibi zorluklar yaşadınız, ve bu zorlukların nasıl üstesinden geldiniz?

.....
3. After you finished university, have you received any training in language testing and assessment (LTA)? If yes, what was the focus of this training? If no, in which domain would you like to receive one?

3. Okulu bitirdikten sonra, hiç Dilde Ölçme ve Değerlendirme ile ilgili eğitim aldınız mı? Eğer aldıysanız, eğitimin konusu neydi? Eğer almadıysanız, ne alanda eğitim almak isterdiniz?

.....

4. How do you usually assess your student's language knowledge in the classroom? What kind of assessment instruments do you use?

4. Öğrencileriniz yabancı dil bilgilerini sınıfta genellikle nasıl ölçersiniz? Ne tür değerlendirme araçları kullanırsınız?

.....

5. Do you know about more recent Language Testing and Assessment (LTA) methods e.g., portfolio assessment, self- or peer-assessment? Have you ever tried them? How often do you use them?

5. Portfolyo değerlendirmesi, öz değerlendirme, akran değerlendirmesi gibi güncel Dilde Ölçme ve Değerlendirme metotları hakkında bilgi sahibi misiniz? Hiç onları kullandınız mı? Onları ne sıklıkta kullanırsınız?

.....

APPENDIX C. Ethics Committee Approval

Evrak Tarih ve Sayısı: 24/03/2020-E.25664



T.C.
KOCAELİ ÜNİVERSİTESİ
Sosyal ve Beşeri Bilimler Etik Kurulu



Sayı : 10017888-044/
Konu : Anketler

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 10/03/2020 tarihli, 21783 sayılı ve "Anketler" konulu yazı

Sosyal ve Beşeri Bilimler Etik Kurulunun 19/03/2020 tarih ve 2020/04 nolu toplantısında alınan 7 sıra sayılı kararı aşağıda sunulmuştur.

Gereğini arz ederim.

Prof.Dr. Adem ÇAYLAK
Kurul Başkanı

Karar No 7: Sosyal Bilimler Enstitüsü Müdürlüğü'nün 10/03/2020 tarih ve 21783 sayılı yazısı görüşüldü. İngiliz Dili Eğitimi yüksek lisans programı öğrencisi Sezgin BALLIDAG'ın, Doç. Dr. Banu İNAN KARAGÜL'ün danışmanlığında yürüttüğü "Türk İngilizce Öğretmenlerinin Dil Değerlendirmesi Okuryazarlığı Araştırması" başlıklı tezi kapsamında internet üzerinden uygulayacağı anketin, bilimsel araştırma ve yayın etiği açısından bir sakınca olmadığına oy birliği ile karar verildi.

CURRICULUM VITAE

Sezgin Ballidag

E-Mail: sballidag@gmail.com

Education:

2000 – 2005 **Hacettepe University**
Bachelor of Arts English Language Teaching

Experience :

2014 - **Yildiz Technical University**
Lecturer
School of Foreign Languages

2006 – 09/2013 **Gebze Technical University**
Foreign Languages Department
Instructor

2005 - 2006 **University of Richmond, the USA**
Fulbright Foreign Language Teaching Assistant (FLTA)
Modern Languages Department

Certificates :

2006 CELTA, British Side

Scholarships :

2005 – 2006 Fulbright Foreign Language Teaching Assistant (FLTA)
University of Richmond, VA, the USA,

Articles Published in Journals:

Ballidag, S. (2020). The effect of using visual aids in cloze tests testing grammar at B1 level: A quasi-experimental study. *Kocaeli University Journal of Education*, 3 (1), 26-40. DOI: 10.33400/kuje.682241

Ballidag, S. (2020). Does social media use affect our real-life relationships: Focus on gender. *European Journal of Education Studies* 6, 163-170.

Ballidag, S. (2018). Using wikis to increase writing skills in writing classes. *The International Journal of Educational Researchers*, 9, 12-17.